



UNION INTERNATIONALE DES TÉLÉCOMMUNICATIONS

UIT-T

SECTEUR DE LA NORMALISATION
DES TÉLÉCOMMUNICATIONS
DE L'UIT

H.263

Appendice III
(06/2001)

SÉRIE H: SYSTÈMES AUDIOVISUELS ET
MULTIMÉDIAS

Infrastructure des services audiovisuels – Codage des
images vidéo animées

Codage vidéo pour communications à faible débit

**Appendice III: Exemples d'implémentation du
codeur/décodeur H.263**

Recommandation UIT-T H.263 – Appendice III

RECOMMANDATIONS UIT-T DE LA SÉRIE H
SYSTÈMES AUDIOVISUELS ET MULTIMÉDIAS

CARACTÉRISTIQUES DES SYSTÈMES VISIOPHONIQUES	H.100–H.199
INFRASTRUCTURE DES SERVICES AUDIOVISUELS	
Généralités	H.200–H.219
Multiplexage et synchronisation en transmission	H.220–H.229
Aspects système	H.230–H.239
Procédures de communication	H.240–H.259
Codage des images vidéo animées	H.260–H.279
Aspects liés aux systèmes	H.280–H.299
SYSTÈMES ET ÉQUIPEMENTS TERMINAUX POUR LES SERVICES AUDIOVISUELS	H.300–H.399
SERVICES COMPLÉMENTAIRES EN MULTIMÉDIA	H.450–H.499

Pour plus de détails, voir la Liste des Recommandations de l'UIT-T.

Recommandation UIT-T H.263

Codage vidéo pour communications à faible débit

APPENDICE III

Exemples d'implémentation du codeur/décodeur H.263

Résumé

Le présent Appendice III à la Rec. UIT-T H.263 est informatif. Il contient plusieurs exemples d'application de codeur/décodeur qui sont présentés aux utilisateurs de la Rec. UIT-T H.263 à titre d'information.

Source

L'Appendice III de la Recommandation H.263 de l'UIT-T, élaboré par la Commission d'études 16 (2001-2004) de l'UIT-T, a été approuvé le 8 juin 2001 selon la procédure définie dans la Résolution 1 de l'AMNT.

AVANT-PROPOS

L'UIT (Union internationale des télécommunications) est une institution spécialisée des Nations Unies dans le domaine des télécommunications. L'UIT-T (Secteur de la normalisation des télécommunications) est un organe permanent de l'UIT. Il est chargé de l'étude des questions techniques, d'exploitation et de tarification, et émet à ce sujet des Recommandations en vue de la normalisation des télécommunications à l'échelle mondiale.

L'Assemblée mondiale de normalisation des télécommunications (AMNT), qui se réunit tous les quatre ans, détermine les thèmes d'étude à traiter par les Commissions d'études de l'UIT-T, lesquelles élaborent en retour des Recommandations sur ces thèmes.

L'approbation des Recommandations par les Membres de l'UIT-T s'effectue selon la procédure définie dans la Résolution 1 de l'AMNT.

Dans certains secteurs des technologies de l'information qui correspondent à la sphère de compétence de l'UIT-T, les normes nécessaires se préparent en collaboration avec l'ISO et la CEI.

NOTE

Dans la présente Recommandation, l'expression "Administration" est utilisée pour désigner de façon abrégée aussi bien une administration de télécommunications qu'une exploitation reconnue.

DROITS DE PROPRIÉTÉ INTELLECTUELLE

L'UIT attire l'attention sur la possibilité que l'application ou la mise en œuvre de la présente Recommandation puisse donner lieu à l'utilisation d'un droit de propriété intellectuelle. L'UIT ne prend pas position en ce qui concerne l'existence, la validité ou l'applicabilité des droits de propriété intellectuelle, qu'ils soient revendiqués par un Membre de l'UIT ou par une tierce partie étrangère à la procédure d'élaboration des Recommandations.

A la date d'approbation de la présente Recommandation, l'UIT avait été avisée de l'existence d'une propriété intellectuelle protégée par des brevets à acquérir pour mettre en œuvre la présente Recommandation. Toutefois, comme il ne s'agit peut-être pas de renseignements les plus récents, il est vivement recommandé aux responsables de la mise en œuvre de consulter la base de données des brevets du TSB.

© UIT 2002

Droits de reproduction réservés. Aucune partie de cette publication ne peut être reproduite ni utilisée sous quelque forme que ce soit et par aucun procédé, électronique ou mécanique, y compris la photocopie et les microfilms, sans l'accord écrit de l'UIT.

TABLE DES MATIÈRES

	Page
Appendice III – Exemples d'implémentation du codeur/décodeur H.263	1
III.1 Introduction	1
III.2 Scénarios applicatifs	2
III.2.1 Codage à débit binaire variable pour canaux de transmission sans erreur....	2
III.2.2 Codage à débit binaire fixe pour la transmission par canaux exempts d'erreur (H.320 ou H.324)	3
III.2.3 Codage à débit binaire fixe pour la transmission par canaux en mode paquet exposés aux pertes (H.323).....	4
III.2.4 Codage à débit fixe pour transmission par canaux fortement exposés aux erreurs sur les bits (Annexe C/H.324).....	6
III.3 Algorithmes communs.....	7
III.3.1 Estimation du mouvement.....	7
III.3.2 Quantification.....	11
III.3.3 Autre mode INTER VLC (Annexe S).....	13
III.3.4 Mode de codage INTRA avancé (Annexe I).....	13
III.3.5 Mode de quantification modifié (Annexe T).....	15
III.4 Algorithmes utilisés pour des scénarios applicatifs individuels.....	15
III.4.1 Décision de mode	15
III.4.2 Commande de débit.....	19
III.5 Post-traitement dans le décodeur	25
III.5.1 Post-filtre de soustraction des suroscillations	25
III.5.2 Post-filtre de dégroupage	26
III.5.3 Détection d'erreur	28
III.5.4 Masquage d'erreur	29
III.6 Acquisition des informations.....	30
III.6.1 Mode de codage par trames PB (Annexes G et M).....	30
III.6.2 Mode d'échelonnabilité (Annexe O)	30
III.6.3 Mode de rafraîchissement à résolution réduite (Annexe Q).....	33
III.6.4 Recherche rapide au moyen d'inégalités mathématiques	39
III.6.5 Commande de la fréquence de trame de codage.....	40
III.6.6 Remarques sur l'utilisation optimisée du mode de sélection d'image de référence améliorée	42
III.7 Références	43

Recommandation UIT-T H.263

Codage vidéo pour communications à faible débit

APPENDICE III

Exemples d'implémentation du codeur/décodeur H.263

III.1 Introduction

Le présent appendice décrit le modèle de test à court terme (TMN, *test model near-term*) pour la version 3 de la Rec. UIT-T H.263 (y compris les Annexes A à W) ainsi que l'Annexe X qui a été approuvée ultérieurement. L'objet de ce modèle de test est d'aider les constructeurs, au moyen d'exemples d'implémentation d'un codeur vidéo, à comprendre le mode d'utilisation de la syntaxe et de la spécification du décodeur de la Recommandation.

La Rec. UIT-T H.263 définit la syntaxe du flux binaire et un décodeur vidéo correspondant de façon que les terminaux vidéo issus de constructeurs différents puissent interfonctionner. La conception du codeur vidéo est en dehors du domaine d'application normatif de la Rec. UIT-T H.263 et laissée aux soins du constructeur. Des méthodes de codage préférées ont cependant émergé lors de l'élaboration de la Recommandation. Ces méthodes donnent de bons résultats en termes de qualité vidéo et d'efficacité de compression à des niveaux de complexité appropriés au fonctionnement avec des processeurs à usage général ou particulier et à différents scénarios de transmission (par exemple dans des réseaux en mode commuté ou en mode paquet). Par ailleurs, les niveaux de performance obtenus par ces méthodes servent souvent de critère pour la recherche et le développement de futures Recommandations UIT-T sur le codage vidéo. Le présent appendice décrit les opérations du codeur et du décodeur qui dépassent le texte normatif de la Recommandation H.263, applicables par exemple en cas de violation de syntaxe en raison d'une transmission par des canaux exposés aux erreurs.

Tous les documents cités en référence dans le présent appendice entrent dans l'une des trois catégories suivantes:

- Recommandations de l'UIT: ces Recommandations sont citées en référence au moyen de leur dénomination abrégée, par exemple "H.323" ou "BT.601". Dans certains cas, la version de la Recommandation est significative et est incluse sous une forme appropriée, comme l'année de publication ou le numéro de version. Les Recommandations de l'UIT-T peuvent être obtenues directement auprès de l'UIT-T. Voir <http://www.itu.int/> pour les détails.
- Contributions du groupe d'experts en codage vidéo de la CE 16 de l'UIT-T: ces documents sont cités en référence au moyen de leur désignation abrégée, par exemple [Q15-D-58]. Pour la plupart des contributions citées en référence dans le présent appendice, des hyperliens sont utilisés afin de pointer sur les documents dans un site officiel de téléchargement ftp Q.6/16, dont l'adresse actuelle est la suivante: <ftp://standard.pictel.com/video-site>.
- Autres publications académiques: ces documents sont cités en référence au moyen d'une chaîne d'abréviation fondée surtout sur les noms d'auteur et sur les dates de publication. Ces documents peuvent être trouvés dans la publication citée les concernant.

La conception de la H.263 est fondée sur un codeur de blocs à transformation hybride avec compensation de mouvement, comprenant l'estimation et la compensation du mouvement, la transformation DCT, la quantification, le codage de longueur de séquence, le codage à longueur variable (VLC) ou fixe (FLC). Plusieurs modes de fonctionnement au choix sont par ailleurs autorisés par la H.263 et sont définis par des annexes de cette Recommandation. Le présent appendice part du principe que le lecteur possède une certaine familiarité avec la H.263, avec ses

modes à option et avec le codage vidéo en général. L'on trouvera dans la référence [GFS97] un didacticiel sur la version 1 de la H.263 et sur ses modes à option, tandis qu'une publication correspondante pour la version 2 de la H.263 a été reproduite dans la référence [Q15-D-58].

Le présent appendice décrit les méthodes de codage au moyen d'unités fonctionnelles normalement employées par un codeur. Il donne également des informations sur l'application des diverses méthodes dans des scénarios spécifiques, appelés *scénarios applicatifs*. Il s'agit des codages suivants:

- codage à débit binaire variable avec valeur fixe du quantificateur (voir § 4.2.4/H.263) pour transmission par canaux exempts d'erreur (couramment utilisé aux fins de la recherche en codage vidéo);
- codage à débit binaire fixe dans un environnement pratiquement exempt d'erreur (H.320/H.324);
- codage à débit binaire fixe dans un environnement de réseau en mode paquet exposé aux pertes (H.323);
- codage à débit binaire fixe dans un environnement exposé aux erreurs (Annexe C/H.324).

Les scénarios applicatifs sont examinés au § III.2, qui fait référence aux mécanismes décrits dans les paragraphes ultérieurs. Le § III.2 définit également les environnements de simulation pris en hypothèse dans le présent appendice. Les mécanismes communs à tous les scénarios applicatifs sont décrits dans le § III.3. Ils comprennent la recherche de vecteur cinétique (MV, *motion vector*), la quantification et l'utilisation des modes facultatifs H.263 qui sont communs à tous les scénarios applicatifs. Les mécanismes applicables à des scénarios applicatifs spécifiques sont examinés au § III.4. Combinés avec la Recommandation proprement dite, ces paragraphes définissent un cadre pour un codec H.263 qui fonctionne assez bien dans les divers scénarios applicatifs. Le § III.5 analyse le post-traitement dans le décodeur. Le § III.6 traite de l'acquisition de toutes les informations jugées utiles et adoptées pour inclusion dans le présent appendice mais qui ne s'intègrent pas encore dans un des scénarios applicatifs.

III.2 Scénarios applicatifs

Le présent paragraphe examine les scénarios applicatifs envisagés pour le codage vidéo en mode H.263. Son objet est de donner un aperçu général des mécanismes utiles pour chaque scénario et de décrire, le cas échéant, l'environnement de simulation correspondant. La description de l'environnement extérieur au conduit de données vidéo est limitée à un minimum, bien que des références soient fournies pour commodité.

III.2.1 Codage à débit binaire variable pour canaux de transmission sans erreur

Le scénario de codage à débit binaire variable fait appel à une valeur constante du quantificateur afin d'obtenir un niveau de qualité constant pour toutes les images et pour toutes les régions d'image. Ce scénario est utile pour la recherche en codage vidéo et pour les travaux de normalisation associés. Il offre par exemple un cadre d'évaluation de la qualité objective et subjective des flux binaires produits par de nouvelles propositions relatives à l'efficacité de codage. Il existe deux modèles différents pour ce scénario selon que la complexité de celui-ci est faible ou élevée. Ces modèles sont décrits dans les deux paragraphes suivants.

III.2.1.1 Modèle de codage à faible complexité

Le modèle de codage à faible complexité fait appel aux modes facultatifs qui sont définis dans l'Annexe X/H.263, c'est-à-dire le mode de codage INTRA avancé (Annexe I), le mode avec filtre de dégroupage (Annexe J), le mode de quantification modifié (Annexe T) et une partie du mode d'information d'amélioration complémentaire (Annexe L). Noter qu'aucune des capacités améliorées du mode d'information d'amélioration complémentaire n'est actuellement examinée dans le présent appendice. Il convient toutefois de noter qu'elles visent des caractéristiques qui pourraient être

importantes pour certains modèles de produit. L'estimation du mouvement et la décision relative au mode sont effectuées au moyen des méthodes à faible complexité de recherche de vecteur cinétique et de décision relative au mode. L'on pourra trouver des détails au § III.3.1.2 (recherche vectorielle cinétique à faible complexité), au § III.4.1.2 (décision de mode à faible complexité), au § III.3.4 (mode de codage INTRA avancé) et au § III.3.5 (mode de quantification modifié).

III.2.1.2 Modèle de codage de qualité élevée

Le modèle de codage de qualité élevée est conçu de façon à offrir une meilleure qualité d'image reconstituée par rapport au modèle à faible complexité, au prix d'une plus grande complexité des calculs. Le modèle de codage de qualité élevée fait appel aux modes facultatifs qui sont définis dans le profil 5 de l'Annexe X/H.263, laquelle définit les modes du profil 1 utilisé pour le modèle de codage à faible complexité ainsi que le mode de codage par vecteurs cinétiques non restreints (Annexe D), le mode de prédiction avancé (Annexe F) et le mode de sélection d'image de référence améliorée (Annexe U) dans le cas où plusieurs images de référence sont utilisées pour la compensation du mouvement. L'estimation du mouvement et la décision relative au mode sont effectuées au moyen des méthodes à faible complexité de recherche de vecteur cinétique et de décision relative au mode. L'on pourra trouver des détails au § III.3.1.4 (recherche vectorielle cinétique à codage de qualité élevée), au § III.4.1.3 (décision de mode à codage de qualité élevée), au § III.3.4 (mode de codage INTRA avancé), au § III.3.5 (mode de quantification modifié) et au § III.4.2.4 (décision de mode combinée, recherche vectorielle cinétique et sélection d'image de référence améliorée).

III.2.2 Codage à débit binaire fixe pour la transmission par canaux exempts d'erreur (H.320 ou H.324)

Cette application est caractérisée par la nécessité d'obtenir un débit binaire dont la valeur cible est fixe avec un retard raisonnablement faible. Le mécanisme de transport est en mode binaire et offre un environnement qui peut être considéré comme exempt d'erreur dans toutes les situations concrètes. Ce scénario est, par rapport au précédent, plus proche d'une application pratique, de sorte que le compromis complexité/qualité est un point important à résoudre. Afin d'obtenir un débit cible, l'échelon du quantificateur n'est plus fixe mais déterminé par l'algorithme de commande de débit. Par ailleurs, bien qu'une fréquence de trame cible soit généralement spécifiée, l'algorithme de commande de débit possède la possibilité d'abandonner des images individuelles de source lors d'un dépassement du budget de débit binaire.

III.2.2.1 Modèle de codage à faible complexité

Le modèle à faible complexité inclut tous les mécanismes décrits au § III.2.1.1, plus la commande de débit afin d'obtenir le débit binaire cible. L'on pourra trouver des détails au § III.3.1.2 (recherche vectorielle cinétique à faible complexité), au § III.4.1.2 (décision de mode à faible complexité), au § III.3.4 (mode de codage INTRA avancé), au § III.3.5 (mode de quantification modifié) et au § III.4.2 (commande de débit).

III.2.2.2 Modèle de codage de qualité élevée

Le modèle de qualité élevée inclut tous les mécanismes décrits au § III.2.1.2, plus la commande de débit afin d'obtenir le débit binaire cible. Dans le présent appendice, quelques simplifications sont appliquées à l'utilisation de la commande de débit lorsqu'elle est combinée aux algorithmes d'estimation du mouvement et de décision de mode de qualité élevée, comme décrit au § III.4.2.3.

III.2.3 Codage à débit binaire fixe pour la transmission par canaux en mode paquet exposés aux pertes (H.323)

Ce scénario applicatif est caractérisé par la nécessité d'obtenir un débit cible fixe et par une méthode d'empaquetage pour le transport. Les systèmes H.323 font appel à un transport en mode paquet sur base RTP [RFC 1889]. Les principales caractéristiques d'un tel transport sont les suivantes:

- longueur de paquet modulable par l'expéditeur au voisinage de 1,5 kO, afin de réduire le surdébit d'empaquetage et de respecter la longueur des unités de transfert maximales (MTU, *maximum transfer unit*) en protocole IP;
- principaux débits de perte de paquets.

Noter que les valeurs moyennes des débits de perte de paquets à mi-parcours sont disponibles dans le codeur à partir de comptes rendus obligatoires du récepteur RTCP, faisant partie du protocole RTP. Etant donné que celui-ci, ainsi que les couches protocolaires sous-jacentes qui lui sont associées, garantit que les paquets sont acheminés dans la séquence correcte¹ et qu'ils sont exempts d'erreurs sur les bits, la seule source d'erreur est la perte de paquets.

Il est notoire que la communication dans les deux sens faisant appel à des messages par voie de retour peut améliorer grandement la qualité d'image reproduite. La communication dans les deux sens n'est cependant pas souvent possible à cause de la nature éventuellement multidiffusée du transport et de l'application ainsi qu'en raison des éventuelles contraintes de l'application en terme de délai de transmission. La complexité de la simulation d'un environnement réellement bilatéral n'est donc pas triviale. C'est pourquoi seule la communication unilatérale (dans un seul sens) sera prise en considération ci-après.

Premièrement, un algorithme de commande de débit fixe est employé afin de simplifier l'interaction entre mécanismes de commande de débit source et mécanismes de transport. Dans les systèmes réels, bien que la gestion par tampons du RTP puisse lisser les variations à court terme des débits de transmission, le débit cible et la capacité tampon du récepteur font normalement l'objet d'un réglage périodique sur la base de facteurs tels que le débit moyen de perte de paquets et les contraintes de facturation. En deuxième lieu, l'on suppose que les débits moyens de perte de paquets à court terme sont constants, afin de simplifier l'interaction des mécanismes de tolérance aux erreurs et de transport. Dans les systèmes réels, la prise en charge de la tolérance aux erreurs devrait être adaptative, par exemple en fonction de débits moyens de perte de paquets fenêtrés.

Le reste de ce paragraphe examine les questions de mise en et hors paquets ainsi que l'application d'outils de codage vidéo pour ce scénario. Il y a deux modèles différents, correspondant à une complexité faible et à une qualité élevée.

III.2.3.1 Empaquetage et dépaquetage

Le présent paragraphe décrit un système d'empaquetage et dépaquetage faisant appel au format de charge utile RFC 2429 [RFC 2429] pour le protocole RTP, qui a fait la preuve de son bon fonctionnement dans l'environnement pris comme hypothèse dans ce scénario.

L'on part du principe de l'utilisation d'un mécanisme de codage pour réduire la propagation des erreurs temporelles qui sont inévitables dans un environnement à pertes de paquets. Un tel mécanisme, décrit au § III.4.1.1, est une utilisation judicieuse des macroblocs à codage INTRA. L'utilisation du mode structuré par tranches (Annexe K), où les tranches sont de même longueur et de même forme qu'un groupe GOB, est également prise comme hypothèse. Dans ce scénario, ce sont des tranches qui sont utilisées et non des groupes GOB parce que l'Annexe K autorise une structure

¹ Bien que le protocole RTP ne possède pas de fonction garantissant une numérotation séquentielle correcte, son en-tête comporte un numéro de séquence qui peut être utilisé pour assurer un séquençement correct des paquets.

de flux binaire dans laquelle les tranches n'ont pas besoin d'apparaître dans l'ordre de balayage normal. Le système d'empaquetage dépend d'un tel dispositif.

Le système d'empaquetage est fondé sur l'entrelacement de tranches à numérotage pair ou impair, mises en forme de groupes GOB. Il fait suite à deux considérations théoriques. Tout d'abord, étant donné que le surdébit d'empaquetage est d'environ 40 octets par paquet pour les en-têtes IP/UDP/RTP dans l'Internet, il faut utiliser des paquets de longueur suffisante. En deuxième lieu, il convient de ne pas placer des tranches consécutives dans un même paquet afin de permettre un masquage d'erreur effectif, comme décrit au § III.5.4. Ce mécanisme donne deux paquets par image et permet un masquage suffisant des macroblocs manquants si un seul des deux paquets est perdu. Cette méthode peut être étendue à plus de deux paquets par image si la longueur de l'image codée est supérieure à 2800 octets, étant entendu que la longueur maximale de la charge utile est de 1400 octets par paquet et que les unités MTU ont une longueur de 1500 octets.

La perte du contenu de l'en-tête d'image peut dégrader notablement la qualité de l'image décodée. Il est possible d'y remédier en admettant que le contenu de l'en-tête d'image reste inchangé sauf le champ de référence temporelle, qui peut être reconstitué d'après la référence temporelle incluse dans l'en-tête RTP. Lorsque le codeur vidéo modifie le contenu de l'en-tête d'image, l'on fait appel à un mécanisme RFC 2429 permettant d'inclure des copies redondantes de l'en-tête d'image dans l'en-tête de charge utile de chaque paquet afin qu'il soit possible d'effectuer un décodage partiel d'une image (et son masquage ultérieur) dès réception d'un paquet donné.

Un exemple représentatif de ce système d'empaquetage est présenté dans la Figure III.1 pour un flux vidéo codé à 50 kbit/s et à 10 trames par seconde en résolution QCIF. Le surdébit constant d'empaquetage (composé des en-têtes IP/UDP/RTP soit 40 octets au total par paquet) est de 80 octets par image empaquetée. Dans la Figure, le terme GOB désigne une tranche mise en forme de groupe GOB dont la position spatiale correspond au groupe GOB portant le numéro associé.

Paquet 1: contient l'en-tête d'image original et les groupes GOB ayant un numéro impair.
Longueur totale: 371 octets

Paquet 2: contient une copie redondante de l'en-tête d'image et des groupes GOB ayant un numéro pair.
Longueur totale: 303 octets

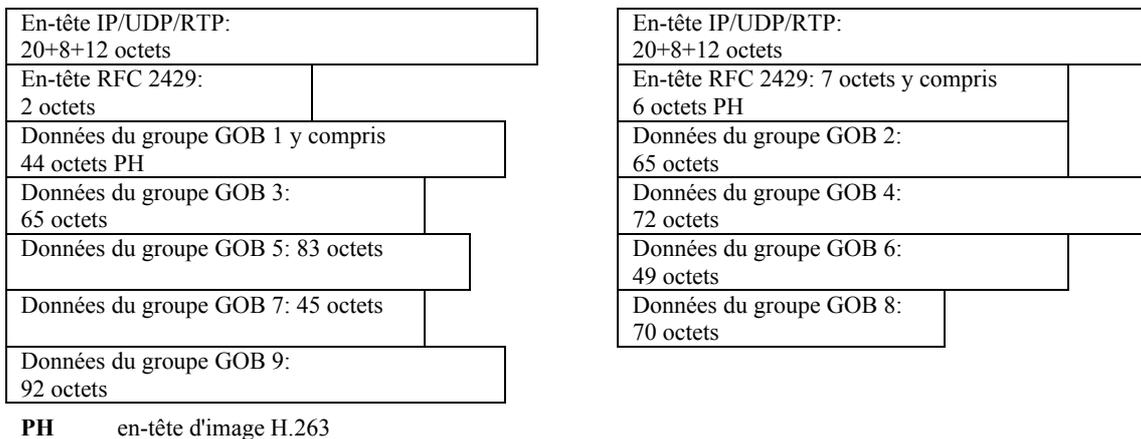


Figure III.1/H.263 – Exemple de deux paquets d'image utilisant le système d'empaquetage entrelacé

L'en-tête minimal de 2 octets défini dans le commentaire RFC 2429 ne contribue pas au surdébit car il remplace les 16 bits du code de déclenchement d'image ou de départ de tranche qui traite chaque image/tranche dans la Rec. UIT-T H.263.

Quatre situations différentes peuvent se produire dans le système d'empaquetage ci-dessus, au cours du dépaquetage et de la reconstruction dans un environnement à pertes de paquets. Si les deux

paquets sont reçus, le décodage est direct. Si seul le premier paquet est reçu, les données disponibles sont décodées et les tranches manquantes sont détectées et masquées, par exemple comme décrit au § III.5.4. Si seul le second paquet est disponible, l'en-tête d'image redondant et les données de charge utile sont concaténés et décodés. Les tranches manquantes sont détectées car elles provoquent des violations de syntaxe. Puis elles sont masquées comme décrit au § III.5.4. Si les deux paquets sont perdus, aucune donnée n'est disponible pour le décodeur et le réaffichage de l'image précédente doit masquer l'image entière.

III.2.3.2 Modèle de codage à faible complexité

Le modèle de codage à faible complexité utilise les mécanismes de codeur décrits au § III.2.2.1 avec deux compléments qui améliorent la tolérance aux erreurs. Des tranches en forme de groupes GOB sont utilisées, l'en-tête de chaque tranche étant inséré au début de celle-ci et servant de marqueur de synchronisation ainsi que de base de réinitialisation du codage spatial prédictif, comme le codage des vecteurs cinétiques et le codage intercoefficient, comme défini dans l'Annexe K/H.263. De même, le débit auquel les macroblocs sont forcés au codage en mode INTRA (rafraîchi) varie en fonction de l'inverse du débit moyen de perte de paquets à mi-parcours. Cette opération est effectuée au moyen du paramètre INTRA_MB_Refresh_Rate du mécanisme de rafraîchissement des macroblocs INTRA, comme décrit au § III.4.1.1.

III.2.3.3 Modèle de codage de qualité élevée

Le modèle de codage de qualité élevée utilise l'empaquetage, le masquage d'erreur et les tranches de forme GOB comme décrit ci-dessus. Il apporte également une extension du mode de qualité élevée décrit au § III.2.1.2, en ce sens que la décision de mode de codage de qualité élevée du § III.4.1.3 est remplacée par celle du § III.4.1.5. Cette décision de mode est motivée par la nature exposée aux erreurs du réseau, par le processus d'empaquetage, par le débit de perte de paquets et par le masquage d'erreur employés par le décodeur pour sélectionner le mode de codage des macroblocs. Des détails peuvent être consultés dans le § III.4.1.5 (décision de mode de qualité élevée pour environnements exposés aux erreurs).

III.2.4 Codage à débit fixe pour transmission par canaux fortement exposés aux erreurs sur les bits (Annexe C/H.324)

Ce scénario applicatif est caractérisé par la nécessité d'obtenir un débit cible fixe dans un environnement fortement exposé aux erreurs sur les bits. Dans de tels environnements H.223, le protocole de transport utilisé par le système Annexe C/H.324 est essentiellement optimisé pour le fonctionnement à bas débit. Le mode H.223 ne peut pas assurer l'acheminement sans erreur de la charge utile, même lorsque les algorithmes facultatifs de réexpédition sont utilisés. Le décodeur vidéo doit donc avoir la capacité de détecter et de traiter les erreurs sur les bits.

Pour des raisons pratiques, plusieurs hypothèses et simplifications sont nécessaires concernant la simulation du transport:

- le mode à trames verrouillées du mode H.223 est employé avec des unités SDU AL3 pour données H.263 (ce qui permet au contrôle CRC 16 bits de détecter les erreurs);
- l'hypothèse d'une communication unilatérale est retenue, c'est-à-dire qu'aucun algorithme de réexpédition ou qu'aucun mécanisme de voie de retour n'est utilisé, en raison de la sévérité des contraintes de temps de propagation.

Deux modèles sont définis: un modèle de codage à faible complexité qui n'est fondé que sur les outils de codage disponibles dans la version 2 (1998) de la H.263, et un modèle de codage à haute tolérance aux erreurs comportant les outils disponibles dans la version 3 de la Rec. UIT-T H.263.

III.2.4.1 Modèle de codage à faible complexité

Le modèle à faible complexité utilise les outils de codage vidéo décrits au § III.2.2.1 pour le modèle à faible complexité et débit fixe en environnement sans erreur. Afin d'améliorer la tolérance aux

erreurs, une méthode d'empaquetage et une méthode de codage INTRA forcé sont également incluses. Chaque groupe GOB est codé avec un en-tête GOB qui sert de marqueur de synchronisation et d'élément de réinitialisation du codage spatial prédictif, comme le codage des vecteurs cinétiques et le codage intercoefficient définis dans la Rec. UIT-T H.263. Chaque groupe GOB codé est empaqueté dans une unité SDU AL3. Toutes les unités SDU AL3 reçues avec CRC défectueux ne sont pas traitées par le décodeur mais sont masquées. Aucun essai n'est tenté pour décoder partiellement les groupes GOB corrompus en raison de la difficulté que représente la définition exacte du fonctionnement d'un décodeur dans un cas semblable. Par ailleurs, ce problème se complique du fait que les erreurs qui ne sont pas détectées peuvent se traduire par des motifs très visibles comme des blocs de couleurs différentes. La seule violation de syntaxe qui est autorisée, et utilisée pour détecter les groupes GOB manquants dans le décodeur, est la numérotation hors séquence d'un groupe GOB.

Une fréquence de rafraîchissement fixe des macroblocs INTRA est utilisée, dont la détermination suit la règle élémentaire suivante. A chaque $1/p$ ième instant (arrondi à la valeur d'entier la plus proche) où un macrobloc contenant des coefficients est codé, ce macrobloc est codé en mode INTRA, p étant la probabilité moyenne de perte pour tous les macroblocs d'une séquence et pour une caractéristique d'erreur donnée. En d'autres termes, si la probabilité de perte déterminée est de 0,1, un macrobloc contenant des informations de coefficient sera codé en mode INTRA tous les 10^e instants. Cet algorithme doit être implémenté au moyen du paramètre INTRA_MB_Refresh_Rate du mécanisme de rafraîchissement de macrobloc INTRA, comme décrit au § III.4.1.1. Contrairement à l'algorithme recommandé au § 5.1.4.3 de la version 2 H.263, le fanion indiquant le type d'arrondissement de la structure PLUSTYPE est mis à "0", quel que soit le type d'image [Q15-I-26].

III.2.4.2 Modèle de codage à tolérance d'erreur élevée

Le modèle de codage à tolérance d'erreur élevée se fonde sur le mode de tranche à subdivision des données comme indiqué au § III.4.2.4. Par ailleurs, le mécanisme de "répétition de l'en-tête d'image précédent" de l'Annexe W est utilisé comme indiqué au § III.4.2.6.

III.3 Algorithmes communs

Les algorithmes communs à tous les scénarios applicatifs seront examinés dans le présent paragraphe.

III.3.1 Estimation du mouvement

Dans le présent appendice, trois algorithmes d'estimation du mouvement sont décrits:

- un algorithme de recherche rapide à faible complexité, fondé sur un nombre réduit d'emplacements de recherche;
- un algorithme de recherche complète, de complexité moyenne;
- un algorithme de recherche complète, de complexité élevée, qui tient compte du débit binaire des vecteurs cinétiques résultats, en plus de la qualité de la concordance.

De ces algorithmes, ceux de qualité faible et élevée sont les plus fréquemment utilisés pour les modèles correspondants de scénarios applicatifs à qualité faible et élevée. L'algorithme de qualité moyenne peut remplacer l'algorithme de qualité faible et fournit une qualité légèrement supérieure en terme de distorsion du débit au prix d'une complexité de calcul un peu plus élevée. Noter que tout algorithme de recherche par blocs peut être employé. Les trois algorithmes décrits dans ce modèle d'essai sont cependant couramment utilisés et sont reconnus comme donnant de bons résultats pour les divers modèles de scénario applicatif.

Le protocole H.263 peut utiliser un ou quatre vecteurs cinétiques par macrobloc, selon les modes facultatifs qui sont activés. L'on trouvera dans les Annexes D, F et J de la Rec. UIT-T H.263 une

description des diverses limitations des vecteurs cinétiques. De même, l'étendue possible du domaine de recherche dépend du sous-mode de l'Annexe D qui est employé.

III.3.1.1 Mesure de distorsion par la somme des différences absolues

Les deux algorithmes de recherche de vecteurs cinétiques (à pixels entiers et à demi-pixels) utilisent la somme des différences absolues (SAD, *sum of absolute difference*) en tant que mesure de distorsion. L'on calcule la somme SAD entre tous les pixels de luminance des macroblocs candidats et des macroblocs cibles. Dans certains cas, l'algorithme de décision de mode favorise la somme SAD pour le vecteur (0,0). La somme SAD pour un vecteur cinétique candidat est calculée comme suit:

$$SAD(u, v) = \sum_{j=0}^{M-1} \sum_{i=0}^{M-1} \left| \tilde{X}_{t-1}(i+u, j+v) \right|$$

où:

$$M = \begin{cases} 8 & \text{pour } 8 \times 8 \text{ vecteurs cinétiques} \\ 16 & \text{pour } 16 \times 16 \text{ vecteurs cinétiques} \end{cases}$$

X_t trame cible

\tilde{X}_{t-1} précédente trame reconstruite

(i, j) emplacement spatial dans la trame cible

(u, v) vecteur cinétique candidat

Afin de réduire la complexité, le calcul de somme SAD fait appel à la technique de distorsion partielle, qui compare la valeur cumulée de la somme SAD après chaque rangée de M pixels à la somme SAD minimale déjà calculée dans la fenêtre de recherche. Si la SAD cumulée dépasse la SAD minimale déjà constatée, le calcul du vecteur cinétique candidat est arrêté car ce vecteur ne fournira pas de meilleure correspondance, en termes de valeur SAD la plus basse, que la meilleure correspondance déjà trouvée.

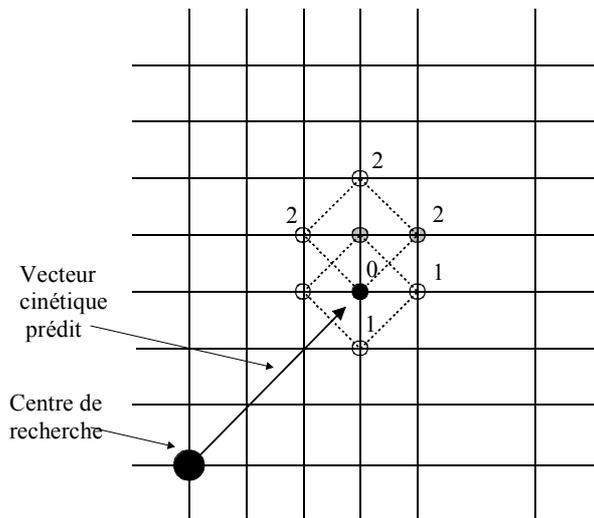
III.3.1.2 Recherche de faible complexité

La recherche de faible complexité a été retenue parce qu'une comparaison approfondie avec l'algorithme de recherche complète a montré une dégradation de performance négligeable ou nulle pour les images en résolution QCIF ou CIF. Les détails sont disponibles dans la référence [GCK99] et des résultats détaillés, comparant la qualité par rapport à l'algorithme de recherche complète, sont disponibles dans la référence [Q15-B-23].

III.3.1.2.1 Procédure de recherche pour blocs 16×16 et précision à pixels entiers

Le centre de la recherche est le vecteur cinétique médian qui est prédit comme défini au § 6.1.1/H.263 et à l'Annexe F.2/H.263. Le vecteur (0,0), s'il diffère du vecteur cinétique prédit, est également recherché. L'on trouve le vecteur (0,0) en retranchant 100 de la somme SAD calculée. L'étendue de recherche admissible est déterminée par le sous-mode de l'Annexe D qui est employé.

L'algorithme procède à une recherche séquentielle des couches en losange, dont chacune contient les quatre voisins immédiats du centre de recherche actuel. La couche $i + 1$ est ensuite centrée sur le point de somme SAD minimale de la couche i . Les couches successives ont donc des centres différents et contiennent au plus trois vecteurs cinétiques candidats qui n'ont pas été vérifiés, sauf pour la première couche autour du vecteur cinétique prédit, qui contient quatre vecteurs cinétiques candidats non vérifiés. La Figure III.2 donne un exemple de cette recherche.



T1609820-01

- 0 Emplacement de recherche (du vecteur cinétique prédit) dans la couche 0
- 1 Emplacement de recherche dans la couche 1
- 2 Emplacement de recherche dans la couche 2

NOTE – Chaque point ombré représente la somme SAD minimale de la couche correspondante

Figure III.2/H.263 – Exemple de recherche rapide à pixels entiers sur deux couches à partir du vecteur cinétique prédit

La recherche n'est arrêtée que si l'une des deux conditions suivantes est satisfaite:

- 1) tous les vecteurs cinétiques candidats de la couche actuelle ont été vérifiés et la valeur minimale de somme SAD de la couche actuelle est supérieure à celle de la couche précédente;
- 2) la recherche a atteint la borne de la région de recherche admissible et tente de la franchir.

Le vecteur cinétique (MV) d'une région à pixels entiers sur 16×16 est ensuite raffiné à demi-pixels comme décrit ci-après.

III.3.1.2.2 Raffinement à demi-pixels dans une région 16×16

La région de recherche à demi-pixels est de ± 1 demi-pixel autour du meilleur vecteur candidat d'une région 16×16 . L'on effectue cette recherche en calculant les valeurs des demi-pixels comme décrit au § 6.1.2/H.263 puis en calculant la somme SAD pour chaque vecteur de demi-pixels possible. Huit vecteurs candidats supplémentaires sont donc vérifiés. L'on retient le vecteur produisant la meilleure correspondance au cours du raffinement à demi-pixels.

III.3.1.2.3 Recherche à pixels entiers dans une région 8×8

Les vecteurs cinétiques dans une région de 8×8 pixels entiers sont transformés en meilleur vecteur cinétique à pixels entiers dans une région 16×16 . Il n'y a pas de recherche additionnelle à pixels entiers dans une région 8×8 . Compte tenu du codage différentiel des vecteurs cinétiques, cette restriction garantit que les vecteurs cinétiques 8×8 choisis pour le macrobloc cible ont un coefficient de corrélation élevé.

III.3.1.2.4 Raffinement à demi-pixels dans une région 8×8

Le raffinement à demi-pixels est effectué pour chacun des blocs entourant le vecteur à pixels entiers d'une région 8×8 . La recherche est pratiquement la même qu'au § III.3.1.2.2, avec des blocs de dimension égale à 8. Noter que les vecteurs (0,0) ne sont pas recherchés ici.

III.3.1.3 Recherche complète de qualité moyenne

III.3.1.3.1 Recherche à pixels entiers dans une région 16×16

Le centre de la recherche est nouveau le vecteur cinétique médian qui est prédit comme défini au § 6.1.1/H.263 et à l'Annexe F.2/H.263. Le vecteur (0,0), s'il diffère du vecteur cinétique prédit, est également recherché. L'on trouve le vecteur (0,0) en retranchant 100 de la somme SAD calculée. L'étendue de recherche admissible est déterminée par le sous-mode de l'Annexe D qui est employé.

L'algorithme procède à une recherche séquentielle des couches selon un motif centrifuge en spirale, par rapport au vecteur cinétique prédit, jusqu'aux limites absolues de l'étendue de recherche autorisée, pour la recherche à pixels entiers aussi bien en région 16×16 qu'en région 8×8 . Chaque couche sauf celle qui ne contient que l'origine de la recherche, ajoute donc $\delta \cdot \text{nombre_de_couches}$ vecteurs cinétiques candidats. Dans chaque couche, la recherche commence au coin supérieur gauche de la spirale et continue dans le sens horaire.

Cette recherche en spirale est plus efficace si elle commence par le vecteur cinétique médian qui a été prédit, car celui-ci produira habituellement une bonne correspondance dès le début de la recherche ce qui, en association avec la technique de distorsion partielle, pourra réduire notablement le nombre de sommes SAD à calculer dans chaque rangée. Le raffinement à demi-pixels est identique à l'algorithme de faible complexité, comme décrit au § III.3.1.2.2.

III.3.1.3.2 Recherche à pixels entiers dans une région 8×8

Une région de 8×8 pixels entiers par recherche analyse entièrement une fenêtre d'exploration réduite, à proximité du meilleur vecteur cinétique d'une région à pixels entiers sur 16×16 , de dimension égale à ± 2 unités de pixel entier. Cela permet une variation légèrement plus grande des vecteurs cinétiques 8×8 pour le macrobloc cible. Mais ces vecteurs peuvent alors fournir des correspondances de meilleure qualité. Le raffinement à demi-pixels est identique à l'algorithme de faible complexité, comme décrit au § III.3.1.2.4.

III.3.1.4 Recherche complète de qualité élevée optimisée par distorsion de débit

Afin de régulariser le problème de l'estimation du mouvement, l'on fait appel à une formulation de Lagrange dans laquelle la distorsion est pondérée en fonction du débit au moyen d'un multiplicateur de Lagrange. Si un mode prenant en charge les vecteurs cinétiques 8×8 est activé, par exemple selon l'Annexe F et/ou l'Annexe J, une recherche complète est employée pour les deux vecteurs cinétiques à pixels entiers: 16×16 et 8×8 . Sinon, une recherche complète n'est employée que pour les vecteurs cinétiques à pixels entiers 16×16 .

Pour chaque macrobloc 16×16 ou chaque bloc 8×8 , la recherche de vecteur cinétique à pixels entiers fait appel à un motif d'exploration en spirale, comme décrit au § III.3.1.3.1. Le raffinement à demi-pixels utilise le motif décrit au § III.3.1.2.2. Le vecteur cinétique à demi-pixels pour le bloc 8×8 i s'obtient avant la recherche des vecteurs cinétiques à demi-pixels ou à pixels entiers pour le bloc 8×8 $i + 1$, etc. Cela permet un calcul précis des termes de débit pour les vecteurs cinétiques à pixels entiers ou à demi-pixels pour les blocs 8×8 , car le vecteur cinétique prédit sera complètement connu.

La recherche à pixels entiers optimisée par distorsion de débit sélectionne le vecteur cinétique qui minimise la fonction de potentiel cinétique de Lagrange définie par:

$$J = D + \lambda_{motion} R$$

La distorsion, D , est définie comme étant la somme SAD calculée entre la composante de luminance du macrobloc (ou bloc) cible et le macrobloc (bloc) contenu dans l'image de référence déplacée par le vecteur cinétique candidat. Les calculs de somme SAD font appel aux critères de correspondance par la technique de distorsion partielle en fonction de la somme SAD minimale trouvée jusque là. Le débit, R , est défini comme la somme des débits relatifs aux vecteurs cinétiques candidats verticaux et

horizontaux de macrobloc (ou de bloc) compte tenu du vecteur cinétique prédit qui est défini au § 6.1.1/H.263. Le paramètre λ est choisi comme indiqué ci-dessous. La recherche est centrée sur le vecteur cinétique prédit. Le vecteur (0,0) est recherché mais non privilégié car la minimisation lagrangienne tient déjà compte du débit des vecteurs cinétiques. Les meilleurs candidats 16×16 et 8×8 sont ensuite raffinés à demi-pixels au moyen d'une recherche de ± 1 demi-pixel. La recherche à demi-pixels emploie également la formulation lagrangienne indiquée ci-dessus afin de choisir le vecteur cinétique à demi-pixels qui minimise la fonction de potentiel lagrangienne. L'algorithme de décision de mode utilise ensuite les meilleurs vecteurs 16×16 et 8×8 .

Le choix du paramètre λ_{motion} a une influence assez faible sur le résultat de l'estimation cinétique d'un bloc 16×16 . Mais le résultat de recherche pour les blocs 8×8 est fortement influencé par λ_{motion} , qui est choisi comme suit:

$$\lambda_{motion} = 0,92 \cdot QP$$

où QP est le paramètre de quantification de macrobloc [Q15-D-13]. Si l'application concerne un débit binaire fixe, la relation est modifiée comme décrit au § III.4.2.3.

III.3.1.5 Autres problèmes posés par la recherche de vecteurs cinétiques

En mode d'actualisation de résolution réduite et pour les divers types d'image autorisés avec le mode d'échelonnabilité défini dans l'Annexe O, le modèle d'essai comprend des algorithmes de recherche de vecteur cinétique pour trames PB (si l'Annexe G ou M est activée) qui cependant ne sont utilisés dans aucun scénario applicatif mais sont examinés plus en détail aux § III.6.1.1, III.6.3 et III.6.2.2 respectivement.

III.3.2 Quantification

Les paramètre de quantification $QUANT$ peut avoir une valeur d'entier comprise entre 1 et 31. L'intervalle de reconstruction par quantification pour coefficients non nuls est égal à $2 \cdot QP$

où:

$QP = 4$ pour le coefficient DC du codage INTRA lorsque le mode de codage INTRA avancé n'est pas sélectionné;

$QP = QUANT$ dans le cas contraire.

Si l'on définit les paramètres suivants:

COF coefficient de transformée (ou différence entre coefficients) à quantifier,

$LEVEL$ version quantifiée du coefficient de transformée,

REC valeur de coefficient reconstruite,

"/" division par troncature.

La règle de reconstruction de base par quantification inverse peut être exprimée comme suit pour tous les coefficients quantifiés différents de zéro:

$$|REC| = QP \cdot (2 \cdot |LEVEL| + p) \quad \text{si } QP = \text{"impair"};$$

$$|REC| = QP \cdot (2 \cdot |LEVEL| + p) - p \quad \text{si } QP = \text{"pair"},$$

où:

$p = 1$ pour les coefficients INTER;

$p = 1$ pour les coefficients INTRA non DC hors mode de codage INTRA avancé;

$p = 0$ pour le coefficient INTRA DC hors mode de codage INTRA avancé;

$p = 0$ pour les coefficients INTRA (DC et non DC) en mode de codage INTRA avancé.

Le paramètre p est égal à l'unité lorsque l'espacement de valeur de reconstruction n'est pas uniforme (c'est-à-dire lorsqu'il y a une expansion de l'espacement de reconstruction autour de zéro) et est égal à zéro dans le cas contraire. La règle de quantification du codeur à appliquer est compensée pour tenir compte de l'influence de p sur l'espacement de reconstruction. Afin que la quantification ait une variance optimale, il convient que les seuils de décision de quantification soient espacés de façon que les valeurs de reconstruction forment, pour chaque région, le centre de gravité des valeurs escomptées. Si la fonction de densité de probabilité (pdf, *probability density function*) des coefficients est modélisée par la distribution laplacienne, un simple décalage qui est le même pour chaque intervalle de quantification peut donner cet espacement optimal. Les coefficients sont quantifiés conformément à une telle règle, c'est-à-dire qu'ils utilisent une forme "réduite aux entiers" de la relation:

$$|LEVEL| = \left[|COF| + (f - p) \cdot QP \right] / (2 \cdot QP)$$

où:

$f \in \left\{ \frac{1}{2}, \frac{3}{4}, 1 \right\}$ est un paramètre utilisé pour localiser les seuils de décision du quantificateur de façon que chaque valeur de reconstruction se trouve quelque part entre une opération d'arrondissement au plus proche entier supérieur ($f = 1$) et une opération de reconstruction par le bord gauche ($f = 0$), f étant choisi de façon à correspondre au taux moyen (exponentiel) de décroissance de la pdf de la source à chaque échelon différent de zéro.

III.3.2.1 Quantification de coefficients INTER

Les coefficients INTER (DC ou non DC) sont quantifiés selon la relation suivante:

$$|LEVEL| = \left(|COF| - QUANT / 2 \right) / (2 \cdot QUANT)$$

Ce qui correspond à $f = \frac{1}{2}$ avec $p = 1$.

III.3.2.2 Quantification de coefficients INTRA non DC hors mode de codage INTRA avancé

Les coefficients INTRA non DC hors mode de codage INTRA avancé sont quantifiés selon la relation suivante:

$$|LEVEL| = |COF| / (2 \cdot QUANT)$$

Ce qui correspond à $f = 1$ avec $p = 1$.

III.3.2.3 Quantification du coefficient INTRA DC hors mode de codage INTRA avancé

Le coefficient DC d'un bloc INTRA hors mode de codage INTRA avancé est quantifié selon la relation suivante:

$$LEVEL = (COF + 4) / (2 \cdot 4)$$

Ce qui correspond à $f = 1$ avec $p = 0$. Noter que COF et $LEVEL$ sont toujours non négatifs et que QP est toujours égal à 4 dans ce cas.

III.3.2.4 Quantification de coefficients INTRA en mode de codage INTRA avancé

Les coefficients INTRA en mode de codage INTRA avancé (DC et non DC) sont quantifiés selon la relation suivante:

$$|LEVEL| = \left(|COF| + 3 \cdot QUANT / 4 \right) / (2 \cdot QUANT)$$

Ce qui correspond à $f = \frac{3}{4}$ avec $p = 0$.

III.3.3 Autre mode INTER VLC (Annexe S)

Au cours du codage entropique, le codeur utilise la table de codage INTRA VLC de l'Annexe I afin de coder un bloc INTER si les trois critères suivants sont satisfaits:

- l'autre mode INTER VLC (Annexe S) est utilisé et signalé dans l'en-tête d'image;
- le codage des coefficients d'un bloc INTER au moyen de la table INTRA VLC de l'Annexe I résulte en un moins grand nombre d'éléments binaires que par l'utilisation de la table INTER VLC;
- l'utilisation de la table INTRA VLC de l'Annexe I doit être détectable par le décodeur. Celui-ci part du principe que les coefficients sont codés au moyen de la table INTER VLC. Le décodeur détecte l'utilisation de la table INTRA VLC de l'Annexe I en cas d'adressage de coefficients extérieurs à l'étendue de 64 coefficients d'un bloc 8×8 .

Avec un grand nombre de grands coefficients, cela se produira facilement en raison de la façon dont la table INTRA VLC a été conçue et un grand nombre de bits pourra être économisé aux débits élevés.

III.3.4 Mode de codage INTRA avancé (Annexe I)

Le codage INTRA avancé est une méthode permettant d'améliorer le codage intrabloc au moyen de la prédiction interblocs. L'application de cette technique est décrite dans l'Annexe I/H.263. La procédure revient essentiellement à une prédiction intrabloc suivie d'une quantification comme décrit dans le § III.3.2.4, avec utilisation de différents ordres d'exploration et d'une table VLC différente pour le codage entropique des coefficients quantifiés.

Le codage intrabloc est effectué par le choix d'un des trois modes prédictifs décrits dans la Rec. UIT-T H.263. Noter que celle-ci fait appel aux coefficients DCT *reconstruits* afin d'exécuter la prédiction interblocs, tandis que les coefficients DCT *originaux* sont employés dans le codeur pour la décision quant au mode de prédiction. Le présent modèle d'essai décrit la décision de mode prise dans le décodeur sur la base des coefficients DCT originaux.

Les blocs de coefficients DCT employés au cours de la prédiction sont étiquetés $A(u, v)$, $B(u, v)$ et $C(u, v)$, où u et v sont respectivement les indices de rangée et de colonne. Le terme $C(u, v)$ indique les coefficients DCT du bloc à coder, $A(u, v)$ indique le bloc de coefficients DCT situés immédiatement au-dessus de $C(u, v)$ et $B(u, v)$ indique le bloc de coefficients DCT situé immédiatement à gauche de $C(u, v)$. La capacité d'utiliser les valeurs de coefficient reconstruites pour les blocs A et B dans la prédiction des valeurs de coefficient pour le bloc C dépend de la présence des blocs A et B dans le même segment d'image vidéo que le bloc C. Un segment d'image est défini dans l'Annexe R/H.263. Le terme $E_i(u, v)$ indique l'erreur de prédiction pour le mode INTRA $i = 0, 1, 2$. Les erreurs de prédiction sont calculées comme suit pour les trois modes de codage:

```
mode 0: prédiction DC seulement.
Si (le bloc A et le bloc B sont chacun à codage intra et sont chacun dans le
même segment d'image que le bloc C)
{
E0(0,0) = C(0,0) - ( A(0,0) + B(0,0) )//2
}
sinon
{
Si (le bloc A est à codage intra et est dans le même segment d'image que le
bloc C)
{
```

```

    E0(0,0) = C(0,0) - A(0,0)
  }
  sinon
  {
    Si (le bloc B est à codage intra et est dans le même segment d'image que
    le bloc C)
    {
      E0(0,0) = C(0,0) - B(0,0)
    }
    sinon
    {
      E0(0,0) = C(0,0) - 1024
    }
  }
}

```

$E0(u,v) = C(u,v) \quad u!=0, v!=0, u = 0..7, v = 0..7.$

mode 1: prédiction DC et AC à partir du bloc supérieur.

Si (le bloc A est à codage intra et est dans le même segment d'image que le bloc C)

```

{
  E1(0,v) = C(0,v) - A(0,v)      v = 0..7, et
  E1(u,v) = C(u,v)                u = 1..7, v = 0..7.
}

```

```

}
sinon
{

```

```

  E1(0,0) = C(0,0) - 1024
  E1(u,v) = C(u,v)      (u,v) != (0,0), u = 0,_,7, v = 0,_,7
}

```

mode 2: prédiction DC et AC à partir du bloc de gauche.

Si (le bloc B est à codage intra et est dans le même segment d'image que le bloc C)

```

{
  E2(0,v) = C(u,0) - A(u,0)      u = 0..7, et
  E2(u,v) = C(u,v)                v = 1..7, u = 0..7.
}

```

```

}
sinon
{

```

```

  E2(0,0) = C(0,0) - 1024
  E2(u,v) = C(u,v)      (u,v) != (0,0), u = 0,_,7, v = 0,_,7
}

```

La sélection du mode de prédiction pour le codage INTRA avancé est effectuée par évaluation de la somme absolue de l'erreur de prédiction, $SAD_{\text{mode } i}$, pour les quatre blocs de luminance contenus dans le macrobloc et par sélection du mode avec la valeur minimale.

$$SAD_{\text{mode } i} = \sum_b \left[|E_i(0,0)| + 32 \sum_u |E_i(u,0)| + 32 \sum_v |E_i(0,v)| \right]$$

où:

$$i = 0 \dots 3,$$

$$b = 0 \dots 3,$$

$$u, v = 1 \dots 7.$$

Une fois le mode approprié choisi, la quantification est effectuée. Les blocs sont quantifiés comme décrit dans le § III.3.2.4.

III.3.5 Mode de quantification modifié (Annexe T)

Le mode de quantification modifié de l'Annexe T réduit considérablement certains artefacts chromatiques (en particulier aux bas débits) et augmente l'étendue des coefficients de luminance. Par ailleurs, l'Annexe T permet au codeur de donner à l'échelon de quantification une valeur quelconque à l'échelle des macroblocs, ce qui peut améliorer la performance des algorithmes de commande de débit. L'Annexe T est autorisée pour tous les scénarios applicatifs et est fortement conseillée pour les modèles de produit à bas débit binaire.

III.4 Algorithmes utilisés pour des scénarios applicatifs individuels

Le présent paragraphe examine les algorithmes utilisés pour des scénarios applicatifs spécifiques.

III.4.1 Décision de mode

La H.263 autorise plusieurs types de système de codage de macrobloc, comme le mode INTRA (codage de coefficients DCT non prédits), le mode INTER (codage prédictif utilisant 1 vecteur cinétique) et le mode INTER4V (codage prédictif utilisant 4 vecteurs cinétiques). Le choix du mode approprié est l'une des fonctions clés d'un codeur et la qualité de la décision influence grandement la performance du codeur. Un algorithme de qualité élevée et un algorithme de faible complexité ont été mis au point pour des environnements aussi bien exempts d'erreurs qu'exposés aux erreurs. Ces algorithmes sont décrits dans les paragraphes suivants. L'on décrira d'abord un mécanisme permettant d'effectuer la mise à jour obligatoire du codage INTRA.

III.4.1.1 Rafrâchissement et structure de mise à jour des macroblocs INTRA

Comme prescrit par la H.263, chaque macrobloc doit être codé en mode INTRA au moins toutes les 132 fois que des coefficients sont émis (à moins que la transformation IDCT de l'Annexe W ne soit en cours d'utilisation). Afin d'éviter de grandes salves de macroblocs INTRA pendant de brèves périodes, l'on fait appel à une structure simple de mise à jour des macroblocs afin d'initialiser le compteur de mises à jour de chaque macrobloc à une valeur comprise dans l'étendue [0, 132]. Il convient d'utiliser le générateur de nombres pseudo-aléatoires de l'Annexe A/H.263 pour initialiser la structure aléatoire. Ce processus de décision de mode remplace le mode de codage après tout autre processus de décision de mode. L'algorithme est décrit comme suit.

MB_intra_update[xpos][ypos]: compteur incrémenté d'une unité chaque fois que des informations de coefficient sont envoyées pour le macrobloc situé à la position (xpos, ypos), à savoir le macrobloc MB[xpos][ypos].

INTRA_MB_Refresh_Rate: dans les environnements sans erreur, il s'agit d'une constante (132). Dans les environnements exposés aux erreurs, INTRA_MB_Refresh_Rate peut être une variable à valeur d'entier (adaptée au taux d'erreur), avec une valeur dans l'étendue (1, 132).

Le mode INTRA pour un macrobloc donné est choisi comme suit:

```
Initialisation après une image I: MB_intra_update[xpos][ypos] =
random_of_Annex_A (0, INTRA_MB_Refresh_Rate)
Tant que (des images non I restent à coder)
  Si ((MB_intra_update[xpos][ypos] == INTRA_MB_REFRESH_RATE) &&
      (MB[xpos][ypos] contient coefficient))
  {
    Coder MB[xpos][ypos] en mode INTRA;
  }
  sinon si (MB[xpos][ypos] contient coefficient)
  {
    ++MB_intra_update[xpos][ypos];
  }
```

D'autres détails figurent dans les références [Q15-E-15] et [Q15-E-37].

III.4.1.2 Décision de mode de faible complexité pour environnements sans erreur

La Figure III.3 décrit l'algorithme de décision de mode de faible complexité, qui interagit avec les étapes de la recherche de vecteur cinétique de qualité basse et de qualité moyenne, décrite respectivement dans les § III.3.1.2 et III.3.1.3.

Après avoir effectué l'estimation cinétique à pixels entiers, le codeur sélectionne le mode de prédiction : INTRA ou INTER. Les paramètres suivants sont calculés afin de prendre la décision INTRA/INTER:

$$\mu_{MB} = \frac{1}{256} \sum_{i=1}^{16} \sum_{j=1}^{16} X(i, j)$$

$$A = \sum_{i=1}^{16} \sum_{j=1}^{16} |X(i, j) - \mu_{MB}|$$

Le mode INTRA est choisi si:

$$A < (SAD(u, v) - 500)$$

Si le mode INTRA est choisi, aucune autre étape de décision de mode ou d'estimation cinétique n'est nécessaire. Si la commande de débit est activée, les coefficients DCT des blocs sont quantifiés au moyen du paramètre de quantification déterminé par la commande de débit. Les opérations du mode de codage INTRA avancé sont effectuées comme indiqué dans le § III.3.4. La structure de bloc codée (CBP, *coded block pattern*) est calculée et le flux binaire pour ce macrobloc est produit.

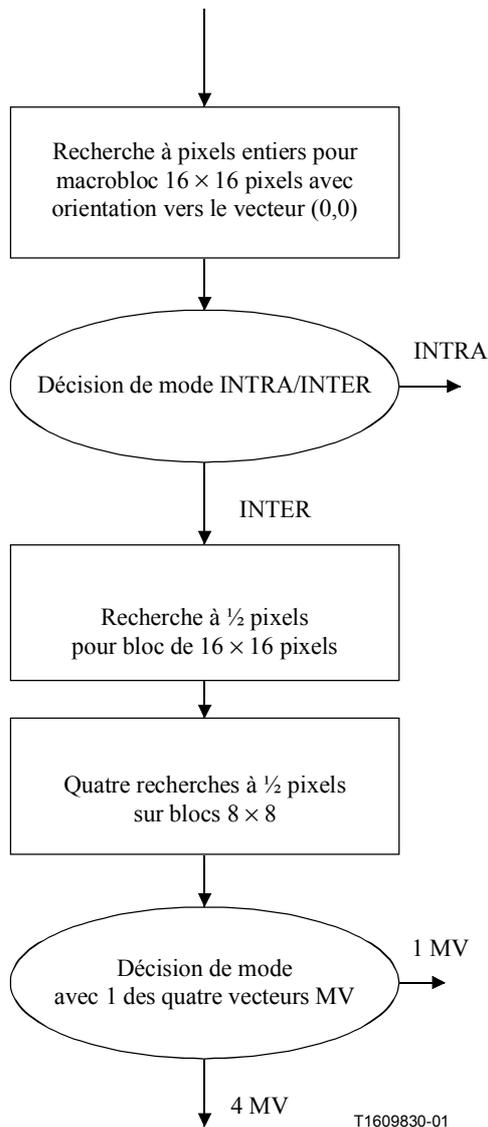


Figure III.3/H.263 – Schéma fonctionnel du processus de décision de mode à faible complexité

Si le mode INTER est choisi, la recherche cinétique continue par un raffinement à demi-pixels autour du vecteur MV à pixels entiers sur 16×16 .

Si des vecteurs cinétiques de bloc 8×8 sont autorisés, c'est-à-dire si l'Annexe F ou l'Annexe J ou les deux sont activées, la recherche cinétique continue également par une recherche à demi-pixels autour des vecteurs MV à pixels entiers sur 8×8 . La décision de mode intervient ensuite pour déterminer s'il convient d'utiliser la prédiction INTER ou INTER4V.

La somme SAD pour le vecteur 16×16 à la précision des demi-pixels [y compris la soustraction de 100 si le vecteur est (0,0)] est la suivante:

$$SAD_{16}(u, v)$$

La somme SAD cumulée pour les 4 vecteurs 8×8 à la précision des demi-pixels est la suivante:

$$SAD_{8 \times 8} = \sum_{k=1}^4 SAD_{8,k}(u_k, v_k)$$

Le mode INTER4V est choisi si:

$$SAD_{8 \times 8} < SAD_{16} - 200$$

sinon, c'est le mode INTER qui est choisi.

III.4.1.3 Décision de mode de qualité élevée pour environnements sans erreur

L'on effectue d'abord l'estimation cinétique de qualité élevée pour les modes INTER et INTER4V, comme décrit au § III.3.1.4. C'est sur la base de ces vecteurs cinétiques que l'on calcule les coûts débit-distorsion globaux pour tous les modes de codage considérés.

Le codeur sélectionne indépendamment le mode de codage de macrobloc qui minimise l'expression:

$$J = D + \lambda_{\text{mode}} \times R$$

La distorsion, D , est définie comme la somme des différences au carré (SSD, *sum of squared difference*) entre les coefficients des composantes de luminance et de chrominance du macrobloc cible et les coefficients des composantes de luminance et de chrominance pour le mode de codage donné. Le débit R est défini comme étant la vitesse de codage du macrobloc cible y compris toutes les informations de commande, de mouvement et de texture, pour le mode codage donné.

Le paramètre λ est choisi par l'expression:

$$\lambda_{\text{mode}} = 0,85 \cdot QP^2$$

où QP est le paramètre de quantification de macrobloc [Q15-D-13].

Les modes de codage explicitement testés sont SKIPPED, INTRA, INTER et INTER4V. Le mode INTER utilise le meilleur vecteur cinétique 16×16 et le mode INTER4V utilise les meilleurs vecteurs cinétiques 8×8 sélectionnés par l'algorithme d'estimation cinétique de qualité élevée. Si l'application vise un débit binaire fixe, la relation est modifiée comme décrit au § III.4.2.3.

Si l'Annexe F est utilisée, la fenêtre de compensation cinétique des blocs superposés n'est pas prise en considération pour prendre les décisions relatives au mode de codage. Elle n'est appliquée qu'une fois que tous les vecteurs cinétiques ont été déterminés et que les décisions de mode ont été prises pour l'image.

III.4.1.4 Décision de mode de faible complexité pour environnements exposés aux erreurs

L'algorithme de décision de mode de faible complexité pour environnements exposés aux erreurs utilise le mécanisme déjà décrit au § III.4.1.2, avec un seul complément. La fréquence de mise à jour des macroblocs INTRA est augmentée par l'utilisation d'une valeur inférieure à 132 pour le paramètre INTRA_MB_Refresh_Rate du § III.4.1.1. L'on pourra trouver des directives sur le choix d'une valeur pour cette variable dans les § III.2.3.2 et III.2.4.1.

III.4.1.5 Décision de mode de qualité élevée pour environnements exposés aux erreurs

L'algorithme de décision de mode de qualité élevée pour environnements exposés aux erreurs est semblable à celui qui a été décrit au § III.4.1.3. Indépendamment pour chaque macrobloc, le codeur sélectionne le mode de codage qui minimise l'expression:

$$J = (1 - p)D_1 + pD_2 + \lambda_{\text{mode}}R$$

où λ_{mode} est la variable déjà indiquée au § III.4.1.3. Pour calculer ce coût, le codeur simule la méthode de masquage d'erreur employée par le décodeur et doit obtenir la probabilité qu'un macrobloc donné soit perdu avec la probabilité p pendant la transmission. La distorsion provient donc de deux origines, D_1 étant la distorsion attribuée à un macrobloc reçu et D_2 la distorsion attribuée à un macrobloc perdu et masqué.

Pour un macrobloc donné, deux distorsions sont calculées pour tous les modes de codage considérés: la distorsion de codage D_1 et la distorsion de masquage D_2 . La distorsion D_1 est pondérée par la probabilité $(1 - p)$ que ce macrobloc soit reçu sans erreur et la distorsion D_2 est pondérée par la probabilité p que le même macrobloc soit perdu et masqué. La distorsion D_2 dépend de la méthode de masquage d'erreur employée. Elle est constante pour tous les modes de codage considérés (et n'a donc pas d'incidence sur le mode qui est choisi au moyen de l'équation ci-dessus). La distorsion D_1 dépendra du mode de codage choisi. Pour le mode INTRA, la distorsion est la même que dans le cas d'un environnement sans erreur. Pour les modes SKIP et INTER, la distorsion D_1 est la distorsion de quantification plus la distorsion de masquage du macrobloc à partir duquel elle est prédite, soit plus précisément:

$$D_1 = D_q + pD_2(v, n-1)$$

où $D_2(v, n-1)$ représente la distorsion de masquage de la zone de l'image précédente $(n-1)$ à l'emplacement spatial actuel du macrobloc déplacé par le vecteur cinétique v , où v est mis à zéro pour le mode SKIP et mis à la valeur du vecteur cinétique du macrobloc considéré pour le mode INTER.

La distorsion D_q est définie comme la somme SSD calculée entre les valeurs des pixels correctement reconstruits du macrobloc codé et les valeurs des pixels originaux. La distorsion D_2 est définie comme la somme SSD calculée entre les valeurs des pixels correctement reconstruits du macrobloc codé et les valeurs des pixels masqués. Le débit R est défini comme étant la vitesse de codage du macrobloc cible, y compris toutes les informations de commande, de mouvement et de texture pour le mode de codage en question.

III.4.2 Commande de débit

Cet algorithme de commande de débit s'applique à tous les scénarios applicatifs à débit binaire fixe. Il se compose d'une couche des trames, dans laquelle on choisit un débit cible pour l'image en cours, et d'une couche des macroblocs, dans laquelle le paramètre de quantification (QP, *quantization parameter*) est adapté afin de parvenir à cette cible. Le développement de cet algorithme est décrit dans la référence [Q15-A-20]. Cet algorithme de commande de débit, souvent appelé commande de débit TMN8, est le seul algorithme de commande de débit qui est utilisé pour les scénarios applicatifs.

Initialement, le nombre de bits contenus dans le tampon W est mis à zéro, soit $W = 0$, et les paramètres K_{prev} et C_{prev} sont initialisés à $K_{prev} = 0,5$ et $C_{prev} = 0$. La première image subit un codage INTRA au moyen d'une valeur fixe du paramètre QP pour tous les macroblocs.

III.4.2.1 Commande de débit dans la couche des trames

Les définitions suivantes sont utilisées pour décrire l'algorithme:

- B' nombre de bits occupés par la précédente image codée;
- R débit binaire cible en bits par seconde (par exemple 10 000 bit/s, 24 000 fps, etc.);
- G fréquence de trame de la séquence vidéo originale, en images par seconde (par exemple 30 fps);
- F fréquence de trame cible en images par seconde (par exemple 7,5 fps, 10 fps, etc.). Le rapport G/F doit être un nombre entier;
- M seuil de saut de trame. Par défaut, l'on a $M = R/F$ (le rapport M/R est le délai de mise en mémoire tampon.)
- A délai de mise en mémoire tampon, de AM s. Par défaut, $A = 0,1$.

Le nombre de bits contenus dans le tampon du codeur est de $W = \max(W + B' - R/F, 0)$. Le paramètre de saut est mis à 1, soit $\text{skip} = 1$.

```
Tant que W > M
{
    W = max (W - R/F, 0)
    skip++
}
```

Saut de codage des $(\text{skip} \cdot (G/F) - 1)$ images de la séquence vidéo originale. Le nombre cible de bits par image est $B = (R/F) - \Delta$, où:

$$\Delta = \begin{cases} \frac{W}{F}, & W > A \cdot M \\ W - A \cdot M, & \text{Sinon} \end{cases}$$

Pour les applications à fréquence de trame fixe, le paramètre de saut est toujours égal à 1. Le saut de trame est constant et égal à $(G/F) - 1$. De même, lors du calcul du nombre cible de bits par image, il y a lieu de prendre $A = 0,5$.

III.4.2.2 Commande de débit dans la couche des macroblocs

Etape 1: initialisation

L'on part du principe que l'estimation du vecteur cinétique a déjà été effectuée.

σ_k^2 est défini comme la variance des valeurs de luminance et de chrominance dans le k ème macrobloc.

Si le k ème macrobloc est de type I (intra), poser $\sigma_k^2 = \sigma_k^2 / 3$.

Soit $i = 1$ et $j = 0$;

$\tilde{B}_1 = B$ le nombre cible de bits comme défini en A.1/H.263;

$N_1 = N$ le nombre de macroblocs dans une image;

$K = K_1 = K_{\text{prev}}$ la valeur initiale des paramètres du modèle;

$C = C_1 = C_{\text{prev}}$ la valeur initiale des paramètres du modèle.

$$S_1 = \sum_{k=1}^N \alpha_k \sigma_k, \text{ où } \alpha_k = \begin{cases} 2 \frac{B}{16^2 N} (1 - \sigma_k) + \sigma_k, & \frac{B}{16^2 N} < 0,5, \\ 1, & \text{Sinon} \end{cases}$$

Etape 2: calculer le quantificateur Q optimisé pour le i ème macrobloc

Si $L = (\tilde{B}_i - 16^2 N_i C) \leq 0$ (épuisement des bits), poser $Q_i^* = 62$.

Sinon, calculer:

$$Q_i^* = \sqrt{\frac{16^2 K}{L} \frac{\sigma_i}{\alpha_i} S_i}$$

Etape 3: trouver QP et coder le macrobloc

QP = arrondir $(Q_i^* / 2)$ au plus proche entier dans l'ensemble 1,2, ..., 31.

DQUANT = QP – QP_prev.

Si DQUANT > 2, poser DQUANT = 2. Si DQUANT < -2, poser DQUANT = -2.

Poser QP = QP_prev + DQUANT.

DCT coder le macrobloc avec le paramètre de quantification QP et poser QP_prev = QP.

Etape 4: mettre à jour les compteurs

Soit B'_i le nombre de bits utilisés pour coder le nième macrobloc, calculer:

$$\tilde{B}_{i+1} = \tilde{B}_i - B'_i, S_{i+1} = S_i - \alpha_i \sigma_i, \text{ et } N_{i+1} = N_i - 1$$

Etape 5: mettre à jour les paramètres de modèle K et C

Les paramètres de modèle mesurés pour le nième macrobloc sont :

$$\hat{K} = \frac{B'_{LC,i}(2QP)^2}{16^2 \sigma_i^2} \text{ et } \hat{C} = \frac{B'_i - B'_{LC,i}}{16^2}$$

où $B'_{LC,i}$ est le nombre de bits dépensés pour la luminance et la chrominance du macrobloc.

Ensuite, mesurer la moyenne des \hat{K} et des \hat{C} calculés jusque là dans l'image.

Si ($\hat{K} > 0$ et $\hat{K} \leq \pi \log_2 e$), poser $j = j + 1$ et calculer $\tilde{K}_j = \tilde{K}_{j-1}(j-1)/j + \hat{K}/j$.

Calculer $\tilde{C}_i = \tilde{C}_{i-1}(i-1)/i + \hat{C}/i$.

Finalement, les mises à jour sont une moyenne pondérée des estimations initiales, K_1 , C_1 , et leur moyenne actuelle:

$$K = \tilde{K}_j(i/n) + K_1(N-i)/N, \quad C = \tilde{C}_i(i/n) + C_1(N-i)/N.$$

Etape 6

Si $i = N$, arrêter l'opération (tous les macroblocs sont codés).

Poser $K_{prev} = K$ et $C_{prev} = C$.

Sinon, soit $i = i + 1$, et passer à l'étape 2.

III.4.2.3 Modèles et commande de débit de qualité élevée

Pour les minimisations de type lagrangien, comme celles qui sont décrites respectivement dans les algorithmes d'estimation de mouvement et de décision de mode de qualité élevée dans les § III.3.1.4 et III.4.1.3, le paramètre lagrangien doit rester constant dans une image entière parce qu'idéalement toutes les régions d'image doivent fonctionner à un point constant sur la pente de leur courbe débit-distorsion [SG88]. Théoriquement, ce point de fonctionnement optimal est trouvé par recherche parmi toutes les valeurs possibles de λ . Dans le présent modèle d'essai, l'on définit une relation simple entre λ et le paramètre de quantification afin d'obtenir une bonne approximation de la valeur optimale de λ [Q15-D-13].

Etant donné que l'algorithme de commande de débit qui a été décrit ci-dessus peut mettre à jour le paramètre de quantification à chaque macrobloc codé, les relations entre λ et le paramètre de quantification doivent être légèrement modifiées. L'on a montré que, pour employer les modèles de qualité élevée dans les scénarios à débit fixe, il convenait de calculer comme suit les paramètres lagrangiens:

$$\lambda_{\text{motion}} = 0,92 \cdot \overline{QP}_{prev} \text{ et}$$

$$\lambda_{\text{mode}} = 0,85 \cdot \overline{QP^2}_{\text{prev}}$$

respectivement, où $\overline{QP}_{\text{prev}}$ est la valeur moyenne, dans tous les macroblocs, du paramètre de quantification issu de l'image précédente à codage du même type que l'image actuelle.

Les algorithmes d'estimation du mouvement et de décision de mode de qualité élevée procèdent ensuite à la détermination des vecteurs cinétiques appropriés et des codes de codage de macrobloc appropriés, comme dans les scénarios à débit variable, c'est-à-dire comme si aucun algorithme de commande de débit n'avait été activé. Une fois que l'on a déterminé les vecteurs cinétiques et les modes optimaux, l'algorithme de commande de débit peut servir à coder l'image sur la base des vecteurs cinétiques et des modes de codage déjà déterminés. Dans les cas particuliers où l'algorithme de décision de mode a sélectionné le mode *SKIPPED* pour le macrobloc cible, l'algorithme de commande de débit n'est pas autorisé à mettre à jour le paramètre de quantification.

III.4.2.4 Sélection d'image de référence améliorée (Annexe U)

Le mode de sélection d'image de référence améliorée (ERPS, *enhanced reference picture selection*) (Annexe U) permet une prédiction à compensation de mouvement (MCP, *motion-compensated prediction*) d'images multiples. En fonctionnement de sélection ERPS, M (avec $M \geq 1$) images déjà décodées sont mises en mémoire tampon dans le codeur et dans le décodeur. Contrairement à la prédiction MCP d'image unique, la prédiction MCP d'images multiples peut exploiter les relations statistiques à long terme qui sont contenues dans les séquences vidéo. Exemples représentatifs de ces relations: coupures de scènes répétées, masquage-dégagement d'arrière-plan, fondu-enchaîné du contenu d'image, etc.

III.4.2.4.1 Choix du nombre d'images de référence

Dans les systèmes du monde réel, le nombre M d'images de référence qui peuvent être prises en charge par le décodeur est négocié ou, si ce n'est pas possible, établi par des moyens externes. Les améliorations de la courbe débit-distorsion par rapport à la prédiction MCP d'image unique dépendent du nombre d'images de référence et de la façon dont elles sont utilisées. Un plus grand nombre d'images de référence se traduira habituellement par une meilleure qualité du codage mais aussi par une augmentation de la complexité et de la mémoire nécessaire. Par intégration des résultats de plusieurs séquences d'essai, l'on a constaté que les améliorations du rapport PSNR à débit fixe étaient généralement proportionnelles au logarithme du nombre ($\log(M)$) d'images de référence. Aux fins du modèle d'essai, 10 images de référence sont donc utilisées afin d'obtenir un bon compromis entre l'efficacité et la complexité du codage. Bien qu'un codeur puisse parvenir à une qualité améliorée au moyen d'un choix optimisé des images de référence qui doivent être mémorisées (et peut-être aussi grâce à un choix optimisé des sous-régions d'image à mémoriser), le présent modèle d'essai se limite à mémoriser et à utiliser les images de référence en mode premier arrivé, premier servi (FIFO, *first-in, first-out*).

III.4.2.4.2 Estimation du mouvement

L'estimation du mouvement consiste à utiliser un nombre inférieur ou égal à M (=10 dans le présent modèle d'essai) images de référence se trouvant en mémoire tampon d'images multiples. Pour déterminer le vecteur cinétique optimal \mathbf{m}_k pour le k ème macrobloc ou bloc, la fonction de coût lagrangien suivante est minimisée:

$$\mathbf{m}_k = \underset{\mathbf{m} \in \text{SR}}{\operatorname{argmin}} D(\mathbf{s}, \mathbf{c}(\mathbf{m})) + \lambda_{\text{motion}} R(\mathbf{m} - \mathbf{p})$$

SR étant l'étendue de recherche englobant normalement l'ensemble des positions de pixels entiers $[-32...31] \times [-32...31] \times [0...M-1]$ dans les plans horizontal, vertical et temporel.

NOTE – La plupart des expériences conduites au cours des essais de la phase de proposition du mode de sélection ERPS ont en fait utilisé une étendue de recherche à échantillon d'entiers égal à ± 16 . Une plus grande étendue est cependant spécifiée ici parce qu'elle est disponible dans le modèle d'essai qui a été conçu.

La distorsion $D(\mathbf{s}, \mathbf{c}(\mathbf{m}))$ est indiquée sous forme de somme SAD et le terme $R(\mathbf{m} - \mathbf{p})$ détermine le débit y compris le déplacement spatial m_x, m_y et le paramètre d'image de référence m_t en fonction du prédicteur $\mathbf{p} = (p_x, p_y, 0)$. Après détermination du vecteur cinétique à pixels entiers du macrobloc ou du bloc, l'on applique le reste des étapes de codage spécifiées dans le mode de qualité élevée du § III.4.2.3. Le paramètre lagrangien λ_{motion} est choisi comme suit: $\lambda_{\text{motion}} = 0,92 \text{ QP}$.

III.4.2.5 Tranches de données partitionnées (Annexe V)

Le mode de tranche de données partitionnées est utilisé dans l'option V3 du scénario applicatif à degré élevé d'exposition aux erreurs sur les bits. L'on estime que ce mode améliorerait également la performance dans les environnements exposés à la perte de paquets, à condition que l'on utilise des procédés d'empaquetage affectant au moins un paquet à chaque partition. Il est possible de réduire la propagation des erreurs entre partitions et l'utilisation de codes RVLC en réorganisant la syntaxe et en formant des tranches à 3 partitions: une pour l'en-tête, une autre pour le vecteur cinétique et la troisième pour les coefficients DCT. Contrairement à d'autres parties du présent modèle d'essai, le présent paragraphe contient des descriptions d'algorithme pour le fonctionnement du codeur comme du décodeur.

III.4.2.5.1 Fonctionnement du codeur

Le codage source des données est inchangé et seule la syntaxe des tranches est modifiée par rapport à l'Annexe K. Afin d'insérer les données vidéo codées dans les tranches, l'on utilise une structure en paquet à longueur pseudofixe. Le codeur prélève un seuil t de longueur de tranche, qu'il convient de choisir en fonction des caractéristiques d'erreur et de surdébit du réseau. Une fois que le seuil t a été choisi, le codeur code un nombre entier de macroblocs de façon que la longueur de tranche correspondante soit inférieure à t . Théoriquement, le seuil t donnera des tranches contenant chacune 1 ou 2 erreurs au plus afin de permettre d'obtenir un avantage maximal pour la récupération par codage RVLC. Une bonne limite inférieure de longueur de paquet est donc égale à $0,5 * 1/\text{BER}$ bits, le BER étant le taux moyen d'erreur sur les bits dans un canal. Cela n'est toutefois pas toujours possible à cause du surdébit mis en jeu dans les très courtes tranches. Compte tenu de tout cela, la réalisation du modèle d'essai de l'Annexe V a choisi un seuil t de 700 bits car il a été démontré que ce nombre produisait de bons résultats pour le canal en modulation WCDMA à 64 kbit/s avec $\text{BER} = 10E - 3$ (ce qui est le canal le moins favorable pris en compte dans les conditions communes [Q15-I-60]).

III.4.2.5.2 Méthode de décodage

Le présent paragraphe décrit la méthode utilisée pour décoder le flux binaire vidéo, y compris toutes les techniques employées pour la détection et le masquage des erreurs. La description algorithmique non formelle remplace pour autant les définitions plus générales du § III.5.3. Le masquage des erreurs est en général indépendant de la méthode décrite ici. Il est donc effectué conformément à la méthode décrite au § III.5.4.

Comme dans le cas de la méthode de codage, le décodage des données vidéo proprement dites est le même, la seule différence concernant la récupération des données et le traitement des erreurs en fonction de la nouvelle structure à données partitionnées.

- 1) Explorer le flux binaire à la recherche d'un code PSC. Lorsqu'un code PSC est reçu, l'on part du principe qu'un en-tête d'image le suit immédiatement.
- 2) Décoder l'en-tête d'image. Le processus est exactement le même que lorsque l'Annexe V n'est pas utilisée. Si l'on détecte une erreur dans l'en-tête d'image, l'on utilise les

informations contenues dans le dernier en-tête d'image sans erreur connu ou dans une copie redondante de l'Annexe W, si disponible.

- 3) Rechercher le code SSC suivant, qui devrait suivre immédiatement l'en-tête d'image et le mot SSTUF. Si ce n'est pas le cas, l'on suppose une erreur et l'on rejette tous les bits compris entre l'en-tête d'image et ce code SSC.

NOTE – Etant donné que le code SSC est un préfixe du code PSC, il faut vérifier si un code PSC existe réellement. Si c'est le cas, la boucle de répétition suivante ne sera évidemment pas exécutée.

- 4) Répéter ce qui suit jusqu'à la détection du prochain code PSC:
 - a) décoder les mots SEPBI et SSBI;
 - b) le mot MBA suit immédiatement le mot SSBI. Deux cas sont possibles:
 - i) si cette tranche suit immédiatement un code PSC et un en-tête d'image, l'on sait que l'indice du premier MB devrait être 0. S'il est différent de zéro, on le corrige pour le mettre à 0 et l'on continue le décodage de la tranche;
 - ii) sinon, l'on sait quel indice de premier MB devrait être fondé sur la tranche immédiatement précédente. Deux cas sont alors possibles:
 - aucune erreur dans la tranche immédiatement précédente: mettre le mot MBA à une valeur prédite d'après la tranche immédiatement précédente;
 - tranche immédiatement précédente manquante ou contenant une erreur: mettre le mot MBA à la valeur lue;
 - c) décoder les mots SEPBI, SQUANT, SWI, SEPBI et GFID;
 - d) rechercher le prochain marqueur d'en-tête (HM, *header marker*);
 - e) décoder la partition des données d'en-tête (HD, *header data*). Ce décodage n'est effectué que vers l'avant. Si une erreur est détectée, l'ensemble de la tranche est rejeté jusqu'au prochain code PSC ou SSC.
- 5) Rechercher un marqueur de mouvement (MVM). Si l'on trouve d'abord un code PSC ou SSC, une erreur est détectée et la tranche est ignorée.
- 6) Décoder la partition des données de vecteurs cinétiques (MVD et LMVV), vers l'avant et vers l'arrière. Conserver tous les vecteurs cinétiques décodés sans erreur dans les deux sens.
- 7) Rechercher le prochain code PSC ou SSC
- 8) Décoder la partition des données de coefficient. Etant donné que cette partition est en codage VLC et non RVLC, l'on ne décode que vers l'avant. Si une erreur est détectée quelque part, l'ensemble des coefficients DCT est mis à 0, ce qui implique logiquement que les informations résiduelles d'erreur sont perdues.

III.4.2.6 Utilisation du précédent algorithme d'en-tête de l'Annexe W

Par opposition à l'algorithme recommandé au § 5.1.4.3/H.263 (Version 2), le bit du type d'arrondissement du champ PLUSTYPE est mis à "0" quel que soit le type d'image [Q15-I-26]. Cela contribue au maintien d'une valeur constante de l'identificateur GFID et permet donc au décodeur de récupérer comme suit un en-tête d'image manquant: si l'identificateur GFID contenu dans le premier en-tête de groupe GOB décodable est le même que dans l'image précédente, les données d'image sont décodées au moyen de l'en-tête d'image extrait de l'image précédente. Sinon (si l'identificateur GFID diffère de celui de l'image précédente), une copie de l'en-tête d'image est insérée dans le champ d'informations supplémentaires d'amélioration de l'image suivante, comme défini au W.6.3.6/H.263 et le décodeur utilise cette copie pour récupérer un en-tête d'image manquant (voir [Q15-J-62] et les références y incluses).

III.5 Post-traitement dans le décodeur

Le présent paragraphe définit le fonctionnement du décodeur vidéo dans les environnements exposés aux erreurs ainsi que les méthodes de post-traitement comme les algorithmes de masquage des erreurs et de post-filtrage. Bien que le post-filtrage soit utile dans les environnements avec ou sans erreur, tous les autres mécanismes décrits ici ne s'appliquent qu'aux environnements exposés aux erreurs.

III.5.1 Post-filtre de soustraction des suroscillations

Le post-filtre de soustraction des suroscillations qui est décrit ci-dessous devrait être utilisé chaque fois que l'Annexe J est appliquée. Sinon, il y a lieu d'employer le post-filtre de dégroupage qui est décrit au § III.5.2. Voir dans l'Annexe J/H.263 la définition du filtre inséré dans la boucle de codage ainsi que la définition des fonctions et de la notation utilisées dans le présent paragraphe.

La version unidimensionnelle du filtre sera décrite. Afin d'obtenir un effet bidimensionnel, le filtre est d'abord utilisé dans le sens horizontal puis dans le sens vertical. Les pixels A, B, C, D, E, F, G, (H) sont alignés horizontalement ou verticalement. Une nouvelle valeur (D1) sera produite par le filtre pour D comme suit:

$D1 = D + \text{Filtre} ((A + B + C + E + F + G - 6D)/8, \text{Puissance 1})$ lors d'un filtrage dans le premier sens.

$D1 = D + \text{Filtre} ((A + B + C + E + F + G - 6D)/8, \text{Puissance 2})$ lors d'un filtrage dans le second sens.

Contrairement au filtre de l'Annexe J, le post-filtre est appliqué à tous les pixels contenus dans l'image. Les pixels périphériques devraient être répétés lorsque le filtre est appliqué aux limites d'image. Les paramètres Puissance1 et Puissance2 peuvent être différents afin de mieux adapter la puissance totale du filtre au paramètre QUANT. La relation entre Puissance1, Puissance2 et QUANT est indiquée dans le Tableau III.1. Les paramètres Puissance1 et Puissance2 peuvent être associés à QUANT pour le macrobloc dont D dépend ou être associés à une certaine valeur moyenne de QUANT dans certaines parties de l'image ou dans la totalité de celle-ci.

Une technique de fenêtre glissante peut être utilisée pour obtenir la somme de 7 pixels (A + B + C + D + E + F + G), ce qui permet de réduire le nombre d'opérations nécessaires pour implémenter le filtre.

Tableau III.1/H.263 – Modèle TMN

QUANT	Puissance	Puissance 1	Puissance 2	QUANT	Puissance	Puissance 1	Puissance 2
1	1	1	1	17	8	3	3
2	1	1	1	18	8	3	3
3	2	1	1	19	8	3	3
4	2	1	1	20	9	3	3
5	3	1	1	21	9	3	3
6	3	2	1	22	9	3	3
7	4	2	1	23	10	3	3
8	4	2	2	24	10	4	3
9	4	2	2	25	10	4	3
10	5	2	2	26	11	4	3
11	5	3	2	27	11	4	3
12	6	3	2	28	11	4	3

Tableau III.1/H.263 – Modèle TMN

QUANT	Puissance	Puissance 1	Puissance 2	QUANT	Puissance	Puissance 1	Puissance 2
13	6	3	2	29	12	4	3
14	7	3	2	30	12	4	3
15	7	3	3	31	12	4	3
16	7	3	3				

III.5.2 Post-filtre de dégroupage

Le présent paragraphe décrit un post-filtre régularisé à deux dimensions permettant de réduire simultanément les artefacts dus au groupage et aux suroscillations. Ce filtre est recommandé lorsque l'Annexe J n'est pas appliquée, de façon à pouvoir éviter des résultats trop flous. Le modèle de dégradation due au processus de quantification peut être écrit sous la forme suivante:

$$g(i, j) = f(i, j) + n(i, j)$$

où g, f, n dénotent l'image reconstruite, l'image originale et le bruit de quantification dans le domaine spatial, et où (i, j) représentent les coordonnées verticales et horizontales de l'image vidéo. Un facteur de lissage à quatre dimensions, M , est incorporé afin de minimiser le bruit de quantification. Sa définition est la suivante:

$$M(f(i, j)) = M_L(f(i, j)) + M_R(f(i, j)) + M_U(f(i, j)) + M_D(f(i, j))$$

$$M_L(f(i, j)) = (1 - \alpha_L(f(i, j)))[f(i, j) - g(i, j)]^2 + \alpha_L(f(i, j))[f(i, j) - f(i, j - 1)]^2$$

$$M_R(f(i, j)) = (1 - \alpha_R(f(i, j)))[f(i, j) - g(i, j)]^2 + \alpha_R(f(i, j))[f(i, j) - f(i, j + 1)]^2$$

$$M_U(f(i, j)) = (1 - \alpha_U(f(i, j)))[f(i, j) - g(i, j)]^2 + \alpha_U(f(i, j))[f(i, j) - f(i, j - 1, j)]^2$$

$$M_D(f(i, j)) = (1 - \alpha_D(f(i, j)))[f(i, j) - g(i, j)]^2 + \alpha_D(f(i, j))[f(i, j) - f(i, j + 1, j)]^2$$

où α_i ($i = L, R, U, D$) représente les paramètres de régularisation régissant la contribution relative en termes de fidélité et de lissage directionnel. La Figure III.4 montre un exemple de position de ces pixels.

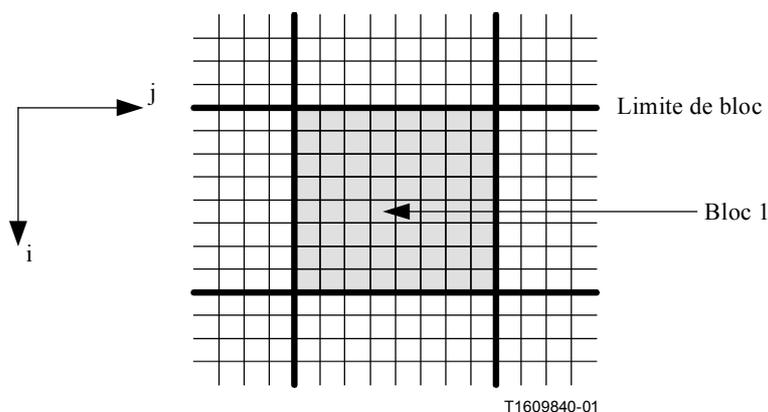


Figure III.4/H.263 – Exemple de positions des pixels

Le bit COD sert à estimer la corrélation d'un macrobloc entre une trame actuelle et l'image codée précédente. Différents filtres sont donc utilisés selon le bit COD, ce qui permet de réduire la complexité. La description suivante s'applique aux filtres lorsque le bit COD est 0 (image codée) ou 1 (image non codée).

Condition 1: le bloc 1 appartient à un macrobloc codé (COD = 0)

Si le bloc 1, représenté (par la région grise) sur la Figure III.4 appartient à un macrobloc codé, le filtre suivant est appliqué à tous les pixels du bloc 1. Lorsque le pixel $g(i, j)$ appartient au bloc 1, le pixel $f(i, j)$ est obtenu par minimisation de la fonction suivante:

$$f(i, j) = \frac{(4 - \alpha_{TOT}(f(i, j)))g(i, j) + A}{4}$$

$$\alpha_{TOT}(f(i, j)) = \alpha_L(f(i, j)) + \alpha_R(f(i, j)) + \alpha_U(f(i, j)) + \alpha_D(f(i, j))$$

$$A = \alpha_L(f(i, j))g(i, j-1) + \alpha_R(f(i, j))g(i, j+1) + \alpha_U(f(i, j))g(i-1, j) + \alpha_D(f(i, j))g(i+1, j)$$

Les paramètres de régularisation sont choisis comme suit:

$$\alpha_L(f(i, j)) = \frac{K_L QP_i^2}{[g(i, j) - g(i, j-1)]^2 + K_L QP_i^2}$$

$$\alpha_R(f(i, j)) = \frac{K_R QP_i^2}{[g(i, j) - g(i, j+1)]^2 + K_R QP_i^2}$$

$$\alpha_U(f(i, j)) = \frac{K_U QP_i^2}{[g(i, j) - g(i-1, j)]^2 + K_U QP_i^2}$$

$$\alpha_D(f(i, j)) = \frac{K_D QP_i^2}{[g(i, j) - g(i+1, j)]^2 + K_D QP_i^2}$$

où QP_i indique la valeur du paramètre de quantification (grandeur d'échelon de quantification) du dernier macrobloc codé. Etant donné qu'un plus grand lissage est nécessaire aux limites du bloc et qu'un lissage relativement plus petit est requis à l'intérieur d'un bloc, les constantes, K_i ($i = L, R, U, D$) sont déterminées en fonction de la position de $g(i, j)$ soit:

$$K_L, K_U = \begin{cases} 9 & \text{si } j \bmod 8 = 0 \\ 1 & \text{sinon} \end{cases}$$

$$K_R, K_D = \begin{cases} 9 & \text{si } j \bmod 8 = 7 \\ 1 & \text{sinon} \end{cases}$$

Condition 2: le bloc 1 appartient à un macrobloc non codé (COD = 1)

Si le bloc 1 représenté dans la Figure III.4 (région grisée) appartient à un macrobloc non codé, la règle suivante est appliquée à tous les pixels du bloc 1.

$$f(i, j) = f_p(i, j)$$

où f_p représente les pixels correspondant à l'image précédente qui a été récupérée.

Le fragment de pseudo-code suivant peut être utilisé pour faciliter l'implémentation de l'algorithme décrit:

```
f /* tableau à deux entrées, trame récupérée */
g /* tableau à deux entrées, trame reconstruite */
QP /* le paramètre de quantification */
```

```

K_L /* constante de gauche pour paramètre de régularisation */
K_R /* constante de droite pour paramètre de régularisation */
K_U /* constante du haut pour paramètre de régularisation */
K_D /* constante du bas pour paramètre de régularisation */
alpha_L /* paramètre de régularisation vers la gauche */
alpha_R /* paramètre de régularisation vers la droite */
alpha_U /* paramètre de régularisation vers le haut */
alpha_D /* paramètre de régularisation vers le bas */

QP=(QP*QP);
si(COD == 0)
{
    diff_L=(g(i,j)-g(i,j-1));
    diff_R=(g(i,j)-g(i,j+1));
    diff_U=(g(i,j)-g(i-1,j));
    diff_D=(g(i,j)-g(i+1,j));

    si(i mod 8 ==0){
        K_U=9;
        K_D=1;
    } sinon si(i mod 8 == 7){
        K_U=1;
        K_D=9;
    } sinon{
        K_U=1;
        K_D=1;
    }

    si(j mod 8 ==0){
        K_L=9;
        K_R=1;
    } sinon si(j mod 8 == 7){
        K_L=1;
        K_R=9;
    } sinon{
        K_L=1;
        K_R=1;
    }

    alpha_L=((K_L*QP)/(diff_L*diff_L+K_L*QP));
    alpha_R=((K_R*QP)/(diff_R*diff_R+K_R*QP));
    alpha_U=((K_U*QP)/(diff_U*diff_U+K_U*QP));
    alpha_D=((K_D*QP)/(diff_D*diff_D+K_D*QP));

    f(i,j)=(((4-alpha_L-alpha_R-alpha_U-alpha_D)*g(i,j)+alpha_L*g(i,j-
    )+alpha_R*g(i,j+1)+alpha_U*g(i-1,j)+alpha_D*g(i+1,j))/4);
}

```

III.5.3 Détection d'erreur

Certaines parties du présent paragraphe sont fondées sur le modèle de masquage d'erreur TCON qui est décrit dans la référence LBC-95-186. Dans les environnements exposés aux erreurs, il est nécessaire que le décodeur soit tolérant aux flux binaires reçus par erreur. En pratique, il existe habituellement certains mécanismes dans la hiérarchie du protocole de transport qui fournissent des informations sur les parties détériorées ou perdues du flux binaire, comme la vérification du numéro de séquence RTP ou du CRC AL3 de la Rec. UIT-T H.223. Il est possible de parvenir à une performance raisonnable en subdivisant les images codées en segments (plus ou moins) décodables indépendamment, au moyen d'en-têtes de groupe GOB ou de tranches, même sans mécanismes élémentaires de reprise sur erreur dans le décodeur vidéo proprement dit.

Dans les scénarios applicatifs exposés aux erreurs qui sont décrits aux § III.2.3 et III.2.4, le flux binaire ne contient jamais d'erreurs sur les bits car la pile de transport assure déjà des environnements exempts d'erreur au moyen de contrôles de CRC. La seule erreur syntaxique possible est donc un défaut d'information entre marqueurs de synchronisation (image, en-têtes de

groupe GOB ou de tranche). Cette erreur doit être détectée au cours du décodage et le masquage d'erreur décrit dans le § III.5.4 doit être appliqué aux macroblocs manquants. Si un en-tête d'image est perdu, il est remplacé par l'en-tête de l'image précédente ou par un en-tête d'image redondant, si disponible (par exemple dans l'en-tête de charge utile du protocole RTP).

Si des erreurs sur les bits sont présentes dans le flux binaire reçu, ces erreurs doivent être détectées par le décodeur vidéo au moyen de violations syntaxiques ou sémantiques comme les suivantes:

- Un mot de code illégal est trouvé. C'est l'événement le plus fréquent pour arrêter le décodage.
- Un marqueur de synchronisation n'arrive pas après la reconstruction d'une ligne de macrobloc. (Cette situation ne peut pas être vérifiée si l'on ne sait pas que des codes GBSC sont utilisés dans chaque rangée des macroblocs.)
- Des vecteurs cinétiques pointent à l'extérieur du domaine admissible.
- La position des points du coefficient DCT reconstruit est à l'extérieur de la position 63, lorsque l'Annexe S n'est pas utilisée.
- Des valeurs de coefficient DC de chrominance sont hors du domaine normal.

Les deux premiers points sont les plus importants. La liste des points de contrôle pourrait être augmentée considérablement. Lorsqu'une telle erreur est détectée, le décodeur vidéo recherche le prochain point de synchronisation dans le flux binaire. Toutes les données comprises entre les marqueurs de synchronisation où l'erreur a été détectée (c'est-à-dire non copiées dans la mémoire de trame) sont rejetées et masquées comme décrit au § III.5.4.

III.5.4 Masquage d'erreur

L'on estime que l'algorithme suivant est efficace lorsque l'Annexe J est utilisée avec le post-filtre approprié, comme défini dans le § III.5.1. L'interopérabilité avec le post-filtre examiné au § III.5.2 est à l'étude.

Certaines parties du présent paragraphe sont fondées sur le modèle de masquage d'erreur TCON qui est décrit dans la référence LBC-95-186. Les macroblocs manquants sont masqués au moyen d'un algorithme très simple mais efficace. Dans les parties de l'image où des données sont perdues, des données issues de l'image précédemment décodée sont utilisées pour le masquage. Aucun masquage d'erreur n'est appliqué si une partie de la première image reçue est corrompue. Le vecteur cinétique du macrobloc manquant est mis à la valeur du vecteur cinétique du macrobloc situé au-dessus du macrobloc manquant. Si ce vecteur cinétique n'est pas disponible, le second vecteur est mis à *zéro*. Au moyen du vecteur cinétique approprié, le macrobloc issu de l'image précédente à l'emplacement spatial spécifié par ce vecteur cinétique est copié dans l'image actuelle à l'emplacement du macrobloc manquant.

Cas particuliers pour 4 vecteurs MV et la compensation OBMC:

- si le macrobloc situé au-dessus du macrobloc manquant a été codé avec quatre vecteurs cinétiques (comme ce pourrait être le cas avec l'Annexe F ou J), le masquage d'erreur fonctionne de la même façon que sur la base d'un bloc 8×8 , de façon que:
 - les deux blocs 8×8 situés à gauche du macrobloc manquant utilisent le vecteur cinétique du bloc 8×8 situé en bas à gauche du macrobloc supérieur,
 - et de même que les deux blocs 8×8 situés à droite du macrobloc manquant utilisent le vecteur cinétique du bloc 8×8 situé en bas à droite du macrobloc supérieur.
- Si l'Annexe F est utilisée, la pondération de compensation OBMC est appliquée au moyen des vecteurs cinétiques de masquage.

III.6 Acquisition des informations

Le présent paragraphe contient toutes les informations qui ont été approuvées officiellement par le groupe d'experts en codage vidéo de l'UIT-T pour adoption dans le modèle d'essai mais qui ne s'intègrent pas dans les scénarios applicatifs définis au § III.2. La plupart de ces informations ne s'appliquent qu'aux situations dans lesquelles certains modes opérationnels sont utilisés, comme le mode d'échelonnabilité (Annexe O), les modes de codage par trames PB (Annexes G et M), ou le mode de rafraîchissement à résolution réduite (Annexe Q). Les informations contenues dans le présent paragraphe ne sont pas toutes vérifiées de manière approfondie quant à leur compatibilité avec les algorithmes "courants" qui sont utilisés pour les scénarios applicatifs. Des informations complémentaires sont sollicitées.

III.6.1 Mode de codage par trames PB (Annexes G et M)

L'Annexe M offre de meilleures options de prédiction que l'Annexe G. Il convient de toujours l'employer à la place de l'Annexe G lorsque l'on applique la version 2 de la Rec. UIT-T H.263.

III.6.1.1 Estimation de mouvement et décision de mode améliorées par trames PB améliorées

Les vecteurs cinétiques anticipés et différés pour chacun des blocs contenus dans le macrobloc B sont obtenus par échelonnement du meilleur vecteur cinétique issu du macrobloc P, MV, comme spécifié dans la Rec. UIT-T H.263. Afin de trouver la somme SADbidir, ces vecteurs servent à effectuer une prédiction dans les deux sens, comme décrit dans l'Annexe M/H.263 (codage par trames PB amélioré) mais avec mise à zéro de MVD.

L'on effectue ensuite, pour le macrobloc B 16×16 , une estimation de mouvement normale aux pixels entiers et aux demi-pixels, par rapport à la précédente image P reconstruite. La meilleure somme SADforw pour cette estimation de mouvement est comparée à la somme SADbidir pour la prédiction dans les deux sens. Si ($SADforw < SADbidir - 100$), la prédiction anticipée est choisie pour ce macrobloc. Dans ce cas, le vecteur cinétique anticipé qui a été trouvé dans l'estimation de mouvement ci-dessus est transmis directement dans MVDB, sans prédiction de vecteur cinétique. Si la prédiction dans les deux sens se trouve être la meilleure, aucun champ MVDB n'est transmis.

III.6.2 Mode d'échelonnabilité (Annexe O)

III.6.2.1 Estimation de mouvement et décision de mode par trames B vraies

Les trames B offrent plusieurs options de prédiction et donc de modes de codage possibles. L'estimation de mouvement anticipée est effectuée par rapport à la précédente trame I/P reconstruite et l'estimation de mouvement différée est effectuée par rapport à la prochaine trame I/P reconstruite.

Par ailleurs, les trames B permettent une prédiction de mode direct, qui n'exige pas la transmission de données de vecteur cinétique, ainsi qu'une prédiction de mode dans les deux sens, qui fait appel aux vecteurs cinétiques résultant de l'estimation de mouvement anticipée ou différée. Pour la prédiction de mode direct, c'est le vecteur cinétique situé entre les trames de paire I/P ou P/P environnantes qui est échelonné.

Les modes de codage préférés sont, dans l'ordre: direct, anticipé, différé et bilatéral. Pour refléter ces préférences, la somme SAD est calculée pour chaque mode de prédiction possible puis ajustée. Le mode direct est sélectionné par soustraction de 100 de sa somme SAD. Le mode anticipé est sélectionné par soustraction de 50 de sa somme SAD. La somme SAD du mode différé n'est pas modifiée. La somme SAD du mode bilatéral est pénalisée par addition de 75. Le mode possédant la plus petite somme SAD après ces modifications est choisi comme mode de codage.

III.6.2.2 Estimation de mouvement et décision de mode par trames EI/EP

Les trames EI – et EP – offrent plusieurs options de prédiction et donc de modes de codage possibles. L'estimation de mouvement anticipée est effectuée par rapport à la précédente trame EI/EP reconstruite.

Par ailleurs, les trames EI et EP permettent une prédiction de mode ascendant, qui n'exige pas la transmission de données de vecteur cinétique, ainsi qu'une prédiction de mode dans les deux sens, qui fait appel aux vecteurs cinétiques résultant de l'estimation de mouvement anticipée. Pour la prédiction de mode ascendant, la prédiction est faite à partir du même emplacement spatial de la trame reconstruite et éventuellement suréchantillonnée dans la couche de référence.

Les modes de codage préférés sont, dans l'ordre: ascendant, anticipé et bilatéral. Pour refléter ces préférences, la somme SAD est calculée pour chaque mode de prédiction possible puis ajustée. Le mode ascendant est sélectionné par soustraction de 50 de sa somme SAD. La somme SAD du mode anticipé n'est pas modifiée. La somme SAD du mode bilatéral est pénalisée par addition de 100. Le mode possédant la plus petite somme SAD après ces modifications est choisi comme mode de codage.

III.6.2.3 Commande de débit pour trames P et B

III.6.2.3.1 Niveau des macroblocs

La commande de débit au niveau des macroblocs, décrite au § III.4.2.2, peut être utilisée directement pour les trames B. La seule différence est que, comme les statistiques des trames B sont différentes de celles des trames P, les paramètres de commande de débit K et C (qui sont actualisés à chaque macrobloc) prennent des valeurs comprises dans des étendues différentes. Par conséquent, lorsque l'on utilise des trames P et B, différents paramètres doivent être utilisés, respectivement $\{K_P, C_P\}$ et $\{K_B, C_B\}$ pour les trames P et B.

III.6.2.3.2 Niveau des trames

Dans le § III.4.2.1, la commande de débit au niveau des trames attribue un nombre cible presque constant de bits par trame P (après la première trame I), ce qui est une stratégie efficace pour les communications vidéo à faible délai. Mais dans les scénarios où une ou plusieurs trames B sont insérées entre les trames P, un mécanisme est nécessaire pour attribuer un moins grand nombre de bits aux trames B car celles-ci sont plus faciles à coder.

Le présent paragraphe décrit une technique appropriée à l'attribution d'un nombre cible de bits aux trames P et B. Les calculs de cette méthode sont expliqués dans [Q15-C-19][Q15-D-22]. Ces calculs sont fondés sur le cas typique où la structure des trames est la suivante:

$$I, B, \dots, B, P, B, \dots, B, P, B, \dots, B, P, B, \dots, B, P, \dots$$

La séquence des trames "B, ..., B,P" est répétée périodiquement après la première trame I et est considérée comme un groupe d'images (GOP). Soit M_B le nombre de trames B dans un groupe GOP. Le nombre cible de bits pour l'image P dans ce GOP, T_P , et la cible pour chacune des trames B, T_B , peuvent être calculés comme suit:

$$T_P = T - M_B T_B$$
$$T_B = \frac{T - 16^2 N(C_P - \beta C_B)}{\beta + M_B}$$
$$\beta = 0,9 \beta_{PREV} + 0,1F \frac{E_P}{E_C}$$

les paramètres des équations ci-dessus étant définis ainsi:

- T , M , et N sont, respectivement, le nombre de bits pour le GOP, le nombre de trames pour le GOP, et le nombre de macroblocs dans une trame.
- La valeur de β détermine le nombre de bits assignés à la trame P et le nombre de bits qui sont assignés aux trames B. β augmente avec F et E_P/E_B , ce qui est décrit ci-dessous.
- F détermine la grandeur du rapport PSNR des trames P par rapport à celui des trames B. Par exemple, si F est égal à 1, le PSNR des deux types de trames sera similaire et si F est plus grand que 1 le PSNR du P augmente par rapport à celui des trames B. La formule suivante est utilisée pour déterminer la valeur de F :

$$F = \max \left\{ \min \left(\frac{1,4}{\sqrt{\text{Bpp}}} - 0,3, 5 \right), 1 \right\}$$

où Bpp est le débit en bits par pixel pour la séquence vidéo. L'application de cette relation montre que le PSNR des trames P est en moyenne supérieur d'environ 1 dB à celui des trames B, ce qui semble être un compromis raisonnable.

- E_P est l'énergie pour la trame P dans le précédent GOP, où l'énergie est définie comme la somme des variances erreurs de prédiction du macrobloc, soit:

$$E_P = \sum_{i=1}^N \sigma_i^2$$

où σ_i^2 est la variance du i ème macrobloc dans la (précédente) trame P, comme défini au § III.4.2.2. D'autre part, E_B est la moyenne des énergies pour les trames B dans le précédent GOP, soit:

$$E_B = \frac{1}{M_B} \sum_{m=1}^{M_B} E_{B,m}$$

où $E_{B,m}$ est l'énergie de la même trame B dans le précédent GOP.

- β_{PREV} est mis à F pour le premier GOP et à la précédente valeur de β pour les GOP suivants.
- C_P et C_B sont respectivement les débits cinétique et syntaxique (en bits par pixel) pour les trames P et B. Leurs valeurs sont obtenues d'après la commande de débit au niveau des macroblocs correspondants, comme décrit au § III.4.2.2.

La précédente commande de débit au niveau des trames, décrite au § III.4.2.1, a été conçue pour les groupes GOP de type "P...P". Elle correspond au cas particulier où $E_P = E_B$, $F = 1$, (ou, ce qui est équivalent, $\beta = 1$) et $C_P = C_B$ dans les équations ci-dessus.

Finalement, avant qu'une trame donnée soit codée (avec une cible de T_P ou T_B bits), la valeur Δ est soustraite du débit binaire cible, comme défini au § III.4.2.1, ce qui permet de réagir à la saturation du tampon du codeur et au seuil de saut de trame. Ce seuil a été réglé sur le débit binaire du canal (en bits par seconde) divisé par la fréquence de trame de codage, ce qui est un bon choix pour les scénarios à faible délai. Mais lorsque des trames B sont insérées entre des trames P et que le délai est donc moins important, un seuil de saut de trame plus grand serait plus approprié (ainsi qu'une plus grande valeur de A au § III.4.2.1).

III.6.2.3.3 Rapport SNR et commande de débit dans la couche d'amélioration spatiale

Normalement, le débit binaire offert à chacune des couches d'amélioration est déterminé par l'application spécifique. Dans chaque couche, la commande de débit peut être utilisée de manière équivalente à son utilisation dans la couche de base, avec le débit binaire, la fréquence de trame et la structure de groupe GOP pour chaque couche considérée. La seule différence est qu'il faut des variables différentes pour K et C dans chaque couche car différentes trames ont des statistiques différentes dans chaque couche. Plus précisément, les paramètres K et C devraient être différents pour les différentes couches et, à l'intérieur d'une couche donnée, les paramètres K et C devraient être différents selon chaque type de trame.

III.6.2.4 Utilisation en environnements exposés aux erreurs

En plus des autres méthodes de tolérance aux erreurs présentées dans le présent appendice, les numéros d'image définis dans l'Annexe U (sélection d'image de référence améliorée) et dans l'Annexe W (informations complémentaires d'amélioration) fournissent une méthode de limitation des erreurs appropriée à l'Annexe O. Les numéros d'image font partie intégrante du flux binaire lorsque l'Annexe U est utilisée. Sinon, il est recommandé d'utiliser les numéros d'image tels que définis dans l'Annexe W.

La définition des numéros d'image dans l'Annexe U et dans l'Annexe W peut être résumée comme suit: le numéro d'image doit être incrémenté d'une unité à chaque image I, P, PB, EI et EP codée et transmise, par rapport à l'image déjà mémorisée dans la même couche d'amélioration. Pour les images B, le numéro d'image doit être incrémenté par rapport à la valeur contenue dans l'image non B la plus récemment mémorisée dans la couche de référence de l'image B qui précède l'image B dans l'ordre du flux binaire (image qui est temporairement placée à la suite de l'image B). Si des images adjacentes possèdent la même référence temporelle dans la même couche d'amélioration et si un mode de sélection d'image de référence est utilisé, le décodeur doit considérer cette occurrence comme une indication du fait que des copies redondantes ont été envoyées avec à peu près le même contenu de scène graphique. Toutes ces images doivent partager le même numéro d'image.

Si la différence entre numéros d'image de deux images reçues et mémorisées consécutivement dans la même couche d'amélioration n'est pas égale à l'unité et ces images ne représentent pas approximativement le même contenu de scène graphique comme décrit ci-dessus, il y a lieu que le décodeur en déduise une perte d'images ou une corruption de données. Dans un tel cas, un message sur voie de retour, tel que défini dans l'Annexe U ou une indication similaire de perte d'images peut être envoyé au codeur distant. Si les images perdues se trouvaient dans une des couches d'amélioration, le décodeur peut également effectuer le décodage de la couche qui a subi la perte d'images et celui de toute couche supérieure, surtout si une image de référence redondante appropriée est signalée (voir W.6.3.13/H.263).

III.6.3 Mode de rafraîchissement à résolution réduite (Annexe Q)

Ce paragraphe décrit la façon d'utiliser le mode de rafraîchissement à résolution réduite qui est défini dans l'Annexe Q. Les informations indiquées n'ont pas été vérifiées.

III.6.3.1 Estimation de mouvement et sélection de mode

Dans le mode de rafraîchissement à résolution réduite, l'estimation du mouvement est effectuée sur le macrobloc de luminance 32×32 au lieu du macrobloc 16×16 . La somme SAD (somme des différences absolues) sert de mesure d'erreur. Dans ce mode, chaque composante du vecteur cinétique ou des quatre vecteurs cinétiques du macrobloc est limitée à une valeur de demi-entier ou de zéro afin d'élargir l'étendue de recherche avec la même table MVD.

III.6.3.2 Estimation de mouvement en mode de base (sans options)

III.6.3.2.1 Estimation de mouvement à pixels entiers

La recherche est effectuée avec un déplacement par pixels entiers dans la composante Y. Les comparaisons sont faites entre le macrobloc entrant et le macrobloc déplacé dans la précédente image reconstruite. Une recherche complète est effectuée, dont l'étendue est inférieure ou égale à ± 30 pixels dans le sens horizontal et dans le sens vertical autour de la position initiale du macrobloc.

$$SAD(x, y) = \sum_{i=1, j=1}^{32, 32} |original - decoded_previous|_{x, y} \text{ jusqu'à } \pm 30$$

Pour le vecteur zéro, la somme $SAD(0,0)$ est réduite de **400** afin de sélectionner le vecteur zéro lorsqu'il n'y a pas de différence significative.

$$SAD(0,0) = SAD(0,0) - 400$$

La paire (x, y) qui produit la plus petite somme SAD est choisie comme vecteur cinétique à pixels entiers, MV0. La somme SAD correspondante est $SAD(x, y)$.

III.6.3.2.2 Décision de mode INTRA/INTER

Après l'estimation de mouvement à pixels entiers, le codeur prend une décision quant à l'emploi de la prédiction INTRA ou INTER dans le codage. Les paramètres suivants sont calculés afin de prendre la décision INTRA/INTER:

$$MB_moyen = \frac{\sum_{i=1, j=1}^{32, 32} original}{1024}$$
$$A = \sum_{i=1, j=1}^{32, 32} |original - MB_moyen|$$

Le mode INTRA est choisi si: $A < (SAD(x, y) - 2000)$

Remarquer que si la somme $SAD(0,0)$ est utilisée, il s'agit de la valeur qui a déjà été réduite de 400 ci-dessus.

Si le mode INTRA est choisi, aucune autre opération n'est nécessaire pour la recherche cinétique. Si le mode INTER est choisi, la recherche cinétique se poursuit à demi-pixels autour de la position MV0.

III.6.3.2.3 Recherche à demi-pixels

La recherche à demi-pixels est effectuée sur la composante Y du macrobloc au moyen de la précédente trame reconstruite. La zone de recherche est de ± 1 demi-pixel autour de la matrice cible 32×32 pointée par le vecteur MV0, à condition que chaque composante du vecteur candidat pour la recherche à demi-pixels soit une moitié d'entier ou zéro. Pour le vecteur zéro $(0,0)$, la somme $SAD(0,0)$ est réduite de 400 comme pour la recherche à pixels entiers.

Les valeurs à demi-pixels sont calculées comme décrit dans le § 6.1.2/H.263.

Le vecteur produisant la meilleure correspondance au cours de la recherche à demi-pixels est nommé MV. Il est formé de composantes horizontales et verticales (MV_x, MV_y) , mesurées chacune en demi-pixels.

III.6.3.3 Estimation de mouvement en mode de prédiction avancé (Annexe F)

Le présent paragraphe ne s'applique que si le mode de prédiction avancé est choisi.

III.6.3.3.1 Estimation de mouvement à pixels entiers

La recherche à ± 2 pixels entiers dans une région $[-31, 30]$ est effectuée pour les **blocs 16×16** autour du vecteur 32×32 à pixels entiers.

III.6.3.3.2 Recherche à demi-pixels

La recherche à demi-pixels est effectuée pour chacun des blocs entourant le vecteur 16×16 pixels entiers. La zone de recherche est de $\pm 0,5$ pixel autour du vecteur 16×16 pixels entiers du bloc correspondant, en respectant la condition que chaque composante du vecteur candidat pour la recherche à demi-pixels soit une moitié d'entier ou zéro dans une région $[-31,5, 30,5]$.

III.6.3.3.3 Décision par un vecteur MV sur quatre en prédiction avancée

Le présent paragraphe ne s'applique que si le mode de prédiction avancé est choisi.

La somme SAD pour le meilleur vecteur à demi-pixels d'un macrobloc **32×32 MB** (y compris la soustraction de **400** si le vecteur est $(0,0)$):

$$SAD_{32}(x, y)$$

Somme SAD pour les meilleurs vecteurs à demi-pixels d'un **bloc 16×16** de macrobloc entier:

$$SAD_{4 \times 16} = \sum_1^4 SAD_{16}(x, y)$$

La règle suivante s'applique:

si: $SAD_{4 \times 16} < SAD_{32} - 800$, choisir la prédiction par **16×16 bloc**

sinon: choisir la prédiction par macrobloc **32×32 MB**.

III.6.3.4 Estimation de mouvement dans le mode de codage par vecteurs cinétiques non restreints (Annexe D)

Le présent paragraphe n'est applicable que si l'étendue élargie de vecteurs cinétiques est choisie dans le mode UMV.

III.6.3.4.1 Limitation de la fenêtre de recherche

Etant donné que, dans ce mode, la fenêtre de vecteurs cinétiques corrects est centrée sur le prédicteur de vecteurs cinétiques pour le macrobloc actuel, quelques restrictions sont appliquées à la recherche de vecteurs cinétiques à pixels entiers afin de s'assurer que les vecteurs cinétiques trouvés seront transmissibles.

Avec ces restrictions, les deux vecteurs **32×32 MB** et **bloc 16×16** trouvés par la procédure décrite ci-dessous seront transmissibles, quel que finisse par devenir le prédicteur de vecteurs cinétiques à la précision des demi-pixels pour le macrobloc ou pour chacun des quatre blocs.

III.6.3.4.2 Recherche à pixels entiers

L'on trouve d'abord le prédicteur de vecteurs cinétiques pour le vecteur **32×32 MB** fondé seulement sur les vecteurs cinétiques à pixels entiers. La recherche sur macrobloc **32×32 MB** est alors centrée sur le prédicteur tronqué, avec une fenêtre de recherche partiellement limitée. Si la recherche s'effectue sur quatre vecteurs, la fenêtre de recherche **32×32 MB** est limitée à l'étendue **$29 - (2 * 16 \times 16_block_search_window + 1)$** . Etant donné que, dans ce modèle, la fenêtre de recherche d'un **bloc $16 \times 16_block$** est **2,5**, la fenêtre de recherche par défaut d'un macrobloc **32×32 MB** dans le mode UMV se révèle égale à ± 23 positions de pixels entiers. Les recherches sur **bloc $16 \times 16_block$** sont donc centrées sur le meilleur vecteur **32×32 MB** et la recherche à ± 2 pixels est effectuée dans chaque **bloc $16 \times 16_block$** .

III.6.3.4.3 Recherche à demi-pixels

Les recherches à demi-pixels sont effectuées comme dans les autres modes. La zone de recherche est de $\pm 0,5$ pixel autour du meilleur vecteur de pixels entiers du macrobloc/bloc correspondant, à condition que chaque composante du vecteur candidat à la recherche à demi-pixels soit une valeur de demi-entier ou de zéro.

III.6.3.5 Sous-échantillonnage de l'erreur de prédiction

Après compensation cinétique sur blocs 16×16 , chaque bloc d'erreur de prédiction 16×16 est sous-échantillonné jusqu'au bloc d'erreur de prédiction 8×8 à résolution réduite. Afin de simplifier l'implémentation, le filtrage est limité à un bloc permettant un suréchantillonnage sur blocs individuels. La Figure III.5 montre le positionnement des échantillons. La procédure de sous-échantillonnage est définie à la Figure III.6 pour les pixels de luminance et de chrominance. Le filtrage est effectué sans tenir compte des limites du bloc. Dans la Figure III.6, le caractère "/" indique une division par troncature.

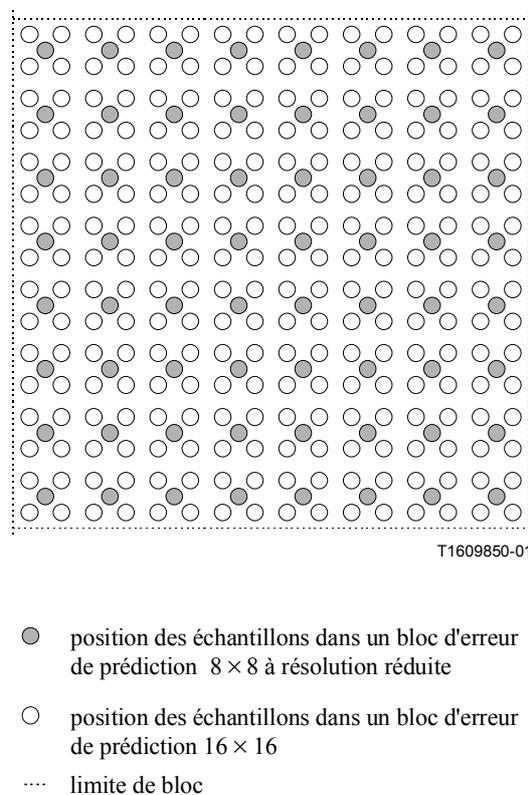


Figure III.5/H.263 – Positionnement des échantillons dans un bloc d'erreur de prédiction 8×8 à résolution réduite et dans un bloc d'erreur de prédiction 16×16

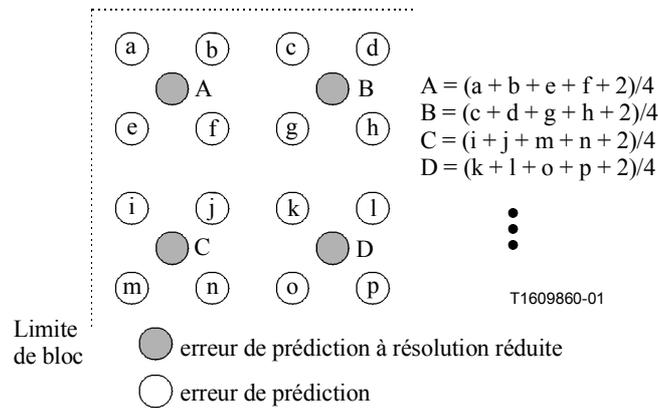


Figure III.6/H.263 – Création d'un bloc d'erreur de prédiction à résolution réduite pour pixels internes

III.6.3.6 Transformation et quantification

Une transformée discrète en cosinus (DCT, *discrete cosine transform*) à 2 dimensions indépendantes est appliquée aux blocs 8×8 d'erreur de prédiction à résolution réduite comme dans le mode par défaut. La quantification est ensuite effectuée comme décrit dans le mode par défaut.

III.6.3.7 Commutation de mode

Le présent paragraphe décrit un algorithme simple de commutation de mode pour l'Annexe Q.

NOTE – Cet algorithme pourrait s'appliquer, au prix d'une légère modification, à la partie "facteur de 4" de l'Annexe P.

III.6.3.7.1 Algorithme de décision relative à la résolution

Afin de déterminer la résolution pour l'Annexe Q, un simple algorithme décisionnel est utilisé sur la base des termes \overline{QP}_{i-1} et B_{i-1} .

\overline{QP}_{i-1} le paramètre QP moyen de la précédente trame codée;

B_{i-1} le nombre de bits utilisés dans la précédente trame codée.

En admettant que la relation entre \overline{QP}_{i-1} et B_{i-1} est proche de la proportion inverse, le produit de \overline{QP}_{i-1} et B_{i-1} peut être considéré comme un indice de la complexité approchée de la trame codée.

La commutation vers le mode de rafraîchissement à résolution réduite est effectuée si la trame actuelle est en mode par défaut et si le produit de \overline{QP}_{i-1} et B_{i-1} est plus grand que le seuil TH1.

Si la trame actuelle est en mode de rafraîchissement à résolution réduite, la commutation vers le mode par défaut est effectuée si ce produit de \overline{QP}_{i-1} et B_{i-1} est plus petit que le seuil TH2.

Commutation du **mode par défaut** au **mode de rafraîchissement à résolution réduite**

si($\overline{QP}_{i-1} * B_{i-1} > TH1$) {
commuter sur
mode de rafraîchissement à résolution réduite;
}

Commutation du **mode de rafraîchissement à résolution réduite** au **mode par défaut**

$$\text{si}(\overline{QP}_{i-1} * B_{i-1} < TH2)\{$$

*commuter sur **mode par défaut**;*

$$\}$$

Le seuil TH1 est déterminé par l'équation suivante, dans laquelle QP1 et FR1 représentent la qualité subjective la plus basse dont le codage soit autorisé dans le mode par défaut.

$$TH1 = QP1 * (\text{Target_Bit rate} / FR1)$$

De même, le seuil TH2 est déterminé par l'équation suivante, dans laquelle QP2 et FR2 représentent la qualité subjective la plus basse dont le codage soit autorisé dans le mode par défaut.

$$TH2 = QP2 * (\text{Target_Bit rate} / FR2)$$

Les valeurs de QP1, FR1, QP2 et FR2 peuvent dépendre du format de source, de la fréquence de trame cible et du débit binaire cible. Si le format de source indique CIF, la fréquence de trame cible est 10 fps et le débit binaire cible est 48 kbit/s, soit respectivement QP1 = 16, FR1 = 7, QP2 = 7, FR2 = 9.

III.6.3.8 Limitation des coefficients DCT lors d'une commutation de résolution réduite à résolution normale

Une fois que le mode de résolution réduite est choisi, le détail de l'image est susceptible d'être perdu. Si le mode revient au mode par défaut, le détail de l'image doit être reproduit, ce qui consomme une grande quantité de bits. Cette soudaine augmentation des bits de codage provoque souvent des sauts de trame intempestifs. Par ailleurs, comme l'algorithme de décision relative à la résolution décrit ci-dessus utilise le produit du QP moyen et du nombre de bits, cette soudaine augmentation binaire provoque le retour au mode de rafraîchissement à résolution réduite et une oscillation entre les deux modes se produit souvent. Afin d'éviter cette dégradation, une limitation des coefficients DCT à envoyer est introduite dans les quelques trames qui suivent la commutation du mode de rafraîchissement à résolution réduite au mode par défaut. Dans la première trame après la commutation vers le mode par défaut, seuls les coefficients contenus dans une région 4×4 à basse fréquence peuvent être envoyés, puis de même les coefficients 5×5 dans la deuxième trame, 6×6 dans la troisième et 7×7 dans la quatrième. Cet algorithme "d'atterrissage en douceur" peut supprimer efficacement le saut de trame intempestif et l'oscillation des modes.

III.6.3.8.1 Commande de débit

La commande de débit est identique au mode par défaut, sauf que le nombre de macroblocs est réduit d'un quart.

III.6.4 Recherche rapide au moyen d'inégalités mathématiques

L'on connaît plusieurs méthodes d'accélération de la recherche cinétique, sur la base d'inégalités mathématiques [LS95] et [LT97]. Ces inégalités, par exemple l'inégalité du triangle, placent une borne inférieure à la norme des différences entre vecteurs. Lors d'une correspondance de bloc, les critères de recherche très souvent utilisés pour la distorsion sont la somme des différences absolues (SAD) ou la somme des différences au carré (SSD) entre la prédiction compensée cinétiquement $c[x, y]$ et le signal original $s[x, y]$. L'insertion de l'inégalité triangulaire dans les sommes SAD et SSD donne:

$$D(s, c) = \sum_{[x, y] \in B} |s[x, y] - c[x, y]|^p \geq \hat{D}(s, c) = \left| \left(\sum_{[x, y] \in B} |s[x, y]|^p \right)^{1/p} - \left(\sum_{[x, y] \in B} |c[x, y]|^p \right)^{1/p} \right|^p \quad (\text{III-6.4.1})$$

en faisant varier le paramètre $p = 1$ pour SAD et $p = 2$ pour SSD. Noter que pour $p = 2$, l'inégalité utilisée dans [LT97] diffère de l'équation (III-6.4.1) ci-dessus. Empiriquement, peu de différence a été trouvée entre ces inégalités. Pour certains blocs, l'inégalité utilisée dans [LT97] fournit une borne plus précise tandis que pour d'autres blocs l'inégalité triangulaire donne de meilleurs résultats. L'ensemble B contient les positions d'échantillonnage des blocs considérés, par exemple d'un bloc de 16×16 échantillons.

Supposons que D_{min} soit la plus petite valeur de distorsion calculée précédemment dans la recherche cinétique sur blocs. La distorsion $D(s, c)$ d'un autre bloc c dans notre étendue de recherche est assurée de dépasser D_{min} si la borne inférieure de $D(s, c)$ dépasse D_{min} . Plus précisément, le bloc c sera rejeté si:

$$\hat{D}(s, c) \geq D_{min} \quad (\text{III-6.4.2})$$

La structure spéciale du problème de l'estimation de mouvement permet d'établir une méthode rapide de calcul des valeurs de norme de tous les blocs $c[x, y]$ dans les trames précédemment décodées [LS95]. L'extension à un critère d'estimation de mouvement contraint en débit est simple [CM97].

III.6.4.1 Ordre de recherche

Il est évident qu'une petite valeur de D_{min} , déterminée au début de la recherche, conduit au rejet ultérieur d'un grand nombre de blocs et réduit donc les calculs. L'ordre dans lequel les blocs sont vérifiés dans l'étendue de recherche a donc une grande influence sur la durée de calcul. Par exemple, si les tables de codes de Huffman des vecteurs cinétiques sont fournies comme informations préalables au sujet de l'espace de recherche, l'ordre d'exploration devra suivre les débits croissants de ces vecteurs cinétiques. Cela augmentera la probabilité de trouver une bonne correspondance dès le début du processus de recherche. Une bonne approximation de ces probabilités est la recherche en spirale décrite au § III.3.1.3.1.

III.6.4.2 Inégalités triangulaires multiples

Conformément à [WZG99], des inégalités triangulaires multiples peuvent être employées. Soit une partition de l'ensemble B en sous-ensembles B_n de façon que:

$$B = \bigcup_n B_n \text{ et } \bigcap_n B_n = \phi \quad (\text{III-6.4.3})$$

L'inégalité triangulaire (III-6.4.1) vaut pour tous les sous-ensembles B_n possibles. Le remaniement de la formule pour $D(s, c)$ donne:

$$\sum_{[x,y] \in B} |s[x,y] - c[x,y]|^p = \sum_n \sum_{[x,y] \in B_n} |s[x,y] - c[x,y]|^p \quad (\text{III-6.4.4})$$

et l'application de l'inégalité triangulaire à tous les sous-ensembles B_n donne:

$$D(s, c) = \sum_{[x,y] \in B} |s[x,y] - c[x,y]|^p \geq \sum_n \left(\left(\sum_{[x,y] \in B_n} |s[x,y]|^p \right)^{1/p} - \left(\sum_{[x,y] \in B_n} |c[x,y]|^p \right)^{1/p} \right)^p \quad (\text{III-6.4.5})$$

Noter que l'équation (III-6.4.5) est une borne inférieure plus stricte que l'équation (III-6.4.1) mais qui exige plus de calculs, ce qui constitue donc le compromis entre la précision de la borne inférieure et la complexité des calculs.

Une question importante reste à ce propos le choix des partitions B_n . L'équation (III-6.4.5) vaut naturellement pour tous les sous-ensembles possibles qui satisfont la relation (III-6.4.4). Mais comme les valeurs de norme de tous les blocs de l'espace de recherche doivent être précalculées, la méthode rapide qui est décrite en [LS95] peut être appliquée. Une subdivision aléatoire de l'ensemble B en n sous-ensembles arbitraires pourrait donc ne pas être le meilleur choix. En revanche, une subdivision symétrique de B serait sans doute préférable en termes de calculs. Dans la référence [LT97], il est proposé de diviser un bloc carré 16×16 en deux éléments distincts. La première subdivision produit 16 sous-ensembles B_n dont chacun est une des 16 lignes contenant 16 échantillons. La deuxième subdivision produit 16 sous-ensembles B_n , chacun étant une des 16 colonnes contenant 16 échantillons.

Comme la H.263 permet des blocs 8×8 , la méthode proposée en [LC95] est appliquée, dans laquelle un bloc 16×16 est décomposé en sous-blocs. Le bloc 16×16 est subdivisé de façon à passer d'un ensemble de 16×16 échantillons à 4 sous-ensembles de 8×8 échantillons. Les diverses inégalités triangulaires (de sous-ensemble) sont successivement appliquées dans l'ordre des instants de calcul afin d'évaluer chacune d'elles, c'est-à-dire que l'inégalité triangulaire relative au bloc 16×16 est vérifiée en premier, puis l'on vérifie les inégalités relatives aux blocs de taille 8×8 . Au niveau des blocs 8×8 , seule l'inégalité triangulaire 8×8 est vérifiée.

III.6.5 Commande de la fréquence de trame de codage

Le présent paragraphe contient deux algorithmes qui peuvent être utilisés à la place de l'algorithme de commande de débit dans la couche des trames, défini au § III.4.2.1. Alors que l'algorithme de fréquence de trame fixe n'est utile que dans les environnements où une fréquence de trame absolument fixe est plus importante que la qualité perceptive et que le délai, l'algorithme de commande adaptative augmente la qualité perceptive tout en conservant des caractéristiques de délai raisonnables en ajustant la fréquence de trame en fonction du contenu scénographique.

III.6.5.1 Commande adaptative de la fréquence de trame de codage

L'objectif de cet algorithme de commande de la fréquence de trame de codage est de conserver la qualité des trames P dans un domaine tolérable en cas de variation cinétique brusque et d'environnements avec canaux de communication variant dans le temps sans dégradation évidente de la continuité cinétique perçue. L'algorithme ajuste la fréquence de trame de codage de manière adaptative en fonction des informations cinétiques contenues dans les données vidéo sous-jacentes afin de maintenir dans un domaine tolérable la qualité d'image de chaque trame P. Comme il est difficile d'assurer une bonne qualité en résolution aussi bien spatiale que temporelle aux très bas débits binaires (en termes de continuité cinétique), une commande adaptative de la fréquence de trame de codage est adoptée en tant que compromis entre la qualité spatiale/temporelle fondée sur les

informations cinétiques vidéo et la largeur de bande disponible dans les canaux. Il convient de remarquer que l'œil humain est sensible aux variations brusque de la fréquence (ou de l'intervalle) des trames de codage. Le présent procédé vise à réduire la dégradation temporelle en termes de mouvements saccadés perçus par des êtres humains. En même temps, un faible délai de codage est imposé pour le traitement en temps réel. La commande de la fréquence de trame de codage en présence de débit CBR variant dans le temps et la relation avec l'algorithme de commande de débit dit TMN8 du § III.4.2 sont expliquées ci-après. Le développement de cet algorithme est décrit dans la référence [Q15-G-22].

Les définitions suivantes sont utilisées:

i : indice de trame codée.

a, b, a' et b' : coefficients du modèle R-D de couche trames.

f_{i-1} : trame de référence reconstruite à l'instant actuel.

f_i : image non comprimée à l'instant actuel.

\overline{QP}_i : QP moyen de tous les macrobloks dans une trame.

$\hat{B}(Q), \hat{D}(Q)$: modèles de débit et de distorsion dans une trame, respectivement.

$MAD(f_{i-1}, f_i)$: moyenne des différences absolues entre f_{i-1} et f_i .

R_i : largeur de canal actuellement disponible.

F_i : intervalle actuel des trames de codage dans l'hypothèse où la caméra capture les trames à une fréquence de G fps.

Etape 0: choisir les paramètres initiaux de façon que la fréquence de trame de codage initiale soit F_0 , et que le seuil d'ajustement de l'intervalle des trames de codage soit $c = 0,04$. De même, les dimensions de la table R-D de la couche des trames, T_0 et T_{max} sont choisies respectivement à 0 et 20. Finalement, l'itération commence avec $i = 1$.

Etape 1: décaler la table de la base de données R-D en posant: $B_k = B_{k+1}, D_k = D_{k+1}, \overline{QP}_k = \overline{QP}_{k+1}$ pour $k = 1, \dots, T_{i-1} - 1$.

Ajouter ensuite une nouvelle entrée à la fin de la table selon: $B_{T_{i-1}} = B_{i-1}, D_{T_{i-1}} = D_{i-1}, \overline{QP}_{T_{i-1}} = \overline{QP}_{i-1}$.

Etape 2: calculer les modèles de débit et de distorsion par rapport au paramètre QP moyen des trames. Calculer d'abord les coefficients du modèle et débit et distorsion au moyen de la table R-D ci-dessus.

$$\begin{aligned}\hat{B}(QP_i) &= (a \cdot QP_i^{-1} + b \cdot QP_i^{-2}) \cdot MAD(f_{i-1}, f_i), \\ \hat{D}(QP_i) &= a' QP_i + b'.\end{aligned}\tag{III-6.5.1}$$

$$a = \frac{\sum_{k=1}^{T_{i-1}} \{R_k QP_k^{-1} - b QP_k^{-2}\}}{\sum_{k=1}^{T_{i-1}} QP_k^{-2}}, b = \frac{\left(\sum_{k=1}^{T_{i-1}} R_k QP_k^{-2} \right) \left(\sum_{k=1}^{T_{i-1}} QP_k^{-2} \right) - \left(\sum_{k=1}^{T_{i-1}} R_k QP_k^{-1} \right) \left(\sum_{k=1}^{T_{i-1}} QP_k^{-3} \right)}{\left(\sum_{k=1}^{T_{i-1}} QP_k^{-4} \right) \left(\sum_{k=1}^{T_{i-1}} QP_k^{-2} \right) - \left(\sum_{k=1}^{T_{i-1}} QP_k^{-3} \right)^2},\tag{III-6.5.2}$$

$$a' = \frac{\sum_{k=1}^{T_{i-1}} d_k \sum_{k=1}^{T_{i-1}} QP_k - T_{i-1} \sum_{k=1}^{T_{i-1}} d_k QP_k}{\left(\sum_{k=1}^{T_{i-1}} QP_k \right)^2 - T_{i-1} \sum_{k=1}^{T_{i-1}} QP_k^2}, b' = \frac{\sum_{k=1}^{T_{i-1}} d_k - a' \sum_{k=1}^{T_{i-1}} QP_k}{T_{i-1}}, \quad (\text{III-6.5.3})$$

Supprimer ensuite les points aberrants au moyen de la vérification suivante pour $k = 1 : T_{i-1}$. Si $|\hat{B} - B_k| > \sigma_B$ ou $|\hat{D} - D_k| > \sigma_D$ alors désactiver temporairement cette donnée et répéter l'étape 2 avec les données raffinées.

Etape 3: calculer la distorsion estimée par:

$$\hat{D} = a' \frac{a \cdot MAD(f_{i-1}, f_i) + \sqrt{(a \cdot MAD(f_{i-1}, f_i))^2 + 4b \cdot B(F_{i-1}) \cdot MAD(f_{i-1}, f_i)}}{2B(F_{i-1})} + b', \quad (\text{III-6.5.4})$$

$$B(F_{i-1}) = \frac{F_{i-1}}{G} \cdot R_i.$$

Etape 4: pour déterminer l'intervalle des trames de codage, calculer d'abord les valeurs de seuil selon: $TH_{d1} = (1 + c) \cdot D_{avg}$, $TH_{d2} = (1 - c/2) \cdot D_{avg}$, où D_{avg} est la distorsion moyenne des 5 trames codées précédentes. Noter que, pour une valeur importante de c , la commande de la fréquence de trame de codage est désactivée et seule la commande de débit dans la couche des macroblocs fonctionne selon le § III.4.2.2. Ajuster ensuite l'intervalle des trames de codage comme indiqué dans l'équation (III-6.5.5) avec $\Delta F_{i-1} = [0,3 \cdot F_{i-1}]$.

Etape 5: appeler la commande de débit dans la couche des macroblocs du § III.4.2.2 avec le débit binaire cible \tilde{B}_i , ce qui renverra \overline{QP}_i, B_i et D_i

$$\tilde{B}_i = \begin{cases} (F_{i-1} + \Delta F_{i-1}) \cdot R_i / G & \text{si } \hat{D} > TH_{d1}, \\ (F_{i-1} + \Delta F_{i-1}) \cdot R_i / G & \text{si } \hat{D} < TH_{d2}, \\ F_{i-1} \cdot R_i / G & \text{sinon,} \end{cases} \quad (\text{III-6.5.5})$$

Etape 6: si toutes les trames sont codées, arrêter l'opération. Sinon passer à l'étape 1 avec $T_i = \min\{T_{i-1} + 1, T_{max}\}$ et $i = i + 1$.

III.6.6 Remarques sur l'utilisation optimisée du mode de sélection d'image de référence améliorée

III.6.6.1 Techniques de recherche rapide

Etant donné que l'estimation cinétique est élargie d'une à plusieurs images, des techniques de recherche rapide sont recommandées. Voir la référence [Q15-D-55] pour les détails sur la possibilité de réduire le temps de calcul.

III.6.6.2 Codage tolérant aux erreurs

Comme pour la prédiction MCP sur image unique, les pertes de contenu d'image et le masquage peuvent donner lieu à différentes images de référence dans le codeur et dans le décodeur lors d'une prédiction MCP sur images multiples, entraînant une propagation d'erreurs temporelles. Dans les deux méthodes, le codage de macroblocs en mode INTRA peut arrêter la propagation d'erreurs temporelles. Mais en MCP sur images multiples, le choix du vecteur cinétique et du paragraphe de référence d'image peut également avoir une influence notable sur la performance. Voir les détails relatifs à une stratégie de codage possible dans la référence [Q15-H-24].

III.7 Références

- [CM97] COBAN (M.), MERSEREAU (R.M.): "Computationally Efficient Exhaustive Search Algorithm for Rate-Constrained Motion Estimation" (Algorithme mathématiquement efficace de recherche exhaustive pour estimation cinétique contrainte en débit), dans *Proc. ICIP*, Santa Barbara, USA, octobre 1997.
- [GCK99] GALLANT (M.), CÔTÉ (G.), KOSENTINI (F.): "A Computation Constrained Block-based Motion Estimation Algorithm for Low Bit-rate Video Coding" (Algorithme d'estimation cinétique sur blocs avec contraintes de calcul pour codage vidéo à bas débit binaire), *IEEE Transactions on Image Processing*, décembre 1999.
- [GFS97] GIROD (B.), FÄRBER (N.), STEINBACH (E.): "Performance of the H.263 Video Compression Standard" (Performance de la norme de compression vidéo H.263), *Journal of VLSI Signal Processing: Systems for Signal, Image, and Video Technology. Special Issue on Recent Development in Video: Algorithms, Implementation and Applications*. Vol. 17, No. 2/3, p. 101-111, novembre 1997.
- [LBC-95-186] Telenor Research: Definition of an Error Concealment Model (TCON), *Contribution LBC-95-186 to meeting of ITU-T SGXV Experts Group for Very Low Bit Rate Visual Telephony (Période d'études 1993-1996)*, Boston, juin 1995.
- [LC95] LEE (C.-H.), CHEN (L.-H.): "A Fast Search Algorithm for Vector Quantization Using Mean Pyramids of Codewords" (Algorithme de recherche rapide pour quantification de vecteurs au moyen de pyramides moyennes de mots de code), dans *IEEE TR-COM*, Vol. 43, No. 2/3/4, p. 604-612, février/mars/avril 1995.
- [LS95] LI (W.), SALARI (E.): "Successive Elimination Algorithm for Motion Estimation" (Algorithme d'élimination successive pour estimation cinétique), *IEEE Trans. Image Proc.*, p.105-107, janvier 1995.
- [LT97] LIN (Y.-C.), TAI (S.-C.): "Fast Full-Search Block-Matching Algorithm for Motioncompensated Video Compression" (Algorithme de recherche complète et rapide par correspondance de blocs pour compression vidéo cinétiquement compensée), dans *IEEE TR-COM*, Vol. 45, No. 5, p. 527-531, mai 1997.
- [Q15-A-20] RIBAS-CORBERA (J.), LEI (S.): Rate Control for Low-Delay Video Communications, *Contribution Q15-A-20 to ITU-T Video Coding Experts Group Meeting (UIT-T SG16 Q.15 Période d'études 1997-2000)*, Portland, Oregon, USA juin 1997.
- [Q15-B-23] CÔTÉ (G.), GALLANT (M.), KOSENTINI (F.): Experimental Results for Integer Pixel Fast Search Motion Estimation of TMN 8, *Contribution Q15-B-23 to ITU-T Video Coding Experts Group Second Meeting (UIT-T SG16 Q.15 Période d'études 1997-2000)*, Sun River, Oregon, USA, septembre 1997.
- [Q15-C-19] RIBAS-CORBERA (J.), LEI (S.): Extension of TMN8 rate control to B frames and enhancement layer, *Contribution Q15-C-19 to ITU-T Video Coding Experts Group Third Meeting (UIT-T SG16 Q.15 Période d'études 1997-2000)*, Eibsee, Allemagne, décembre 1997.
- [Q15-D-13] WIEGAND (T.), ANDREWS (B.): An Improved H.263 Coder Using Rate-Distortion Optimization, *Contribution Q15-D-22 to ITU-T Video Coding Experts Group Fourth Meeting (UIT-T SG16 Q.15 Période d'études 1997-2000)*, Tampere, Finlande, Avril 1998.

- [Q15-D-22] RIBAS-CORBERA (J.), LEI (S.): An improvement on the extension of TMN8 rate control to B frames, *Contribution Q15-D-22 to ITU-T Video Coding Experts Group Fourth Meeting (UIT-T SG16 Q.15 Période d'études 1997-2000)*, Tampere, Finlande, Avril 1998.
- [Q15-D-55] WIEGAND (T.), GIROD (B.), LINCOLN (B.), ANDREWS (B.): Fast Search for Long-Term Memory Motion-Compensated Prediction, *Contribution Q15-D-55 to ITU-T Video Coding Experts Group Fourth Meeting (UIT-T SG16 Q.15 Période d'études 1997-2000)*, Tampere, Finlande, Avril 1998.
- [Q15-D-58] CÔTÉ (G.), EROL (B.), GALLANT (M.), KOSENTINI (F.): H.263+: Video Coding at Low Bit Rates, *Contribution Q15-D-58 to ITU-T Video Coding Experts Group Fourth Meeting (UIT-T SG16 Q.15 Période d'études 1997-2000)*, Tampere, Finlande, Avril 1998.
- [Q15-E-15] WENGER (S.), CÔTÉ (G.): Intra-Macroblock Refresh in Packet (Picture) Lossy Scenarios, *Contribution Q15-E-15 to ITU-T Video Coding Experts Group Fifth Meeting (UIT-T SG16 Q.15 Période d'études 1997-2000)*, Whistler, British Columbia, Canada, juillet 1998.
- [Q15-E-37] CÔTÉ (G.), WENGER (S.): Effects of standard-compliant macroblock intra refresh on rate-distortion performance, *Contribution Q15-E-37 to ITU-T Video Coding Experts Group Fifth Meeting (UIT-T SG16 Q.15 Période d'études 1997-2000)*, Whistler, British Columbia, Canada, juillet 1998.
- [Q15-G-22] SONG (H.), KIM (J.), KUO (C.-C.J.): Performance analysis of real-time encoding frame rate control proposal, *Contribution Q15-G-22 to ITU-T Video Coding Experts Group Seventh Meeting (UIT-T SG16 Q.15 Période d'études 1997-2000)*, Monterey, Californie, USA, février 1999.
- [Q15-H-24] WIEGAND (T.), FÄRBER (N.), GIROD (B.): Error-Resilient Video Transmission Using Long-Term Memory Motion Compensation, *Contribution Q15-H-24 to ITU-T Video Coding Experts Group Eighth Meeting (UIT-T SG16 Q.15 Période d'études 1997-2000)*, Berlin, Allemagne, août 1999.
- [Q15-I-26] HANNUKSELA (M.), LEDISCHKE (M.), ZHANG (J.): Results from Error Resilient Header Repetition Core Experiment, *Contribution Q15-I-26 to ITU-T Video Coding Experts Group Ninth Meeting (UIT-T SG16 Q.15 Période d'études 1997-2000)*, Red Bank, New Jersey, USA, octobre 1999.
- [Q15-I-60] WENGER (S.): Common Conditions for Video Performance Evaluation in H.324/M error-prone systems, *Contribution Q15-I-60 to ITU-T Video Coding Experts Group Ninth Meeting (UIT-T SG16 Q.15 Période d'études 1997-2000)*, Red Bank, New Jersey, USA, octobre 1999.
- [Q15-J-62] HANNUKSELA (M.): Picture header recovery for H.263 Test Model for H.324/M Use, *Contribution Q15-J-62 to ITU-T Video Coding Experts Group Tenth Meeting (UIT-T SG16 Q.15 Période d'études 1997-2000)*, Osaka, Japon, mai 2000.
- [RFC 1889] SCHULZRINNE (H.), CASNER (S.), FREDERICK (R.), JACOBSON (V.): RTP: A transport Protocol for Real-Time Applications, *IETF RFC 1889*, janvier 1996.
- [RFC 2429] BORMANN (C.), CLINE (L.), DEISHER (G.), GARDOS (T.), MACIOCCO (C.), NEWELL (D.), OTT (J.), SULLIVAN (G.), WENGER (S.), ZHU (C.): RTP Payload Format for the 1998 Version of ITU-T Rec. H.263 Video (H.263+), *IETF RFC 2429*, octobre 1998.

- [SG88] SHOHAM (Y.), GERSHO (A.): "Efficient Bit Allocation for an Arbitrary Set of Quantizers" (Attribution efficace des bits pour un ensemble arbitraire de quantificateurs), dans *IEEE Trans. On Acoustics, Speech, and Signal Processing*, Vol. 36, No. 9, p. 1445-453, septembre 1988.
- [WZG99] WIEGAND (T.), ZHANG (X.), GIROD (B.): "Long-Term Memory Motion-Compensated Prediction" (Prédiction cinétiquement compensée à long terme de mémorisation), dans *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 9, No. 1, p. 70-84, février 1999.

SÉRIES DES RECOMMANDATIONS UIT-T

Série A	Organisation du travail de l'UIT-T
Série B	Moyens d'expression: définitions, symboles, classification
Série C	Statistiques générales des télécommunications
Série D	Principes généraux de tarification
Série E	Exploitation générale du réseau, service téléphonique, exploitation des services et facteurs humains
Série F	Services de télécommunication non téléphoniques
Série G	Systèmes et supports de transmission, systèmes et réseaux numériques
Série H	Systèmes audiovisuels et multimédias
Série I	Réseau numérique à intégration de services
Série J	Réseaux câblés et transmission des signaux radiophoniques, télévisuels et autres signaux multimédias
Série K	Protection contre les perturbations
Série L	Construction, installation et protection des câbles et autres éléments des installations extérieures
Série M	RGT et maintenance des réseaux: systèmes de transmission, circuits téléphoniques, télégraphie, télécopie et circuits loués internationaux
Série N	Maintenance: circuits internationaux de transmission radiophonique et télévisuelle
Série O	Spécifications des appareils de mesure
Série P	Qualité de transmission téléphonique, installations téléphoniques et réseaux locaux
Série Q	Commutation et signalisation
Série R	Transmission télégraphique
Série S	Equipements terminaux de télégraphie
Série T	Terminaux des services télématiques
Série U	Commutation télégraphique
Série V	Communications de données sur le réseau téléphonique
Série X	Réseaux de données et communication entre systèmes ouverts
Série Y	Infrastructure mondiale de l'information et protocole Internet
Série Z	Langages et aspects généraux logiciels des systèmes de télécommunication