RECOMMANDATION UIT-R BT.1683

Techniques de mesure objective de la qualité vidéo perceptuelle pour la télédiffusion numérique à définition normale en présence d'une image de référence complète

(Question UIT-R 44/6)

(2004)

L'Assemblée des radiocommunications de l'UIT,

considérant

a) qu'il n'est plus à démontrer depuis longtemps que la possibilité de mesurer automatiquement la qualité d'une séquence vidéo diffusée constitue un atout précieux pour l'industrie;

b) que les méthodes objectives classiques ne conviennent plus parfaitement pour mesurer la qualité vidéo perçue de systèmes vidéo numériques utilisant la compression;

c) que les mesures objectives de la qualité vidéo perçue viendront compléter les méthodes d'essai objectives classiques;

d) que les méthodes d'évaluation subjectives classiques qui existent actuellement sont chronophages et coûteuses et ne sont, en général, pas adaptées aux conditions d'exploitation;

e) que les mesures objectives de la qualité vidéo perçue peuvent utilement compléter les méthodes d'évaluation subjective,

recommande

1 d'utiliser les lignes directrices, les paramètres et les limites indiqués dans l'Annexe 1 pour appliquer les modèles d'évaluation objective de la qualité vidéo décrits dans les Annexes 2 à 5;

2 d'utiliser les modèles d'évaluation objective de la qualité vidéo décrits dans les Annexes 2 à 5 pour effectuer des mesures objectives de la qualité vidéo perçue.

Annexe 1

Résumé

La présente Recommandation spécifie les méthodes à utiliser pour estimer la qualité vidéo perçue d'un système de transmission vidéo unidirectionnel. Elle s'applique aux signaux en bande de base. Les méthodes d'estimation qui y sont décrites sont valables pour:

– l'évaluation du codec, ses spécifications et les essais d'homologation;

– le contrôle de qualité pendant le service, éventuellement en temps réel, à la source;

– le télécontrôle de qualité à la destination, lorsqu'une copie de la source est disponible;

– la mesure de qualité d'un système d'archivage vidéo ou d'un système numérique vidéo qui utilise des techniques de compression et de décompression numériques, par application unique ou concaténation de ces techniques.

Introduction

Il n'est plus à démontrer depuis longtemps qu'il est précieux pour l'industrie de pouvoir mesurer automatiquement la qualité d'une séquence vidéo diffusée. En effet, les radiodiffuseurs ont besoin de tels outils pour remplacer ou compléter les essais d'évaluation subjective de la qualité qui soit coûteux et chronophages. Jusqu'à présent les mesures objectives de la qualité ont été faites par calcul de la valeur de crête du rapport signal/bruit (PSNR). Ce rapport est certes un indicateur utile de la qualité mais il a été montré qu'il donnait une représentation moins que satisfaisante de la qualité perceptuelle. Pour s'affranchir des limites liées au rapport PSNR, on s'est orienté dans les recherches vers la définition d'algorithmes permettant de mesurer la qualité perceptuelle d'une séquence vidéo diffusée. De tels outils de mesure objective de la qualité perceptuelle peuvent être utilisés pour tester les performances d'un réseau de radiodiffusion, comme aide d'achat d'équipements et pour la mise au point de nouvelles techniques de codage vidéo de radiodiffusion. Ces dernières années beaucoup de travaux ont été consacrés à la mise au point d'outils fiables et précis susceptibles d'être utilisés pour mesurer objectivement la qualité perceptuelle d'une séquence vidéo diffusée. La présente Recommandation définit des modèles de calcul objectif dont il a été montré qu'ils étaient de meilleurs outils de mesure automatique que le rapport PSNR pour évaluer la qualité d'une séquence vidéo diffusée. Les modèles ont été testés sur des séquences à 525 lignes et à 625 lignes conformes à la Recommandation UIT-R BT.601 qui caractérise la distribution secondaire de signaux vidéo numériques de qualité télévision.

Les résultats obtenus avec les modèles d'évaluation de la qualité perceptuelle ont été évalués dans le cadre de deux évaluations parallèles de la séquence vidéo soumise à l'essai[1]. Dans la première évaluation, on a utilisé une méthode d'évaluation subjective normalisée, la méthode d'échelle de qualité continue à double stimulus (DSCQS) pour obtenir auprès de groupes d'observateurs humains des indices subjectifs de la qualité de la séquence vidéo. Voir la Recommandation UIT-R BT.500 – Méthode d'évaluation subjective de la qualité des images de télévision. Dans la seconde évaluation, les indices objectifs ont été calculés à l'aide de modèles de calcul objectif. Pour chaque modèle, on a procédé à plusieurs calculs pour mesurer l'exactitude et la cohérence avec lesquelles les indices objectifs permettent de prévoir les indices subjectifs. Trois laboratoires indépendants se sont chargés de la partie évaluation subjective du test. Deux laboratoires, Communications Research Center (CRC, Canada) et Verizon (Etats-Unis d'Amérique), ont effectué les tests avec des séquences 525/60 Hz et un troisième laboratoire Fondazione Ugo Bordoni (FUB, Italie) avec des séquences 625/50 Hz. Plusieurs laboratoires ont élaboré des modèles de calcul objectif de la qualité vidéo des mêmes séquences vidéo testés avec des observateurs humains par CRC, Verizon et FUB. Les résultats des tests sont donnés dans l'Appendice 1.

La présente Recommandation comprend les modèles de calcul objectif indiqués dans le Tableau 1.

Une description complète des quatre modèles de calcul objectif est donnée dans les Annexes 2 à 5.

On peut utiliser les équipements d'essai de la qualité vidéo existants en attendant que de nouveaux équipements d'essai utilisant l'un quelconque des quatre modèles ci-dessus soient facilement disponibles.

---

[1] Document UIT-R 6Q/14 [septembre 2003] Final Report from the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment, Phase II (FR-TV2).

TABLEAU 1

Numéro du modèle	Nom	Initiateur Groupe d'experts en qualité vidéo (VQEG)	Pays	Annexe
1	British Telecom	D	Royaume-Uni	2
2	Yonsei University/Radio Research Laboratory/SK Telecom	E	Corée (Rép. de)	3
3	Centre de recherche et de développement en télécommunication (CPqD)	F	Brésil	4
4	National Telecommunications and Information Administration (NTIA)/Institute for Telecommunication Sciences (ITS)	H	Etats-Unis d'Amérique	5

Pour envisager l'inclusion d'un modèle quelconque dans la partie normative de la présente Recommandation, le modèle doit être vérifié par un organe indépendant ouvert (par exemple, le VQEG) qui effectuera l'évaluation technique dans le respect des lignes directrices et des critères de performance fixés par la Commission d'études 6 des radiocommunications. L'intention de la Commission d'études 6 des radiocommunications est de recommander à terme une seule méthode de référence complète normative.

1 Champ d'application

La présente Recommandation spécifie les méthodes à utiliser pour évaluer la qualité vidéo perçue d'un système vidéo unidirectionnel. Elle s'applique aux signaux en bande de base. Les estimateurs de la qualité vidéo objective sont définis pour la qualité de bout en bout entre les deux points. Les méthodes d'estimation sont basées sur le traitement d'une séquence vidéo à composante numérique à 8 bits telle qu'elle est définie dans la Recommandation UIT-R BT.601[2]. Le codeur peut utiliser diverses méthodes de compression (par exemple Groupe d'experts pour les images animées (MPEG), Recommandation UIT-T H.263, etc.). Les modèles proposés dans la présente Recommandation peuvent être utilisés pour évaluer un codec (combinaison codeur/décodeur) ou une concaténation de diverses méthodes de compression et dispositifs d'archivage de mémoire. Le calcul des estimateurs de qualité objectifs décrit dans la présente Recommandation aura peut-être tenu compte des dégradations dues aux erreurs (erreurs sur les bits, perte de paquets) mais on ne dispose pas actuellement de résultats d'essai indépendants permettant de valider l'utilisation des estimateurs pour des systèmes présentant des dégradations dues à des erreurs. Le matériel d'essai de validation ne contenait pas d'erreurs sur les canaux.

---

[2] Cela n'exclut pas la mise en oeuvre de la méthode de mesure pour des systèmes vidéo unidirectionnels qui utilisent une entrée vidéo et des sorties vidéo composites. Les spécifications de la conversion entre le domaine composite et le domaine à composante n'entrent pas dans le cadre de la présente Recommandation. Par exemple, la norme SMPTE 170M spécifie une méthode pour effectuer cette conversion dans le cas d'un système NTSC.

1.1 Application

La présente Recommandation donne des estimations de la qualité vidéo pour différentes classes de télévision (TV0-TV3), et pour la classe vidéo multimédia (MM4) définie dans l'Annexe B de la Recommandation UIT-T P.911. Les applications des modèles d'estimation décrits dans la présente Recommandation sont notamment les suivantes:

– évaluation du codec, spécification du codec, essai d'homologation, contenu de la précision limitée décrite ci-dessous;

– contrôle de la qualité pendant le service, éventuellement en temps réel, à la source;

– télécontrôle de la qualité au point de destination lorsqu'on dispose d'une copie de la source;

– mesures de qualité d'un système d'archivage ou de transmission qui utilise des techniques de compression ou de décompression vidéo, par passage unique ou concaténation de telles techniques.

1.2 Limitations

Les modèles d'estimation décrits dans la présente Recommandation ne peuvent être utilisés pour remplacer les essais subjectifs. Les valeurs de corrélation entre deux essais subjectifs conçus et exécutés avec soin (par exemple dans deux laboratoires différents) se situent normalement dans la fourchette 0,92-0,97. La présente Recommandation ne donne pas de moyens permettant de quantifier d'éventuelles erreurs d'estimation. Les utilisateurs de la présente Recommandation devraient comparer les résultats des évaluations subjectives et objectives disponibles pour avoir une idée de la fourchette des erreurs d'estimation des indices de qualité vidéo.

Les performances prévues des modèles d'estimation ne sont pas actuellement validées pour des systèmes vidéo comportant des dégradations dues à des erreurs sur les canaux de transmission.

Annexe 2

Modèle 1

TABLE DES MATIÈRES

Page

1 Introduction

L'outil d'évaluation automatique de la qualité vidéo avec une image de référence complète de BT (BTFR, *BT full-reference*) permet d'avoir des prévisions de la qualité vidéo qui sont représentatives des jugements de qualité de l'être humain. Cet outil de mesure objective simule numériquement les caractéristiques du système visuel humain (HVS, *human visual system*) pour donner des prévisions précises de la qualité vidéo et constitue une alternative viable aux évaluations subjectives classiques qui sont coûteuses et chronophages.

Une mise en oeuvre logicielle du modèle a été intégrée dans les tests VQEG2 et les résultats correspondants ont été présentés dans un rapport sur les essais[1].

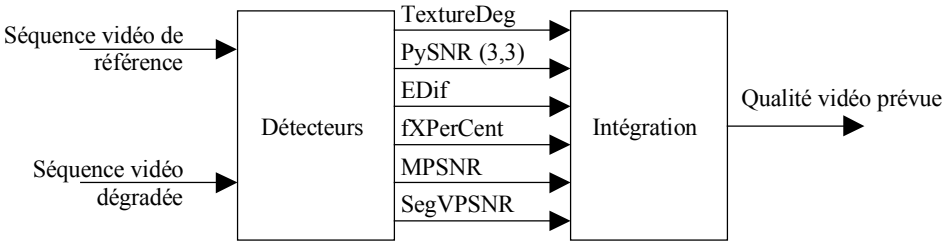
2 Modèle d'image de BTFR

L'algorithme BTFR effectue une détection suivie d'une intégration (voir la Fig. 1). Par détection, on entend le calcul d'un ensemble de paramètres du détecteur perceptuellement significatifs à partir de la séquence vidéo non déformée (de référence) et de la séquence vidéo déformée (dégradée). Ces paramètres constituent alors les données d'entrée pour l'intégrateur qui donne une estimation de la qualité vidéo perçue avec une pondération appropriée. Le choix des détecteurs et des facteurs de pondération est fonction de caractéristiques de masquage spatial et temporel connues du HVS et déterminé par étalonnage.

Le modèle accepte des séquences vidéo d'entrée de type 625 (720 × 576) entrelacées à 50 trames/s et 525 (720 × 486) entrelacées à 59,94 trames/s en format *YUV*422.

FIGURE 1

Modèle d'évaluation de la qualité vidéo avec une image de référence complète



MPSNR: rapport PSNR adapté

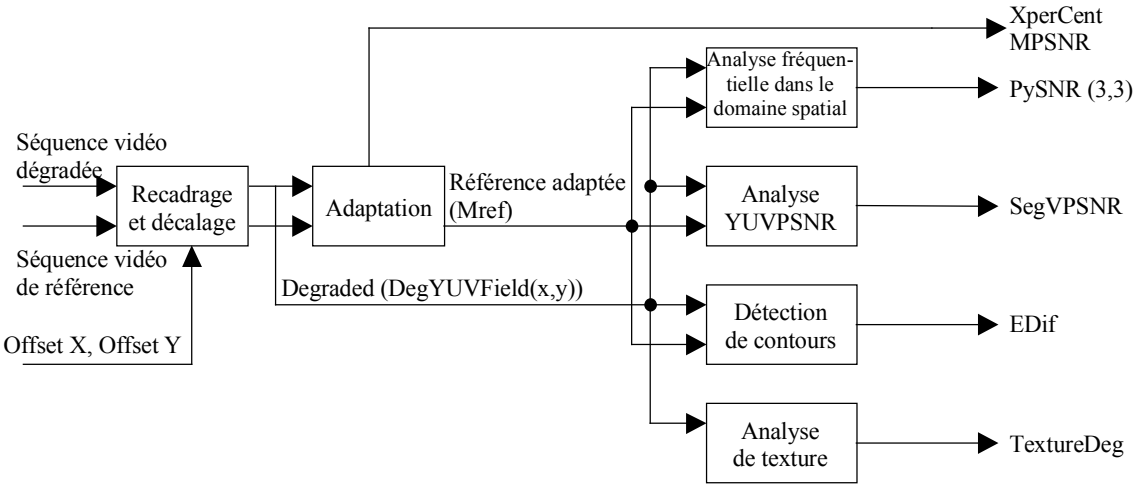
1683-01

3 Détecteurs

Le module de détection de l'algorithme BTFR effectue un certain nombre de mesures fréquentielles dans le domaine temporel et le domaine spatial à partir des séquences d'entrée formatées *YUV* (voir la Fig. 2).

FIGURE 2

Détection



1683-02

3.1 Conversion des séquences d'entrée

Tout d'abord, les séquences d'entrée sont converties du format entrelacé *YUV*422 à un format desentrelacé de bloc *YUV*444 de sorte que chaque trame successive est représentée par des tableaux *RefY*, *RefU* et *RefV*:

$$RefY(x, y) \qquad x = 0...X-1, \qquad y = 0...Y-1 \tag{1}$$

$$RefU(x, y) \qquad x = 0...X-1, \qquad y = 0...Y-1 \tag{2}$$

$$RefV(x, y) \qquad x = 0...X-1, \qquad y = 0...Y-1 \tag{3}$$

où:

X: nombre de pixels horizontaux dans une trame

Y: nombre de pixels verticaux.

Pour une séquence d'entrée *YUV*422, chaque valeur de *U* et chaque valeur de *V* doivent être répétées pour obtenir les matrices (2) et (3) avec une résolution complète.

3.2 Recadrage et décalage

Cette routine recadre, avec décalage, la séquence d'entrée dégradée et recadre, sans décalage, la séquence d'entrée de référence. Les paramètres de décalage *XOffset* et *YOffset*, déterminés extérieurement, définissent de combien de pixels horizontaux et verticaux la séquence est décalée par rapport à la séquence de référence. L'origine de l'image est située dans le coin supérieur gauche, avec un déplacement positif horizontal vers la droite et vertical vers le bas. Une valeur du paramètre *XOffset* de 2 indique que les trames dégradées sont décalées vers la droite de 2 pixels et une valeur du paramètre *YOffset* de 2 indique un décalage vers le bas de 2 pixels. Pour une trame d'entrée avec des valeurs *YUV* archivées en format *YUV*444 (§ 3.1) dans des tableaux *InYField*, *InUField* et *InVField*, la séquence de sortie recadrée et décalée est calculée selon les expressions (4) à (20).

$$XStart = -XOffset \tag{4}$$

$$\text{si } (XStart < C_x) \qquad \text{alors } XStart = C_x \tag{5}$$

$$XEnd = X - 1 - XOffset \tag{6}$$

$$\text{si } (XEnd > X - C_x - 1) \qquad \text{alors } XEnd = X - C_x - 1 \tag{7}$$

$$YStart = -YOffset \tag{8}$$

$$\text{si } (YStart < C_y) \qquad \text{alors } YStart = C_y \tag{9}$$

$$YEnd = Y - 1 - Yffset \tag{10}$$

$$\text{si } (YEnd > Y - C_y - 1) \qquad \text{alors } YEnd = Y - C_y - 1 \tag{11}$$

X et Y donnent respectivement la dimension de trame horizontale et la dimension de trame verticale et C_x et C_y le nombre de pixels à recadrer depuis la gauche et la droite ainsi que le haut et le bas.

Pour des séquences à 625 lignes,

$$X = 720, \qquad Y = 288, \qquad C_x = 30, \qquad C_y = 10 \tag{12}$$

Pour des séquences à 525 lignes,

$$X = 720, \qquad Y = 243, \qquad C_x = 30, \qquad C_y = 10 \tag{13}$$

Xstart, *Xend*, *Ystart* et *Yend* définissent maintenant la région de chaque trame qui sera copiée. Les pixels situés en dehors de cette région sont initialisés selon les équations (14) et (15), dans lesquelles *YField*, *UField* et *VField* sont les tableaux de pixels de sortie *XxY* contenant respectivement les valeurs *Y*, *U* et *V*.

Les barres verticales à gauche et à droite de la trame sont initialisées selon:

$$YField(x, y) = 0 \qquad x = 0...XStart - 1, XEnd + 1...X - 1 \qquad y = 0...Y - 1 \tag{14}$$

$$UField(x, y) = VField(x, y) = 128 \qquad x = 0...XStart - 1, XEnd + 1...X - 1 \qquad y = 0...Y - 1 \tag{15}$$

Les barres horizontales en haut et en bas de la trame sont initialisées selon:

$$YField\,(x, y) = 0 \qquad x = XStart\,...XEnd\,, \qquad y = 0...YStart\, - 1, YEnd\, + 1...Y - 1 \tag{16}$$

$$UField(x, y) = VField(x, y) = 128 \qquad x = XStart...XEnd \qquad y = 0...YStart - 1, YEnd + 1...Y - 1 \tag{17}$$

Enfin, les valeurs des pixels sont copiées selon:

$$YField(x, y) = InYField(x + XOffset, y + YOffset) \quad x = XStart...XEnd \quad y = YStart...YEnd \tag{18}$$

$$UField(x, y) = InUField(x + XOffset, y + YOffset) \quad x = XStart...XEnd \quad y = YStart...YEnd \tag{19}$$

$$VField(x, y) = InVField(x + XOffset, y + YOffset) \quad x = XStart...XEnd \quad y = YStart...YEnd \tag{20}$$

Pour la séquence d'entrée dégradée, le recadrage et le décalage génèrent des tableaux de trame de sortie *DegYField*, *DegUField* et *DegVField* tandis que le recadrage sans décalage pour la séquence de référence génère *RefYField*, *RefUField* et *RefVfield*. Ces tableaux bidimensionnels *XxY* servent de données d'entrée pour les routines de détection décrites ci-après.

3.3 Adaptation

Le processus d'adaptation produit des signaux destinés à être utilisés dans d'autres procédures de détection, ainsi que des paramètres de détection destinés à être utilisés dans la procédure d'intégration. Pour les signaux d'adaptation, on cherche, pour de petits blocs dans chaque trame dégradée, dans une mémoire tampon de trames de référence voisines les trames qui correspondent le mieux. On obtient ainsi une séquence, la séquence de référence adaptée, destinée à être utilisée en lieu et place de la séquence de référence dans certains des modules de détection.

L'analyse d'adaptation est réalisée sur des blocs de pixels 9×9 des tableaux d'intensité *RefYField* et *DegYField*. Si l'on ajoute la dimension nombre de trames aux tableaux d'intensité, le pixel (*Px*, *Py*) de la trame de référence *N* peut être représenté comme suit:

$$Ref(N, Px, Py) = RefYField(Px, Py) \quad \text{trame} \quad N \tag{21}$$

Un bloc de pixels 9×9 avec un pixel central (*Px*, *Py*) dans la $N^{\text{ième}}$ trame peut être représenté comme suit:

$$BlockRef(N, Px, Py) = Ref(n, x, y) \qquad x = Px - 4...Px + 4, \qquad y = Py - 4...Py + 4 \tag{22}$$

Deg(*n, x, y*) et *BlockDeg*(*n, x, y*) peuvent être définis de la même manière.

Pour *BlockDeg*(*N, Px, Py*), on calcule une erreur d'adaptation minimale *E*(*N, Px, Py*) en cherchant les trames de référence voisines selon l'équation:

$$\begin{aligned} E(N, Px, Py) = \text{Min}((1/81) \sum_{j=-4}^{4} \sum_{k=-4}^{4} (&\text{Deg}(N, Px + j, Py + k) \\ &- Ref(n, x + j, y + k))^2) \end{aligned} \tag{23}$$

$$\begin{aligned} n &= N - 4, \ldots, N + 5 \\ x &= Px - 4,..,Px,..,Px + 4 \\ y &= Py - 4,..,Py,..,Py + 4 \end{aligned}$$

où *N* est l'indice de la trame dégradée contenant le bloc dégradé qui fait l'objet de l'adaptation.

Si l'équation (23) permet de déterminer que la meilleure correspondance avec *BlockDeg*(*N, Px, Py*) est *BlockRef*(n_m, x_m, y_m), alors un tableau de référence adaptée *MRef* est mis à jour selon:

$$MRef(N, Px + j, Py + k) = Ref(n_m, x_m + j, y_m + k) \qquad j = -4...4, k = -4...4 \qquad (24)$$

Le processus d'adaptation de recherche de la meilleure correspondance pour un bloc dégradé suivie de la copie du bloc résultant dans le tableau de référence adapté est répété pour l'ensemble de la zone d'analyse souhaitée. Cette zone d'analyse est définie par les points centraux de blocs *Px*() et *Py*() selon:

$$Px(h) = 16 + 8 \times h \qquad h = 0...Qx - 1 \qquad (25)$$

et

$$Py(v) = 16 + 8 \times v \qquad v = 0...Qy - 1 \qquad (26)$$

où *Qx* et *Qy* définissent le nombre de blocs d'analyse horizontaux et verticaux.

L'analyse d'adaptation de la $N^{\text{ième}}$ trame produit donc une séquence de référence adaptée décrite par:

$$BlockMRef\ (N, Px(h), Py\ (v)) \qquad h = 0...Qx - 1, \qquad v = 0...Qy - 1 \qquad (27)$$

et un ensemble de valeurs d'erreur pour la meilleure correspondance:

$$E(N, Px(h), Py(v)) \qquad h = 0...Qx - 1, \qquad v = 0...Qy - 1 \qquad (28)$$

Un ensemble de tableaux de décalage *MatT*, *MatX* et *MatY* peuvent être définis de façon que:

$$BlockMRef(N, Px(h), Py(v))) = BlockRef(MatT(h,v), MatX(h,v), MatY(h,v)) \qquad h = 0...Qx - 1, \qquad v = 0...Qy - 1 \qquad (29)$$

Les paramètres d'adaptation pour des séquences de radiodiffusion à 625 lignes ou 525 lignes sont donnés dans le Tableau 2.

TABLEAU 2

Paramètres de recherche pour la procédure d'adaptation

Paramètre	625	525
Qx	87	87
Qy	33	28

La zone d'analyse définie par les équations (26) et (27) ne couvre pas l'ensemble de la trame. *MRef* doit donc être initialisé selon l'équation (29) de façon à pouvoir être utilisé ailleurs sans restriction.

$$MRef(x, y) = 0 \qquad x = 0...X - 1, \qquad y = 0...Y - 1 \qquad (30)$$

3.3.1 Statistiques d'adaptation

Les statistiques d'adaptation horizontales sont calculées à partir du processus d'adaptation et destinées à être utilisées dans le processus d'intégration. La meilleure correspondance pour chaque bloc d'analyse, déterminée selon l'équation (23), est utilisée dans la construction de l'histogramme *histX* pour chaque trame selon:

$$\begin{aligned} histX(MatX(h,v) - Px(h) + 4 = histX(MatX(h,v) - Px(h) + 4) + 1 \\ h = 0...Qx - 1, \qquad v = 0...Qx - 1 \end{aligned} \tag{31}$$

où le tableau *histX* est initialisé à zéro pour chaque trame. L'histogramme est ensuite utilisé pour déterminer la mesure *fXPerCent* selon:

$$fXPerCent = 100 \times \text{Max}(histX(i)) / \sum_{j=0}^{8} histX(j) \qquad i = 0...8 \tag{32}$$

Pour chaque trame, la mesure *fXPerCent* donne la proportion (%) de blocs adaptés qui interviennent dans la crête de l'histogramme d'adaptation.

3.3.2 Rapport MPSNR

L'erreur minimale, *E*(), pour chaque bloc adapté est utilisé pour calculer un rapport SNR adapté selon:

$$\begin{aligned} &\text{si} \left(\sum_{h=0}^{Qx-1} \sum_{v=0}^{Qy-1} E(N, Px(h), Py(v)) \right) > 0 \qquad \text{alors} \\ &MPSNR = 10 \log_{10} \left(Qx \times Qy \times 255^2 \Big/ \sum_{h=0}^{Qx-1} \sum_{v=0}^{Qy-1} E(N, Px(h), Py(v)) \right) \end{aligned} \tag{33}$$

$$\text{si} \left(\sum_{h=0}^{Qx-1} \sum_{v=0}^{Qy-1} E(N, Px(h), \ Py(v)) \right) = 0 \qquad \text{alors} \quad MPSNR = 10 \log_{10}(255^2) \tag{34}$$

3.3.3 Vecteurs d'adaptation

Le vecteur horizontal, le vecteur vertical et le vecteur retard sont archivés en vue d'une utilisation future selon:

$$SyncT(h,v) = MatT(h,v) - N \qquad h = 0...Qx - 1, \qquad v = 0...Qy - 1 \tag{35}$$

$$SyncX(h,v) = MatX(h,v) - Px(h) \qquad h = 0...Qx - 1, \qquad v = 0...Qy - 1 \tag{36}$$

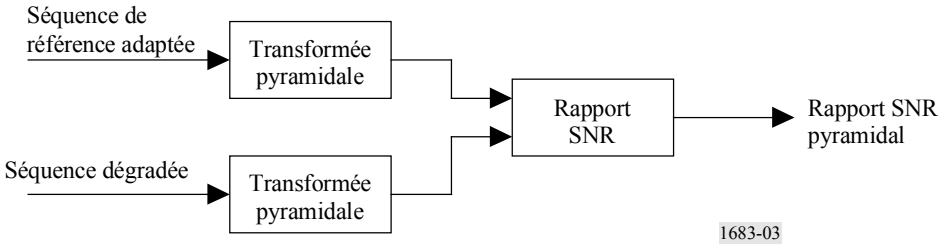
$$SyncY(h,v) = MatY(h,v) - Py(h) \qquad h = 0...Qx - 1, \qquad v = 0...Qy - 1 \tag{37}$$

3.4 Analyse fréquentielle dans le domaine spatial

Le détecteur fréquentiel dans le domaine spatial est basé sur une transformation "pyramidale" des séquences de référence dégradée et adaptée. Tout d'abord, chaque séquence est transformée pour donner un tableau pyramidal de référence et un tableau pyramidal dégradé. Ensuite, les différences entre les tableaux pyramidaux sont calculées à l'aide d'une mesure de l'erreur quadratique moyenne et les résultats en sortie sont présentés sous forme d'un rapport SNR pyramidal.

FIGURE 3

Analyse fréquentielle dans le domaine spatial



3.4.1 Transformée pyramidale

Tout d'abord, la trame d'entrée *F* est copiée dans un tableau pyramidal *P* selon:

$$P(x,y) = F(x,y) \qquad x = 0...X-1, \qquad y = 0...Y-1 \tag{38}$$

Ce tableau pyramidal est ensuite mis à jour par analyse horizontale et verticale en trois étapes (étape = 0..2). L'analyse horizontale *Hpy*(*stage*) est définie par les équations (39) à (43).

Tout d'abord, il est fait une copie temporaire de l'ensemble du tableau pyramidal:

$$PTemp(x,y) = P(x,y) \qquad x = 0...X-1, \qquad y = 0...Y-1 \tag{39}$$

Ensuite les limites *x* et *y* sont calculées selon:

$$Tx = X / 2^{(stage+1)} \tag{40}$$

$$Ty = Y / 2^{stage} \tag{41}$$

Les moyennes et les différences des paires horizontales d'éléments du tableau temporaire sont ensuite utilisées pour mettre à jour le tableau pyramidal selon:

$$P(x,y) = 0{,}5\ (PTemp(2x,y) + PTemp(2x+1,y)) \qquad x = 0...Tx-1, \qquad y = 0...Ty-1 \tag{42}$$

$$P(x+Tx,y) = PTemp(2x,y) - PTemp(2x+1,y) \qquad x = 0...Tx-1 \qquad y = 0...Ty-1 \tag{43}$$

L'analyse verticale *Vpy*(*stage*) est définie par les équations (44) à (48).

$$PTemp(x,y) = P(x,y) \qquad x = 0...X-1, \qquad y = 0...Y-1 \tag{39}$$

$$Tx = X / 2^{stage} \tag{45}$$

$$Ty = Y / 2^{(stage+1)} \tag{46}$$

Les moyennes et les différences des paires verticales d'éléments du tableau temporaire sont ensuite utilisées pour mettre à jour le tableau pyramidal selon:

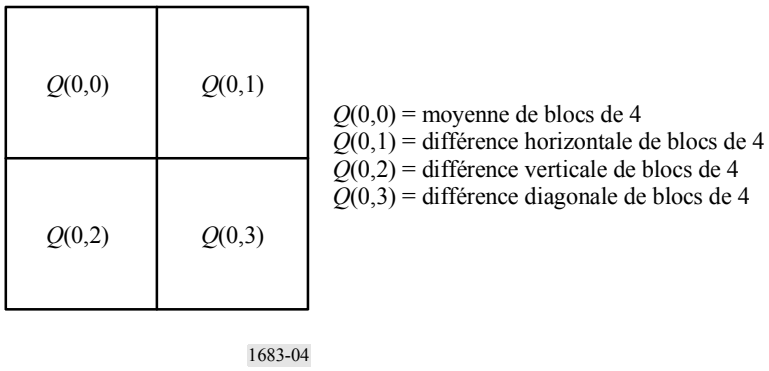
$$P(x, y) = 0{,}5\,(PTemp(x,2y) + PTemp(x,2y+1)) \qquad x = 0...Tx-1, \qquad y = 0...Ty-1 \qquad (47)$$

$$P(x, y+Ty) = PTemp(x,2y) - PTemp(x,2y+1) \qquad x = 0...Tx-1 \qquad y = 0...Ty-1 \qquad (48)$$

Pour l'étape 0, l'analyse horizontale *Hpy*(0) suivie de l'analyse verticale *Vpy*(0) met à jour l'ensemble du tableau pyramidal avec les quatre quadrants *Q*(étape, 0...3) structurés comme suit:

FIGURE 4

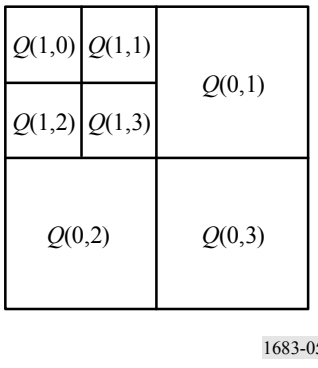
Représentation en quadrants de la sortie de l'analyse, étape 0



L'analyse étape 1 est ensuite réalisée sur *Q*(0,0) pour obtenir les résultats *Q*(1,0...3) qui sont archivés dans la pyramide selon:

FIGURE 5

Représentation en quadrants de la sortie de l'analyse, étape 1



L'analyse étape 2 traite *Q*(1,0) et le remplace par *Q*(2,0...3).

A l'issue des trois stades de l'analyse, le tableau pyramidal résultant comporte un total de 10 blocs de résultats. Trois blocs *Q*(0,1...3) proviennent de l'analyse des pixels 2×2 , étape 0, trois *Q*(1,1...3) de l'analyse des pixels 4×4, étape 1 et 4 *Q*(2,0...3) de l'analyse des pixels 8×8, étape 2.

L'analyse en trois étapes de la séquence de référence adaptée et de la séquence dégradée produit les tableaux pyramidaux *Pref* et *Pdeg*. Les différences entre ces tableaux sont ensuite mesurées dans le module SNR pyramidal.

3.4.2 Rapport SNR pyramidal

On mesure l'erreur quadratique entre le tableau pyramidal de référence et le tableau pyramidal dégradé sur les quadrants 1 à 3 des étapes 0 à 2 selon:

$$E(s,q) = (1/XY^2) \sum_{x=x1(s,q)}^{x2(s,q)-1} \sum_{y=y1(s,q)}^{y2(s,q)-1} (Pref(x,y) - Pdeg(x,y))^2 \quad s = 0...2 \quad q = 1...3 \quad (49)$$

où, $x1$, $x2$, $y1$ et $y2$ définissent les limites horizontales et verticales des quadrants dans les tableaux pyramidaux et sont calculés selon:

$$x1(s,1) = X/2^{(s+1)} \quad x2(s,1) = 2 \times x1(s,1) \quad y1(s,1) = 0 \quad y2(s,1) = Y/2^{(s+1)} \quad (50)$$

$$x1(s,2) = 0 \quad x2(s,2) = X/2^{(s+1)} \quad y1(s,2) = Y/2^{(s+1)} \quad y2(s,2) = 2 \times y1(s,2) \quad (51)$$

$$x1(s,3) = X/2^{(s+1)} \quad x2(s,3) = 2 \times x1(s,3) \quad y1(s,3) = Y/2^{(s+1)} \quad y2(s,3) = 2 \times y1(s,3) \quad (52)$$

Les résultats de l'équation (49) sont ensuite utilisés pour mesurer le PSNR pour chaque quadrant de chaque trame selon:

$$\begin{aligned} &\text{si } (E > 0,0) \qquad PySNR(s,q) = 10 \log_{10}(255^2 / E(s,q)) \\ &\text{sinon} \qquad SNR = 10 \log_{10}(255^2 \; XY^2) \end{aligned} \quad (53)$$

où le nombre d'étapes $s = 0...2$ et le nombre de cadrans pour chaque étape $q = 1...3$.

3.5 Analyse de la texture

On mesure la texture de la séquence dégradée en enregistrant le nombre de points de transition du signal d'intensité sur les lignes horizontales de l'image, selon les équations (54) à (59).

Pour chaque trame, un compteur de points de transition est tout d'abord initialisé selon l'équation (54).

$$sum = 0 \quad (54)$$

Puis, chaque ligne $y = 0...Y - 1$, est traitée pour $x = 0...X - 2$ selon:

$$last\_pos = 0, \qquad last\_neg = 0 \quad (55)$$

$$dif(x) = P(x,y) - P(x+1,y) \quad (56)$$

$$\text{si } ((dif(x) < 0) \text{ et } (last\_neg < last\_pos)) sum = sum + 1 \quad (57)$$

$$\text{si } ((dif(x) > 0) \text{ et } (last\_neg > last\_pos)) sum = sum + 1 \quad (58)$$

$$\text{si } (dif(x) > 0) \;\; last\_pos = x \quad (59)$$

$$\text{si } (dif(x) < 0) \;\; last\_neg = x \quad (60)$$

Quand toutes les lignes d'une trame ont été traitées, le compteur, *sum*, contiendra le nombre de points de transition du signal d'intensité horizontal. Ce nombre est ensuite utilisé pour calculer un paramètre de texture pour chaque trame selon:

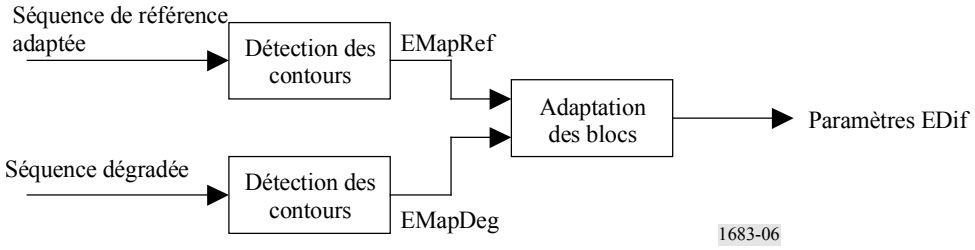
$$TextureDeg = sum \times 100 / XY \quad (61)$$

3.6 Analyse des contours

Chaque trame de la séquence dégradée et de la séquence de référence adaptée subit séparément une routine de détection des bords pour produire des représentations correspondantes des bords de la trame, lesquelles sont ensuite comparées dans une procédure d'adaptation de blocs pour établir les paramètres de détection.

FIGURE 6

Analyse des bords



3.6.1 Détection des contours

On a utilisé un détecteur de contours de Canny [Canny, 1986] pour déterminer les représentations des contours, mais d'autres techniques analogues de détection de contours bords peuvent également être utilisées. Les représentations des contours résultantes, *EMapRef* et *EMapDeg*, sont des représentations de pixels où un contour est indiqué par un 1 et l'absence de contour par un 0,

Pour la détection d'un contour ou pixel (*x, y*):

$$EMap(x,y) = 1 \qquad x = 0...X-1, \qquad y = 0...Y-1 \tag{62}$$

Pour la détection d'une absence de contour ou pixel (*x, y*):

$$EMap(x,y) = 0 \qquad x = 0...X-1, \qquad y = 0...Y-1 \tag{63}$$

3.6.2 Différenciation des contours

La procédure de différenciation des contours permet de mesurer les différences entre les représentations des contours pour la trame dégradée et la trame de référence adaptée correspondante. L'analyse est effectuée dans *NxM* blocs de pixels ne se chevauchant pas selon les équations (64) à (68).

Tout d'abord, on calcule le nombre de pixels marqués par un bord dans chaque bloc d'analyse où *Bh* et *Bv* définissent le nombre de blocs ne se chevauchant pas à analyser dans les directions horizontale et verticale et *X*1 et *Y*1 définissent les décalages par rapport au bord de la trame.

$$Bref(x,y) = \sum_{i=i1}^{i2} \sum_{j=j1}^{j2} EMapRef(Nx + X1 + i, My + Y1 + j) \qquad x = 0...Bh-1, y = 0...Bv-1 \tag{64}$$

$$BDef(x,y) = \sum_{i=i1}^{i2} \sum_{j=j1}^{j2} EMapDeg(Nx + X1 + i, My + Y1 + j) \qquad x = 0...Bh-1, y = 0...Bv-1 \tag{65}$$

Les limites de sommation sont déterminées selon:

$$i1 = -(N \text{ div } 2) \qquad i2 = (N-1) \text{ div } 2 \tag{66}$$

$$j1 = -(M \text{ div } 2) \qquad j2 = (M-1) \text{ div } 2 \tag{67}$$

où l'opérateur, div, représente une division par un nombre entier.

Ensuite, on effectue une mesure des différences sur la totalité de la trame selon:

$$EDif = (1/(N\ M\ Bh\ Bv)) \left(\sum_{x=0}^{Bh-1} \sum_{y=01}^{Bv-1} (B \operatorname{Re} f(x,y) - BDeg(x,y))^{Q} \right)^{1/Q} \tag{68}$$

Pour des trames de 720×288 pixels pour une séquence vidéo à 625 lignes:

$$N = 4, \quad X1 = 6, \quad Bh = 178 \qquad M = 4, \quad Y1 = 10, \quad Bv = 69, \quad Q = 3 \tag{69}$$

Pour des trames de 720×243 pixels pour une séquence vidéo à 525 lignes:

$$N = 4, \quad X1 = 6, \quad Bh = 178 \qquad M = 4, \quad Y1 = 10, \quad Bv = 58, \quad Q = 3 \tag{70}$$

3.7 Analyse du rapport MPSNR

Un rapport SNR adapté est calculé pour les valeurs du pixel *V* en utilisant les vecteurs d'adaptation définis dans les équations (35) à (37). Pour chaque ensemble de vecteurs d'adaptation une mesure d'erreur, *VE*, est calculée selon:

$$VE(h,v) = (1/81) \sum_{i=-4}^{4} \sum_{j=-4}^{4} (DegV(N, Px(h)+i, Py(h)+j) - RefVField(N + SyncT(h,v), Px(h) + SyncX(h,v) + i, Py(v) + SyncY(h,v) + j))^2 \tag{71}$$

On calcule alors une mesure du rapport PSNR segmentaire pour la trame selon:

$$SegVPSNR = (1/Qx\ Qy) \sum_{h=0}^{Qx-1} \sum_{v=0}^{Qy-1} 10 \log_{10}(255^2 / (VE(h,v)+1)) \tag{72}$$

4 Intégration

La procédure d'intégration nécessite tout d'abord une pondération temporelle des paramètres de détection trame par trame selon l'équation (73):

$$AvD(k) = (1/N) \sum_{n=0}^{N-1} D(k,n) \qquad k = 0..5 \tag{73}$$

où:

N: nombre total de trames des séquences testées

$D(k, n)$: paramètre de détection k pour la trame n.

Les paramètres de détection pondérés, $AvD(k)$, sont ensuite combinés pour donner une note de qualité prévue, PDMOS, pour la séquence de trame N selon l'équation (74):

$$PDMOS = Offset + \sum_{k=0}^{5} AvD(k) \times W(k) \tag{74}$$

Les Tableaux 3 et 4 donnent les paramètres de l'intégrateur respectivement pour les séquences à 625 lignes et celles à 525 lignes.

TABLEAU 3

Paramètres d'intégration pour un système de vidéodiffusion à 625 lignes

K	**Nom du paramètre**	***W***
0	*TextureDeg*	–0,68
1	*PySNR*(3,3)	–0,57
2	*EDif*	+58 913,294
3	*fXPerCent*	–0,208
4	*MPSNR*	–0,928
5	*SegVPSNR*	–1,529
Décalage	+176,486	
N	400	

TABLEAU 4

Paramètres d'intégration pour un système de vidéodiffusion à 525 lignes

K	**Nom du paramètre**	***W***
0	*TextureDeg*	+0,043
1	*PySNR*(3,3)	–2,118
2	*EDif*	+60 865,164
3	*fXPerCent*	–0,361
4	*MPSNR*	+1,104
5	*SegVPSNR*	–1,264
Décalage	+260,773	
N	480	

5 Alignement

Le modèle FR nécessite un bon fonctionnement de l'alignement spatial et temporel. Le modèle intègre un alignement inhérent et peut prendre en charge des décalages spatiaux entre la séquence de référence et la séquence dégradée de ±4 pixels et des décalages temporels de ±4 trames. Les décalages spatiaux ou temporels au-delà de ces limites ne sont pas pris en charge par le modèle et il faudra un module d'alignement distinct pour s'assurer que la séquence de référence et la séquence dégradée sont correctement alignées.

6 Références bibliographiques

CANNY, J. [1986] A computational approach to edge detection. *IEEE Trans. Pattern Analysis and Machine Intelligence*. Vol. 8(**6**), p. 679-698.

Annexe 2a

TABLEAU 5

Données objectives et subjectives pour un système à 525 lignes

Nom du fichier	Séquence source (SRC)	Circuit fictif de référence (HRC)	Note subjective moyenne brute	Note prévue par le modèle sur la base des données brutes	Note subjective moyenne corrigée	Note prévue par le modèle sur la base de données corrigées
V2src01_hrc01_525.yuv	1	1	–38,30757576	44,945049	0,5402368	0,69526
V2src01_hrc02_525.yuv	1	2	–39,56212121	38,646271	0,5483205	0,58989
V2src01_hrc03_525.yuv	1	3	–25,9469697	32,855755	0,4024097	0,50419
V2src01_hrc04_525.yuv	1	4	–17,24090909	21,062775	0,3063528	0,36089
V2src02_hrc01_525.yuv	2	1	–35,23636364	31,260744	0,5025558	0,48242
V2src02_hrc02_525.yuv	2	2	–18,01818182	18,732758	0,3113346	0,33715
V2src02_hrc03_525.yuv	2	3	–6,284848485	8,914509	0,1881739	0,25161
V2src02_hrc04_525.yuv	2	4	–6,983333333	4,16663	0,1907347	0,21776
V2src03_hrc01_525.yuv	3	1	–31,96515152	22,348713	0,4682724	0,37461
V2src03_hrc02_525.yuv	3	2	–17,47727273	10,44728	0,3088831	0,26352
V2src03_hrc03_525.yuv	3	3	–1,104545455	2,494911	0,1300389	0,20688
V2src03_hrc04_525.yuv	3	4	–1,171212121	0	0,1293293	0,19158
V2src04_hrc05_525.yuv	4	5	–50,64090909	40,82526	0,6742005	0,6249
V2src04_hrc06_525.yuv	4	6	–28,05454545	32,552322	0,4250873	0,49999
V2src04_hrc07_525.yuv	4	7	–23,87575758	25,286598	0,3762656	0,40764
V2src04_hrc08_525.yuv	4	8	–16,60757576	19,86405	0,2972294	0,3485
V2src05_hrc05_525.yuv	5	5	–31,86969697	30,812616	0,4682559	0,47645
V2src05_hrc06_525.yuv	5	6	–18,56515152	21,413895	0,3203024	0,3646
V2src05_hrc07_525.yuv	5	7	–8,154545455	15,446437	0,2071702	0,306
V2src05_hrc08_525.yuv	5	8	–4,006060606	10,836051	0,1652752	0,26662
V2src06_hrc05_525.yuv	6	5	–41,63181818	37,342789	0,5690291	0,56967
V2src06_hrc06_525.yuv	6	6	–29,48787879	26,660055	0,4370961	0,42391
V2src06_hrc07_525.yuv	6	7	–22,25909091	20,878248	0,3591788	0,35896
V2src06_hrc08_525.yuv	6	8	–12,03181818	16,896168	0,2482169	0,31941
V2src07_hrc05_525.yuv	7	5	–23,89545455	19,086998	0,3796362	0,34067
V2src07_hrc06_525.yuv	7	6	–10,15606061	10,69402	0,2276934	0,26548
V2src07_hrc07_525.yuv	7	7	–4,240909091	4,896546	0,1644409	0,22267
V2src07_hrc08_525.yuv	7	8	–5,98030303	1,555055	0,1819566	0,20099
V2src08_hrc09_525.yuv	8	9	–76,2	52,094177	0,9513387	0,83024
V2src08_hrc10_525.yuv	8	10	–61,34545455	47,395226	0,789748	0,7397
V2src08_hrc11_525.yuv	8	11	–66,02575758	52,457584	0,8405916	0,83753

TABLEAU 5

Données objectives et subjectives pour un système à 525 lignes

Nom du fichier	Séquence source (SRC)	Circuit fictif de référence (HRC)	Note subjective moyenne brute	Note prévue par le modèle sur la base des données brutes	Note subjective moyenne corrigée	Note prévue par le modèle sur la base de données corrigées
V2src08_hrc12_525.yuv	8	12	–37,20454545	37,931854	0,5221555	0,57874
V2src08_hrc13_525.yuv	8	13	–31,23030303	30,95985	0,4572049	0,4784
V2src08_hrc14_525.yuv	8	14	–31,26818182	33,293602	0,4614104	0,51031
V2src09_hrc09_525.yuv	9	9	–64,42878788	54,414772	0,8262912	0,87746
V2src09_hrc10_525.yuv	9	10	–49,92878788	36,080425	0,660339	0,55061
V2src09_hrc11_525.yuv	9	11	–53,73181818	46,338791	0,7100111	0,72031
V2src09_hrc12_525.yuv	9	12	–34,36969697	23,21393	0,4921708	0,38409
V2src09_hrc13_525.yuv	9	13	–22,85454545	16,955978	0,3656559	0,31998
V2src09_hrc14_525.yuv	9	14	–16,41666667	13,694396	0,2960957	0,29046
V2src10_hrc09_525.yuv	10	9	–72,11212121	48,179104	0,9084171	0,75433
V2src10_hrc10_525.yuv	10	10	–43,11666667	30,703861	0,5908784	0,475
V2src10_hrc11_525.yuv	10	11	–56,11969697	52,63887	0,7302376	0,84118
V2src10_hrc12_525.yuv	10	12	–19,55909091	21,95225	0,3345703	0,37033
V2src10_hrc13_525.yuv	10	13	–12,34393939	16,23988	0,2565459	0,31328
V2src10_hrc14_525.yuv	10	14	–16,05	23,201355	0,2953144	0,38395
V2src11_hrc09_525.yuv	11	9	–50,40454545	36,394535	0,6675853	0,55531
V2src11_hrc10_525.yuv	11	10	–54,26212121	37,812542	0,7054929	0,5769
V2src11_hrc11_525.yuv	11	11	–41,73636364	44,128036	0,5761193	0,68087
V2src11_hrc12_525.yuv	11	12	–19,03939394	14,619688	0,32761	0,29857
V2src11_hrc13_525.yuv	11	13	–17,72121212	14,12041	0,310495	0,29417
V2src11_hrc14_525.yuv	11	14	–19,4969697	14,927424	0,331051	0,30132
V2src12_hrc09_525.yuv	12	9	–61,35	40,051254	0,7883371	0,61229
V2src12_hrc10_525.yuv	12	10	–46,84545455	31,128973	0,6295301	0,48066
V2src12_hrc11_525.yuv	12	11	–51,80151515	41,77285	0,6809288	0,6406
V2src12_hrc12_525.yuv	12	12	–22,51969697	20,868282	0,3651402	0,35886
V2src12_hrc13_525.yuv	12	13	–14,17878788	15,040992	0,2714356	0,30234
V2src12_hrc14_525.yuv	12	14	–14,6030303	13,521517	0,2782449	0,28896
V2src13_hrc09_525.yuv	13	9	–55,25	38,691498	0,7211194	0,5906
V2src13_hrc10_525.yuv	13	10	–39,55	33,054504	0,5545722	0,50696
V2src13_hrc11_525.yuv	13	11	–40,03939394	45,9454	0,5525494	0,71318
V2src13_hrc12_525.yuv	13	12	–14	16,631002	0,2708744	0,31692
V2src13_hrc13_525.yuv	13	13	–14,33181818	15,113959	0,27549	0,30299
V2src13_hrc14_525.yuv	13	14	–14,31969697	16,611286	0,2733771	0,31674

TABLEAU 6

Données objectives et subjectives pour un système à 625 lignes

Nom du fichier	SRC	HRC	Note subjective moyenne brute	Note prévue par le modèle sur la base des données brutes	Note subjective moyenne corrigée	Note prévue par le modèle sur la base de données corrigées
V2src1_hrc2_625.yuv	1	2	38,85185185	31,764214	0,59461	0,47326
V2src1_hrc3_625.yuv	1	3	42,07407407	21,868561	0,64436	0,36062
V2src1_hrc4_625.yuv	1	4	23,77777778	12,195552	0,40804	0,27239
V2src1_hrc6_625.yuv	1	6	18,14814815	9,169512	0,34109	0,24887
V2src1_hrc8_625.yuv	1	8	12,92592593	6,738072	0,2677	0,23128
V2src1_hrc10_625.yuv	1	10	11,88888889	2,553883	0,26878	0,20356
V2src2_hrc2_625.yuv	2	2	33,51851852	31,492788	0,54173	0,46985
V2src2_hrc3_625.yuv	2	3	46,48148148	31,1313	0,70995	0,46535
V2src2_hrc4_625.yuv	2	4	13,33333333	20,241726	0,27443	0,34432
V2src2_hrc6_625.yuv	2	6	8,814814815	17,39045	0,22715	0,31721
V2src2_hrc8_625.yuv	2	8	7,074074074	14,914576	0,21133	0,29513
V2src2_hrc10_625.yuv	2	10	3,407407407	7,352309	0,16647	0,23562
V2src3_hrc2_625.yuv	3	2	48,07407407	38,852715	0,73314	0,56845
V2src3_hrc3_625.yuv	3	3	50,66666667	38,244621	0,76167	0,55982
V2src3_hrc4_625.yuv	3	4	32,11111111	27,733229	0,49848	0,42454
V2src3_hrc6_625.yuv	3	6	22,33333333	24,80323	0,38613	0,39159
V2src3_hrc8_625.yuv	3	8	16,33333333	23,296747	0,34574	0,37544
V2src3_hrc10_625.yuv	3	10	11,96296296	16,33028	0,26701	0,30759
V2src4_hrc2_625.yuv	4	2	36,14814815	42,041592	0,58528	0,61514
V2src4_hrc3_625.yuv	4	3	55,03703704	49,283836	0,90446	0,72942
V2src4_hrc4_625.yuv	4	4	39,7037037	38,322186	0,62361	0,56091
V2src4_hrc6_625.yuv	4	6	38,03703704	36,863457	0,61143	0,54053
V2src4_hrc8_625.yuv	4	8	24,40740741	32,46579	0,43329	0,48214
V2src4_hrc10_625.yuv	4	10	12,88888889	25,918123	0,26548	0,40388
V2src5_hrc2_625.yuv	5	2	38,62962963	38,95779	0,61973	0,56995
V2src5_hrc3_625.yuv	5	3	44,18518519	40,076313	0,68987	0,58609
V2src5_hrc4_625.yuv	5	4	24,66666667	23,166002	0,41648	0,37406
V2src5_hrc6_625.yuv	5	6	23,62962963	20,592213	0,4218	0,34778
V2src5_hrc8_625.yuv	5	8	12,40740741	13,763152	0,27543	0,28531
V2src5_hrc10_625.yuv	5	10	7,37037037	8,418313	0,2022	0,24332
V2src6_hrc2_625.yuv	6	2	22,48148148	33,810165	0,38852	0,49949
V2src6_hrc3_625.yuv	6	3	27,07407407	25,004984	0,44457	0,39379
V2src6_hrc4_625.yuv	6	4	13,18518519	20,889347	0,27983	0,35074

TABLEAU 6

Données objectives et subjectives pour un système à 625 lignes

Nom du fichier	SRC	HRC	Note subjective moyenne brute	Note prévue par le modèle sur la base des données brutes	Note subjective moyenne corrigée	Note prévue par le modèle sur la base de données corrigées
V2src6_hrc6_625.yuv	6	6	14,44444444	17,418222	0,28106	0,31747
V2src6_hrc8_625.yuv	6	8	8,740740741	15,486559	0,23726	0,30011
V2src6_hrc10_625.yuv	6	10	5,518518519	11,509192	0,17793	0,2669
V2src7_hrc4_625.yuv	7	4	39,25925926	45,231079	0,59953	0,66412
V2src7_hrc6_625.yuv	7	6	33,85185185	43,131519	0,55093	0,63163
V2src7_hrc9_625.yuv	7	9	27,07407407	39,506535	0,45163	0,57784
V2src7_hrc10_625.yuv	7	10	19,25925926	34,418381	0,35617	0,50749
V2src8_hrc4_625.yuv	8	4	15,85185185	40,408993	0,32528	0,59095
V2src8_hrc6_625.yuv	8	6	17,03703704	38,552574	0,32727	0,56418
V2src8_hrc9_625.yuv	8	9	14,85185185	35,577034	0,30303	0,52297
V2src8_hrc10_625.yuv	8	10	11,48148148	30,278536	0,26366	0,45484
V2src9_hrc4_625.yuv	9	4	28,96296296	30,515778	0,47656	0,45775
V2src9_hrc6_625.yuv	9	6	30,51851852	26,971027	0,49924	0,41577
V2src9_hrc9_625.yuv	9	9	19,66666667	23,351355	0,39101	0,37601
V2src9_hrc10_625.yuv	9	10	20,92592593	17,856861	0,37122	0,32152
V2src10_hrc4_625.yuv	10	4	40,33333333	43,640377	0,70492	0,63942
V2src10_hrc6_625.yuv	10	6	37,33333333	40,552502	0,58218	0,59305
V2src10_hrc9_625.yuv	10	9	30,92592593	36,747391	0,49711	0,53893
V2src10_hrc10_625.yuv	10	10	21,2962963	30,161013	0,37854	0,45341
V2src11_hrc1_625.yuv	11	1	50,25925926	55,909908	0,79919	0,84263
V2src11_hrc5_625.yuv	11	5	35,51851852	44,049999	0,59256	0,64572
V2src11_hrc7_625.yuv	11	7	18,7037037	26,877754	0,34337	0,4147
V2src11_hrc10_625.yuv	11	10	15,07407407	23,420477	0,30567	0,37674
V2src12_hrc1_625.yuv	12	1	36,33333333	43,837097	0,61418	0,64244
V2src12_hrc5_625.yuv	12	5	38,44444444	40,349903	0,6661	0,59008
V2src12_hrc7_625.yuv	12	7	31,11111111	37,254383	0,53242	0,54594
V2src12_hrc10_625.yuv	12	10	26,14814815	28,953564	0,44737	0,43887
V2src13_hrc1_625.yuv	13	1	43,7037037	38,333649	0,74225	0,56108
V2src13_hrc5_625.yuv	13	5	43,2962963	34,290554	0,66799	0,5058
V2src13_hrc7_625.yuv	13	7	25,2962963	26,990025	0,42065	0,41598
V2src13_hrc10_625.yuv	13	10	15,88888889	20,181463	0,33381	0,34373

Annexe 3

Modèle 2

TABLE DES MATIÈRES

1 Introduction

Depuis toujours, on utilise pour évaluer la qualité vidéo un certain nombre d'évaluateurs qui évaluent subjectivement la qualité vidéo. L'évaluation peut être faite avec ou sans séquence vidéo de référence. Dans une évaluation avec séquence de référence, on montre aux évaluateurs deux séquences vidéo: la séquence vidéo de référence (source) et la séquence vidéo traitée qui sera comparée avec la séquence vidéo source. En comparant les deux séquences vidéo, les évaluateurs attribuent des notes subjectives à chacune d'elles. Par conséquent, on parle souvent de test subjectif de qualité vidéo. Le test subjectif est considéré comme la méthode la plus précise étant donné qu'il reflète la perception de l'homme, mais il comporte plusieurs limitations. Tout d'abord il suppose la présence d'un certain nombre d'évaluateurs. Il est donc chronophage et coûteux. Par conséquent, il ne peut être fait en temps réel. On s'est donc beaucoup intéressé à l'élaboration de méthodes objectives de mesure de la qualité vidéo. Un critère important pour une méthode objective de

mesure de la qualité vidéo est que cette méthode donne des résultats cohérents pour toute une série de séquences vidéo qui ne sont pas utilisées au stade de la conception. Dans cette optique, on a élaboré un modèle facile à mettre en oeuvre, suffisamment rapide pour des mises en oeuvre en temps réel et résistant à toute une série de dégradations vidéo. Ce modèle est un produit élaboré conjointement par Yonsei University, SK Telecom et Radio Research Laboratory, République de Corée.

2 Mesure objective de la qualité vidéo basée sur la dégradation des contours

2.1 Rapport PSNR basé sur la dégradation des contours (EPSNR)

Le modèle de mesure objective de la qualité vidéo est une méthode avec une image de référence complète. En d'autres termes, on suppose qu'une séquence vidéo de référence est fournie. En analysant comment les êtres humains perçoivent la qualité vidéo, on observe que le système visuel humain est sensible aux dégradations autour des contours. En d'autres termes lorsque les zones des contours d'une séquence vidéo sont floues, les évaluateurs ont tendance à donner à cette séquence de mauvaises notes même si l'erreur quadratique moyenne globale est faible. On observe en outre que les algorithmes de compression vidéo ont tendance à produire davantage de défauts (artéfacts) autour des zones des contours. Sur la base de cette observation, le modèle fournit une méthode de mesure objective de la qualité vidéo qui permet de mesurer les dégradations autour des contours. Dans ce modèle on applique tout d'abord un algorithme de détection des bords à la séquence vidéo source pour localiser les zones des bords. Ensuite on mesure la dégradation de ces zones des bords en calculant l'erreur quadratique moyenne. A partir de cette erreur on calcule le rapport EPSNR, rapport que l'on utilise comme mesure de la qualité vidéo après post-traitement.

Dans le modèle, il faut tout d'abord appliquer un algorithme de détection de contours pour localiser les régions des contours. On peut utiliser n'importe quel algorithme de détection de contours même s'il peut y avoir des différences minimes dans les résultats. Par exemple, on peut utiliser n'importe quel opérateur gradient pour localiser les régions des contours. Un certain nombre d'opérateurs gradient ont été proposés. Dans de nombreux algorithmes de détection de contours, on calcule tout d'abord à l'aide d'opérateur gradient l'image du gradient horizontal $g_{horizontal}(m,n)$ et l'image du gradient vertical $g_{vertical}(m,n)$. On peut ensuite calculer l'image du gradient d'amplitude $g(m,n)$ comme suit:

$$g(m,n) = \left|g_{horizontal}(m,n)\right| + \left|g_{vertical}(m,n)\right|$$

Enfin, on applique un seuillage à l'image du gradient d'amplitude $g(m,n)$ pour trouver les régions des contours. En d'autres termes, les pixels dont les gradients d'amplitude dépassent une valeur seuil sont considérés comme étant les régions des contours.

Les Fig. 7 à 11 illustrent cette procédure. La Fig. 7 montre une image source. La Fig. 8 montre une image du gradient horizontal $g_{horizontal}(m,n)$, laquelle est obtenue par application d'un opérateur gradient horizontal à l'image source de la Fig. 7. La Fig. 9 montre une image du gradient vertical $g_{vertical}(m,n)$, laquelle est obtenue par application d'un opérateur gradient vertical à l'image source de la Fig. 7. La Fig. 10 montre l'image du gradient d'amplitude (image des contours) et la Fig. 11 l'image binaire des contours (image de masquage), lesquelles sont obtenues par application d'un seuillage à l'image du gradient d'amplitude de la Fig. 10.

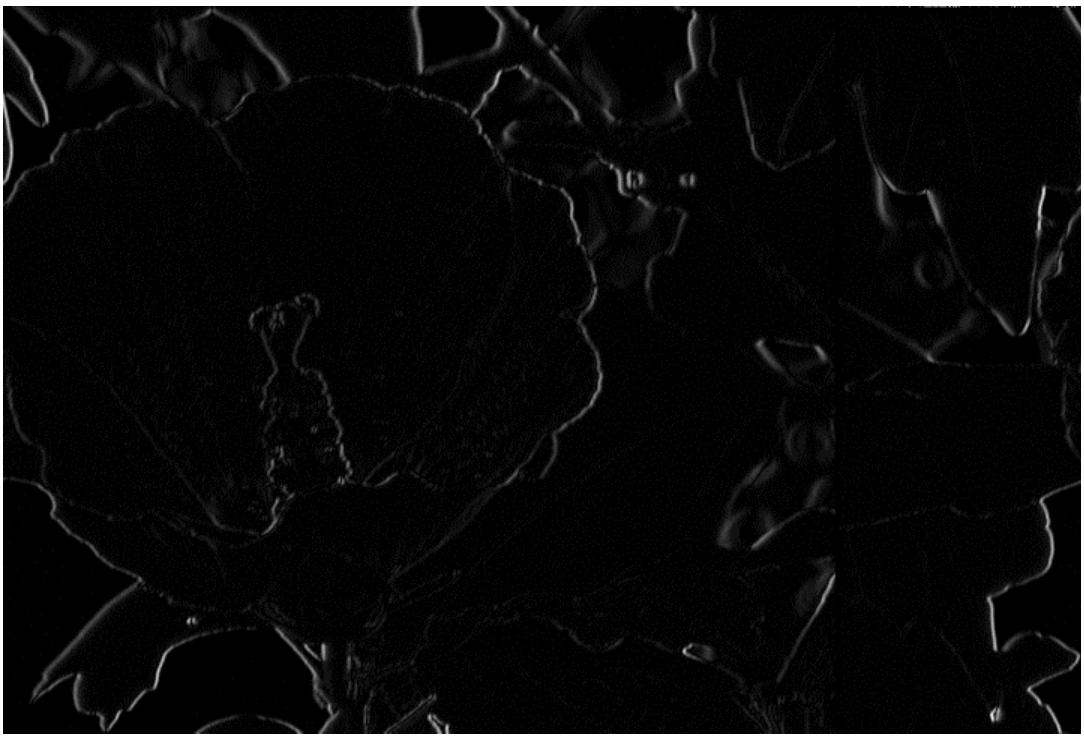
FIGURE 7

Image source (image d'origine)

1683-07

FIGURE 8

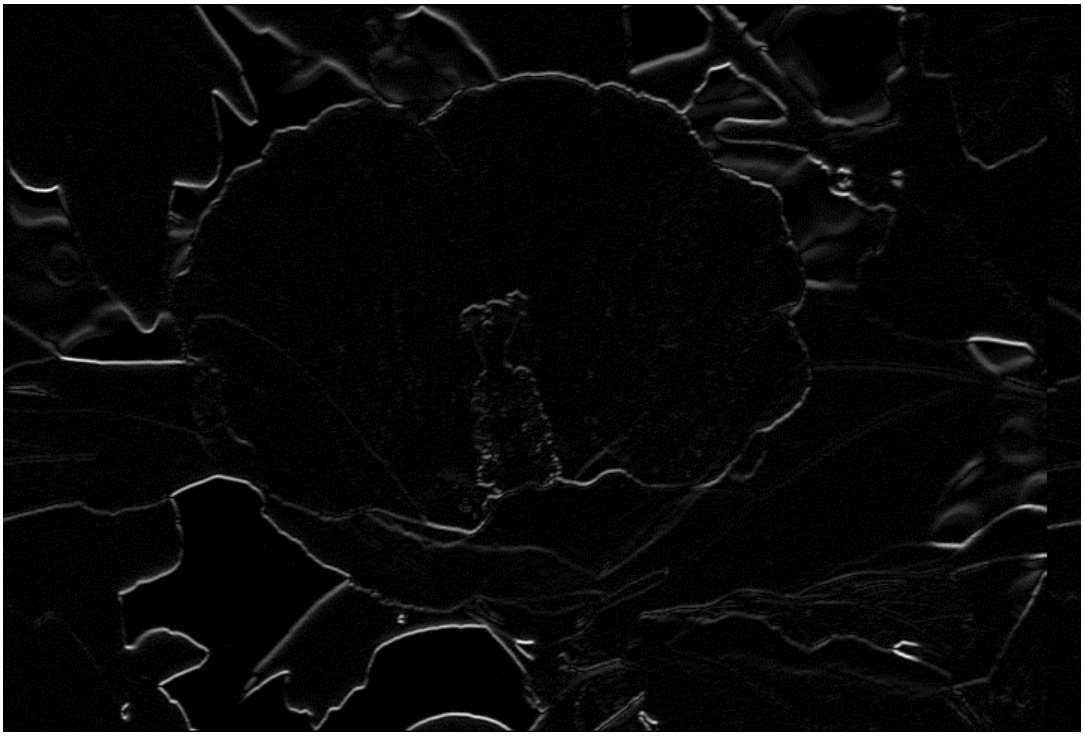
Image du gradient horizontal, laquelle est obtenue par application d'un opérateur gradient horizontal à l'image source de la Fig. 7



1683-08

FIGURE 9

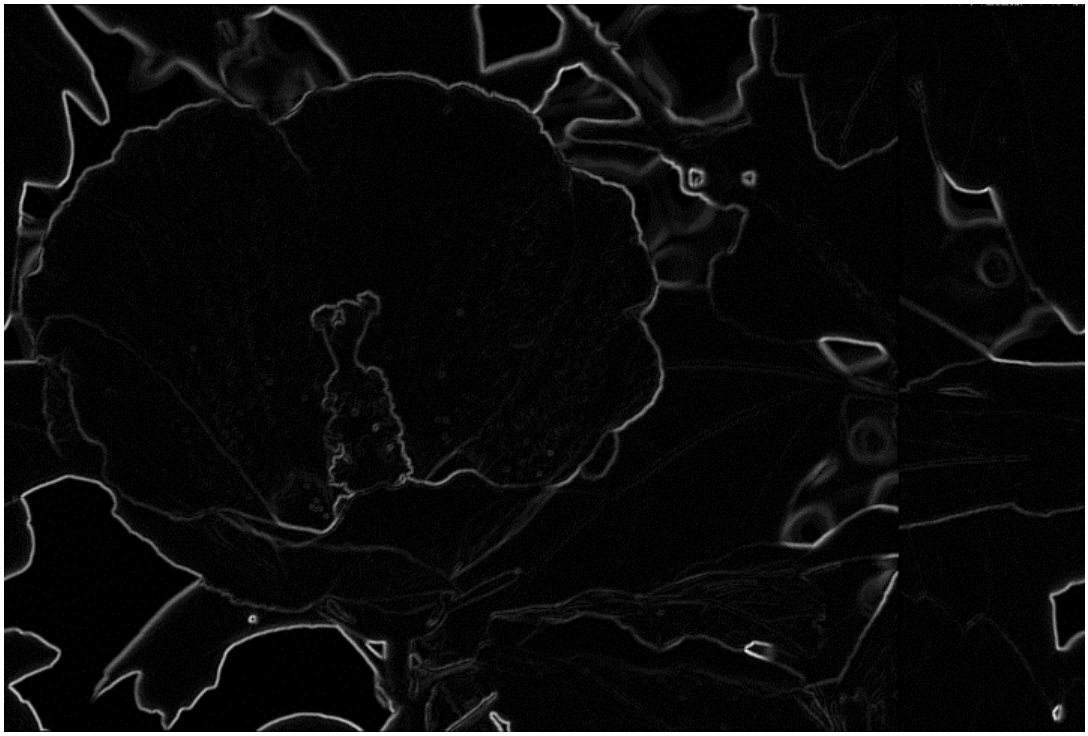
Image de gradient vertical, laquelle est obtenue par application d'un opérateur gradient vertical à l'image source de la Fig. 7



1683-09

FIGURE 10

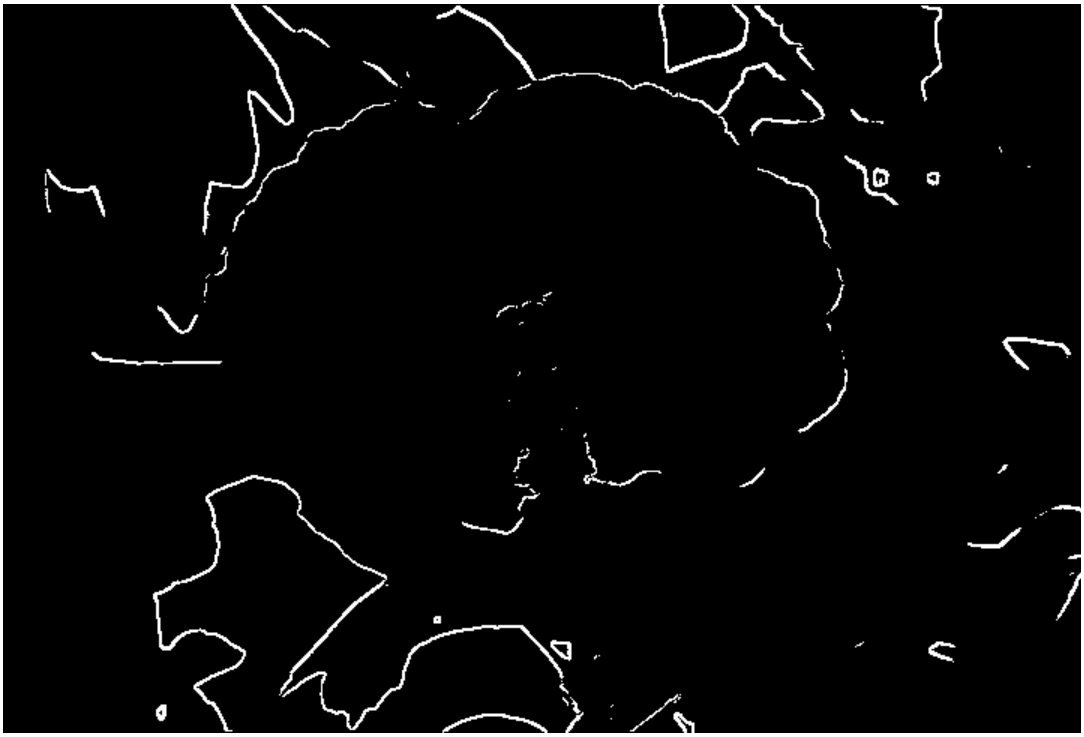
Image du gradient d'amplitude



1683-10

FIGURE 11

Image binaire des contours (image masque) obtenue par application d'une opération de seuillage à l'image du gradient d'amplitude de la Fig. 10



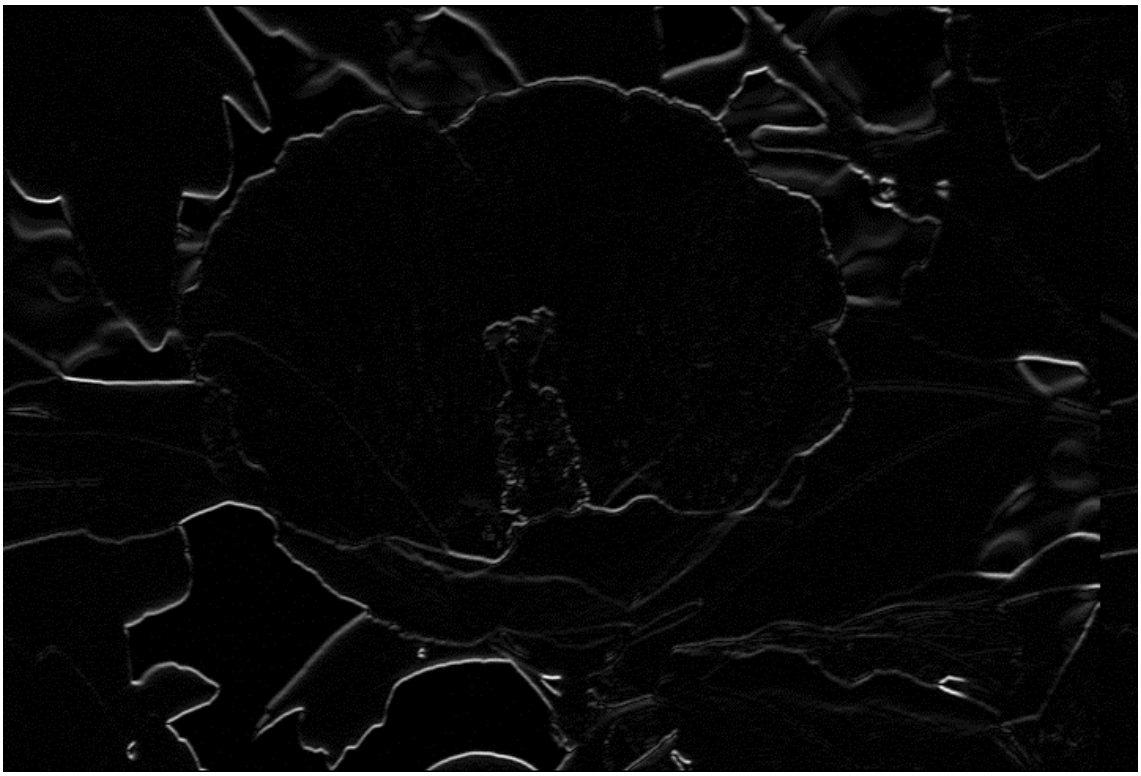
1683-11

On peut également utiliser une procédure modifiée pour localiser les régions des contours. Par exemple, on peut tout d'abord appliquer un opérateur gradient vertical à l'image source, ce qui donne l'image du gradient vertical. On applique ensuite un opérateur gradient horizontal à l'image du gradient vertical, ce qui donne une image du gradient successif modifié (image du gradient horizontal et du gradient vertical). Enfin, on peut appliquer un seuillage à l'image du gradient successif modifié pour trouver les régions des contours. En d'autres termes, les pixels de l'image du gradient successif modifié qui dépassent une valeur seuil sont considérés comme étant les zones des contours. Les Fig. 12 à 15 illustrent la procédure modifiée. La Fig. 12 montre une image du gradient vertical $g_{vertical}(m,n)$, laquelle est obtenue par application d'un opérateur gradient vertical à l'image source de la Fig. 7. La Fig. 13 montre une image du gradient successif modifié (image du gradient horizontal et du gradient vertical), laquelle est obtenue par application d'un opérateur gradient horizontal à l'image du gradient vertical de la Fig.12. La Fig. 14 montre l'image binaire des contours (image masque) obtenue par application d'un seuillage à l'image du gradient successif modifié de la Fig. 13.

On notera que les deux méthodes peuvent être considérées comme un algorithme de détection de contours. On peut choisir n'importe quel algorithme de détection de contours selon la nature des séquences vidéo et des algorithmes de compression. Toutefois, certaines méthodes peuvent donner de meilleurs résultats que d'autres.

FIGURE 12

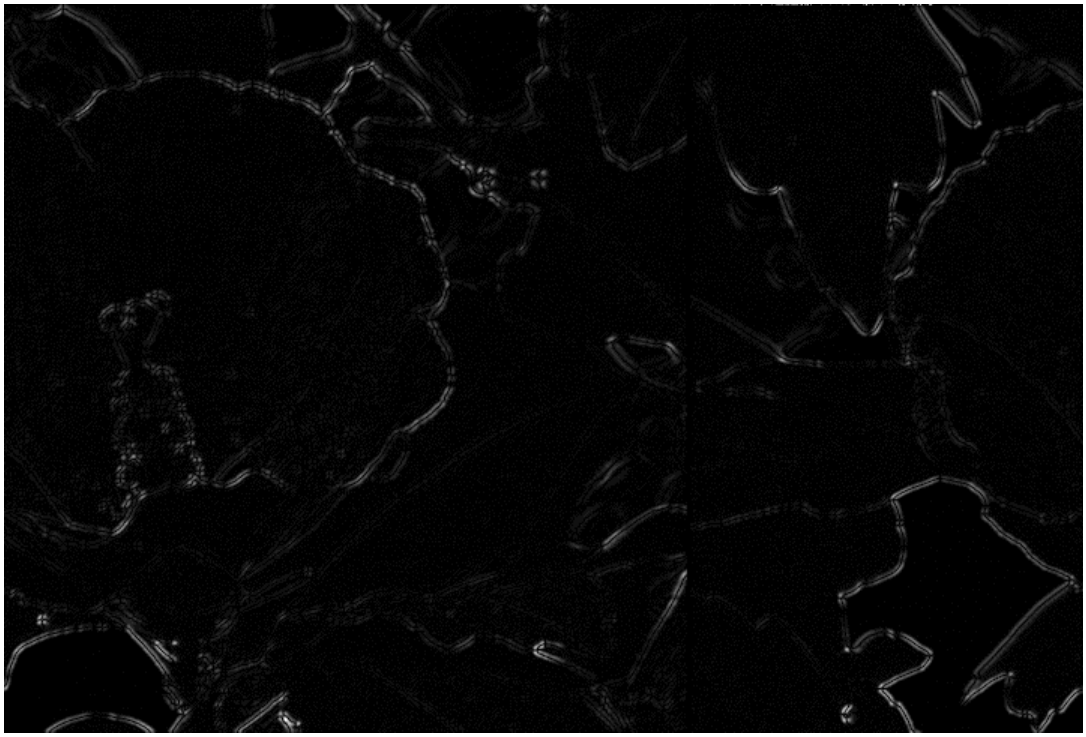
Image du gradient vertical, laquelle est obtenue par application d'un opérateur gradient vertical à l'image source de la Fig. 7



1683-12

FIGURE 13

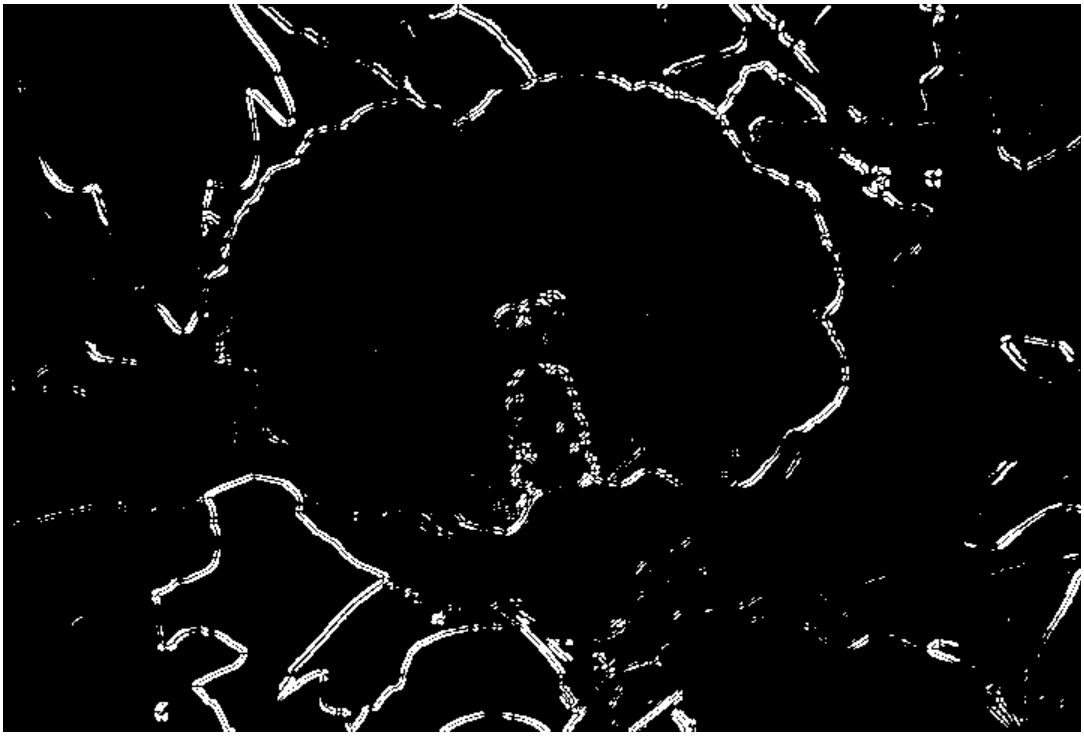
Image des gradients successifs modifiée (image des gradients horizontal et vertical), laquelle est obtenue par application d'un opérateur gradient horizontal à l'image du gradient vertical de la Fig. 12



1683-13

FIGURE 14

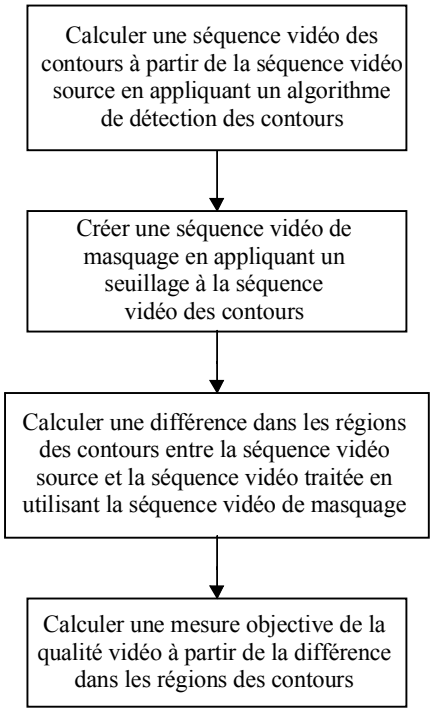
Image binaire des contours (image masque), obtenue par application d'un seuil à l'image des gradients successifs modifiée de la Fig. 13



1683-14

FIGURE 15

Schéma fonctionnel d'un rapport EPSNR



1683-15

Ainsi, dans le modèle, on applique tout d'abord un opérateur de détection de contours; ce qui permet d'obtenir des images des contours (voir les Fig. 10 et 13). Ensuite, on crée une image de masquage (image binaire des contours) en appliquant un seuillage à l'image des contours (voir les Fig. 11 et 14). En d'autres termes, les pixels de l'image des contours dont la valeur est inférieure au seuil, t_e, sont mis à zéro et les pixels dont la valeur est égale ou supérieure à ce seuil sont positionnés à une valeur autre que zéro. Les Fig. 11 et 14 donnent des exemples d'images de masquage. On notera que cet algorithme de détection des contours est appliqué à l'image source. On peut appliquer l'algorithme de détection des contours aux images traitées mais il est plus exact de l'appliquer aux images source. Etant donné qu'une séquence vidéo peut être considérée comme une séquence d'images ou de trames, la procédure susmentionnée peut être appliquée à chaque image ou à chaque trame de séquence vidéo. Etant donné que le modèle peut être utilisé pour des séquences vidéo composées de trames ou d'images, on utilisera le terme «d'image» pour parler indifféremment de trame ou d'image.

Ensuite, on calcule les différences entre la séquence vidéo source et la séquence vidéo traitée correspondant aux pixels ayant une valeur autre que zéro de l'image de masquage. En d'autres termes, l'erreur quadratique des régions des contours de la $l^{\text{ième}}$ trame est calculée comme suit:

$$se_e^l = \sum_{i=1}^{M}\sum_{j=1}^{N}\{S^l(i,j) - P^l(i,j)\}^2 \quad \text{si } \left|R^l(i,j)\right| \neq 0 \tag{75}$$

où:

$S^l(i,j)$: $l^{\text{ième}}$ image de la séquence vidéo source

$P^l(i,j)$: $l^{\text{ième}}$ image de la séquence vidéo traitée

$R^l(i,j)$: $l^{\text{ième}}$ image de la séquence vidéo de masquage

M: nombre de rangées

N: nombre de colonnes.

Lorsque le modèle est mis en oeuvre, on peut sauter la génération de la séquence vidéo de masquage. En fait, sans créer la séquence vidéo de masquage, l'erreur quadratique des régions des contours de la *l*ième image est calculée comme suit:

$$se_e^l = \sum_{i=1}^{M}\sum_{j=1}^{N}\{S^l(i,j) - P^l(i,j)\}^2 \quad \text{si } \left|Q^l(i,j)\right| \geq t_e \tag{76}$$

où:

$Q^l(i,j)$: $l^{\text{ième}}$ image de la séquence vidéo des contours

t_e: un seuil.

L'erreur quadratique moyenne est utilisée à l'équation (75) pour calculer la différence entre la séquence vidéo source et la séquence vidéo traitée mais on peut utiliser tout autre type de différence. Par exemple, on peut également utiliser la différence absolue. Dans le modèle soumis aux tests VQEG Phase II, t_e a été mis à 260 et l'algorithme de détection des contours modifié a été utilisé avec l'opérateur de Sobel.

Cette procédure est répétée pour l'ensemble des séquences vidéo et l'erreur quadratique moyenne des contours est calculée comme suit:

$$mse_e = \frac{1}{K}\sum_{l=1}^{L} se_e^l \tag{77}$$

où:

L: nombre d'images (trames ou images)

K: nombre total de pixels des contours.

Enfin, le rapport PSNR des zones des contours (EPSNR) est calculé comme suit:

$$EPSNR = 10 \log_{10}\left(\frac{P^2}{mse_e}\right) \tag{78}$$

où:

P: valeur crête des pixels.

Dans le modèle, ce rapport EPSNR est utilisé comme note objective de base de la qualité vidéo. La Fig. 15 donne un schéma fonctionnel de calcul du rapport EPSNR.

2.2 Postajustements

2.2.1 Désaccentuation d'un rapport EPSNR élevé

Lorsque le rapport EPSNR a une valeur supérieure à 35, il surestime, semble-t-il, la qualité perceptuelle. On utilise par conséquent la mise à l'échelle linéaire par paliers suivante:

$$EPSNR = \begin{cases} EPSNR & \text{si} \quad 0 \le EPSNR \le 35 \\ EPSNR \times 0{,}9 & \text{si } 35 < EPSNR \le 40 \\ EPSNR \times 0{,}8 & \text{si} \quad EPSNR > 40 \end{cases} \tag{79}$$

2.2.2 Prise en considération de contours floutés

On observe que lorsque les contours sont très flous dans des séquences vidéo de qualité médiocre, les évaluateurs ont tendance à donner des notes subjectives médiocres. En d'autres termes, si les régions des contours de la séquence vidéo traitée sont nettement plus petites que celles de la séquence vidéo source, les évaluateurs donnent de moins bonnes notes. Par ailleurs, on observe que certaines séquences vidéo ont un très petit nombre de pixels ayant des composantes haute fréquence. En d'autres termes, le nombre de pixels des régions des contours est très faible. Pour tenir compte de ces problèmes, les régions des contours de la séquence vidéo source et de la séquence vidéo traitée sont calculées et le rapport EPSNR est modifié comme suit:

$$MEPSNR = \begin{cases} MEPSNR - 60 \times \left(0{,}1{,}225 - \left(\frac{EP_{common}}{EP_{src}}\right)^2\right) \\ MEPSNR \end{cases}$$

$$\text{si} \quad EPNSR < 25 \quad \text{et} \quad \left(\frac{EP_{common}}{EP_{scr}}\right)^2 < 0{,}35 \tag{80}$$

$$\text{et} \quad \frac{EP_{hrc}}{EP_{src}} < 0{,}13 \text{ dans les autres cas}$$

où:

$MEPSNR$: EPSNR modifié.

EP_{common}: nombre total de pixels des contours communs dans les séquences vidéo SRC et HRC (c'est-à-dire pixels des contours apparaissant au même endroit)

EP_{src}: nombre total des pixels des contours dans la séquence vidéo (source) SRC.

Pour certaines séquences vidéo, EP_{src} peut être très faible. Si EP_{src} est inférieur à 10 000 pixels (environ 10 000/240 = 41,7 pixels par trame pour des séquences vidéo 525 lignes de 8 s et d'environ 10 000/200 = 50 pixels par trame pour des séquences vidéo 625 lignes de 8 s), l'utilisateur peut réduire le seuil t_e dans l'équation (76) de 20 jusqu'à ce que EP_{src} soit supérieur ou égal à 10 000 pixels. Si EP_{src} est inférieur à 10 000 pixels même lorsque t_e est réduit à 80, on ne procède pas au postajustement à l'aide de l'équation (80). Dans ce cas, on calcule le rapport EPSNR en utilisant $t_e = 60$. Si cette option est retenue, l'utilisateur peut supprimer la condition $EP_{hrc}/EP_{src} < 0{,}13$ dans l'équation (80).

2.2.3 Mise à l'échelle

Ensuite, les notes objectives sont remises à l'échelle de façon à être comprises entre 0 (non distinguable de séquence vidéo d'origine) et 1.

$$VQM = 1 - MEPSNR \times 0{,}02 \tag{81}$$

Cette mesure VQM est utilisée comme la note objective du modèle.

2.3 Précision d'alignement

La précision d'alignement recommandée pour le modèle est une précision d'un demi-pixel dans les séquences vidéo entrelacées, ce qui équivaut à une précision d'un quart de pixel dans un format vidéo progressif. L'interpolation spline cubique [Lee et autres, 1998], voire mieux, est fortement recommandée pour calculer les valeurs des sous-pixels.

2.4 Schéma fonctionnel du modèle

La Fig. 16 donne le schéma fonctionnel complet du modèle.

3 Données objectives

Le modèle a été appliqué aux données vidéo des tests VQEG Phase I[1] [Document UIT-R 6Q/14 (septembre 2003) – Final Report from the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment, Phase II (FR-TV2).]. Toutefois, une fois le modèle soumis, des erreurs d'alignement et d'opérateur ont été relevées. Les données objectives présentées dans la présente Annexe sont les mêmes que celles indiquées dans le Rapport final Phase II du VQEG. Par conséquent, lorsque la méthode décrite dans la présente Annexe est correctement mise en oeuvre, l'utilisateur peut obtenir différentes données objectives de la présente Annexe. Les Tableaux 7 et 8 donnent les données objectives pour les ensembles de données vidéo 525 et 625 lignes.

4 Conclusion

Un nouveau modèle de mesure objective de la qualité vidéo basé sur la dégradation des contours est proposé. Ce modèle est extrêmement rapide. Une fois la représentation binaire générée, le modèle est plusieurs fois plus rapide que le rapport PSNR classique, d'où une amélioration importante. Par conséquent, le modèle convient bien pour les applications qui nécessitent une évaluation de la qualité vidéo en temps réel.

5 Références bibliographiques

LEE, C., EDEN, M. et UNSER, M. [1998] High quality image resizing using oblique projection operators. *IEEE Trans. Image Processing*, Vol. 5, p. 679-692.

FIGURE 16

Schéma fonctionnel complet du modèle

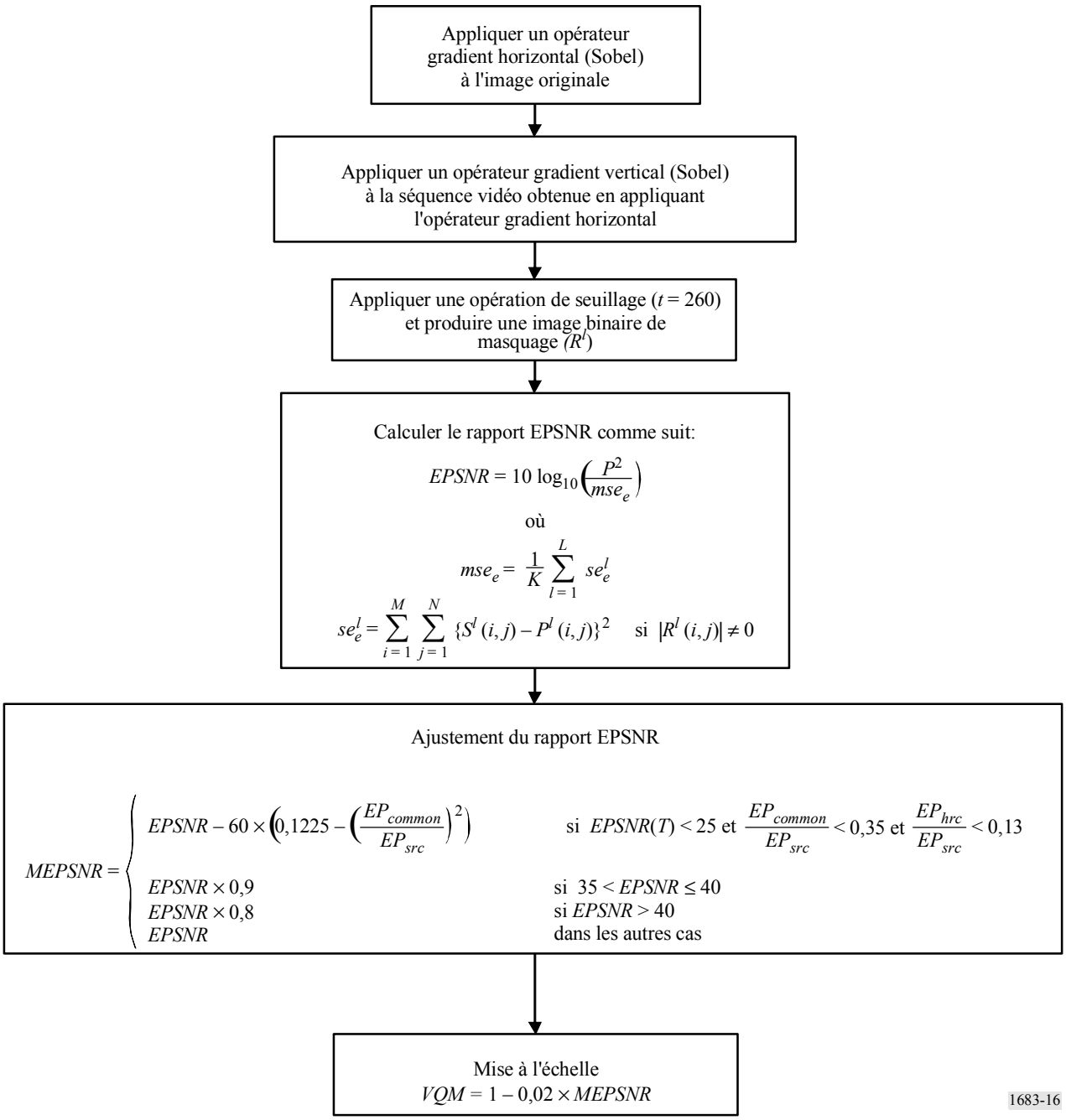


TABLEAU 7

Matrice VQM à 525 lignes (yonsei_1128c.exe)[1]

SRC (Image)	HRC																											
	1		2		3		4		5		6		7		8		9		10		11		12		13		14	
1	1	0,679	4	0,525	7	0,512	10	0,419																				
2	2	0,431	5	0,365	8	0,313	11	0,342																				
3	3	0,558	6	0,452	9	0,340	12	0,305																				
4									13	0,668	17	0,581	21	0,556	25	0,535												
5									14	0,543	18	0,485	22	0,443	26	0,410												
6									15	0,631	19	0,477	23	0,441	27	0,411												
7									16	0,467	20	0,415	24	0,376	28	0,346												
8																	29	0,787	35	0,734	41	0,740	47	0,551	53	0,520	59	0,537
9																	30	0,848	36	0,559	42	0,723	48	0,495	54	0,462	60	0,465
10																	31	0,552	37	0,449	43	0,542	49	0,352	55	0,308	61	0,377
11																	32	0,610	38	0,628	44	0,633	50	0,475	56	0,471	62	0,498
12																	33	0,576	39	0,539	45	0,577	51	0,470	57	0,436	63	0,448
13																	34	0,554	40	0,569	46	0,517	52	0,399	58	0,382	64	0,412

[1] Une fois le modèle soumis, des erreurs d'alignement et d'opérateur ont été relevées. Les données objectives présentées dans la présente Annexe sont les mêmes que celles qui figurent dans le Rapport final des tests VQEG Phase II. Par conséquent, lorsque la méthode décrite dans la présente Annexe est correctement mise en oeuvre, l'utilisateur peut obtenir des données objectives différentes de celles figurant dans le Tableau 7.

TABLEAU 8

Matrice VQM à 625 lignes (yonsei_1128c.exe)[1]

SRC (Image)	HRC																			
	1		2		3		4		5		6		7		8		9		10	
1			4	0,612	10	0,531	16	0,452			29	0,434			42	0,436			52	0,382
2			5	0,544	11	0,540	17	0,451			30	0,437			43	0,440			53	0,363
3			6	0,572	12	0,571	18	0,497			31	0,479			44	0,478			54	0,418
4			7	0,601	13	0,656	19	0,557			32	0,547			45	0,526			55	0,472
5			8	0,603	14	0,621	20	0,500			33	0,492			46	0,444			56	0,390
6			9	0,591	15	0,520	21	0,483			34	0,469			47	0,461			57	0,423
7							22	0,576			35	0,555					48	0,531	58	0,501
8							23	0,512			36	0,500					49	0,482	59	0,457
9							24	0,507			37	0,487					50	0,468	60	0,436
10							25	0,610			38	0,594					51	0,575	61	0,540
11	1	0,753							26	0,594			39	0,508					62	0,485
12	2	0,643							27	0,556			40	0,550					63	0,496
13	3	0,669							28	0,524			41	0,481					64	0,441

(1) Une fois le modèle soumis, des erreurs d'alignement et d'opérateur ont été relevées. Les données objectives présentées dans la présente Annexe sont les mêmes que celles qui figurent dans le Rapport final des tests VQRG Phase II. Par conséquent, lorsque la méthode décrite dans la présente Annexe est correctement mise en oeuvre, l'utilisateur peut obtenir des données objectives différentes de celles figurant dans le Tableau 8.

Annexe 4

Modèle 3

TABLE DES MATIÈRES

1 Introduction

La présente Annexe présente une méthode d'évaluation de la qualité vidéo à l'aide de paramètres objectifs basés sur une segmentation de l'image. Les scènes naturelles sont segmentées en différentes zones (plans, contours et texture) et un ensemble de paramètres objectifs est attribué à chacune de ces zones. On définit un modèle perceptuel qui permet de donner des notes subjectives en calculant la corrélation entre les mesures objectives (Recommandation UIT-R BT.500 et Recommandation UIT-R BT.802 – Images et séquences pour l'évaluation subjective des codecs numériques véhiculant des signaux produits conformément à la Recommandation UIT-R BT.601) et les résultats des tests d'évaluation subjective, et ce pour un ensemble de scènes naturelles traitées par des codecs vidéo MPEG-2. Dans ce modèle, la corrélation entre chaque paramètre objectif traité par plusieurs systèmes de compression (par exemple des codecs MPEG-2 et MPEG-1) et le niveau

de dégradation subjective est approximée par une courbe logistique, et on obtient une estimation du niveau de dégradation pour chaque paramètre. Le résultat final est une combinaison linéaire des niveaux de dégradation estimés, la pondération de chaque niveau de dégradation étant proportionnelle à sa fiabilité statistique.

Au § 2, une description générale du système d'évaluation des images basée sur la segmentation (CPqD-IES, *image evaluation based on segmentation*) est présentée. Au § 3, les mesures à suivre pour comptabiliser les défauts d'alignement spatial ou temporel ainsi que la correction de gain sont décrites. Au § 4, l'algorithme de segmentation des images en zones de plans, de contours et de texture est expliqué. Au § 5, la mesure objective pour chaque zone et chaque composante d'image est décrite. Au § 6, on décrit la façon dont la base de données des modèles de dégradation a été établie. Les calculs des paramètres sont eux aussi décrits dans ce même paragraphe. Le § 7 décrit comment on estime l'indice de qualité vidéo à partir des paramètres figurant dans la base de données des modèles de dégradation. L'Annexe 4a présente les résultats des valeurs d'indice de qualité vidéo (VQR, *video quality rating*) qui ont été évaluées pendant les tests Phase II du VQEG[1].

2 Description générale du système IES

La Fig. 17 donne une description générale de l'algorithme CPqD-IES pour des scènes naturelles. Chaque scène naturelle est représentée par une scène originale (de référence) ***O*** et une scène dégradée ***I*** qui résulte de l'application d'un codec à la scène ***O***. Des corrections de décalage et de gain sont appliquées à la scène ***I*** pour créer une scène dégradée corrigée ***I′***, de telle sorte que chaque trame f de ***I′*** correspond à la trame de référence f de ***O*** pour $f = 1, 2, ..., n$ (voir le § 3.2).

Les scènes d'entrée ***I*** et ***O*** pour l'algorithme CPqD-IES ont un format YCbCr4:2:2 conformément à la Recommandation UIT-R BT.601 – Paramètres de codage en studio de la télévision numérique pour des formats standards d'image 4:3 (normalisé) et 16:9 (écran panoramique).

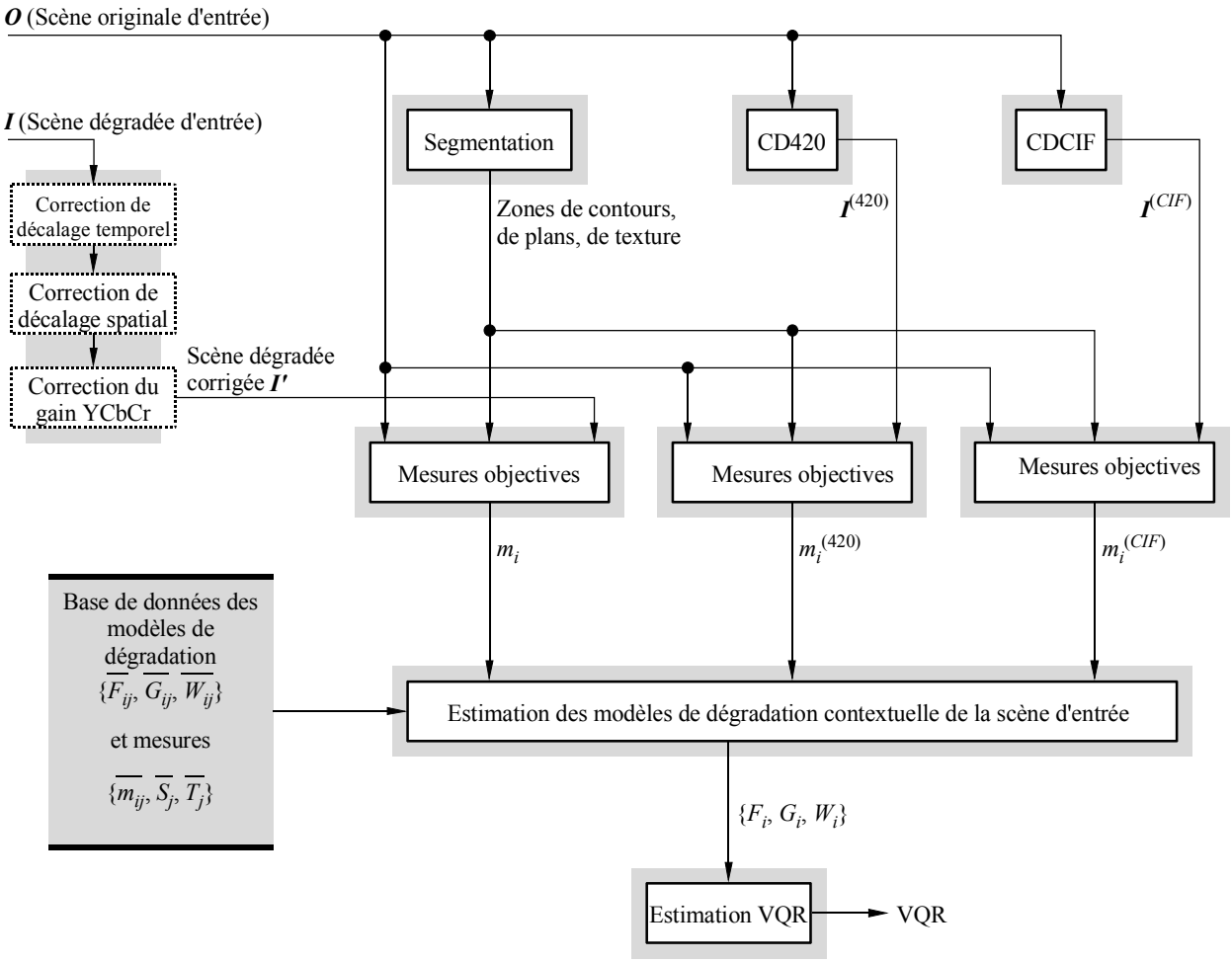
La composante Y de chaque trame d'image f de ***O*** est segmentée en trois catégories: texture, contours et plans (voir le § 4). Une mesure objective est calculée à partir de la différence entre les trames correspondantes de ***O*** et ***I′*** pour chacune de ces zones et pour chaque composante d'image YCbCr, formant ainsi un ensemble de 9 mesures objectives $\{m1, m2, ..., m9\}$ pour chaque trame d'image f (voir le § 5). Pour chaque mesure objective m_i, $i = 1, 2, ..., 9$, on obtient un niveau de dégradation contextuelle, L_i, basé sur son modèle d'estimation de la dégradation qui est donné par:

$$L_i = 100 / \left[1 + \left(\frac{F_i}{m_i} \right)^{G_i} \right] \tag{82}$$

où F_i et G_i sont deux paramètres calculés (voir le § 7) à partir d'une base de données de modèles de dégradation (voir le § 6), de l'attribut spatial S et de l'attribut temporel F (voir le § 5) et des mesures objectives $m_i^{(420)}$ et $m_i^{(CIF)}$ pour la trame f résultant des opérations des codecs CD420 et CDCIF appliqués à ***O*** (voir le § 7). Les deux références des codecs de dégradation, CD420 (codeur/décodeur MPEG-2 4:2:0) et CDCIF (codeur/décodeur MPEG-1 CIF), sont totalement basées sur les routines extraites directement des MPEG2 (Recommandation UIT-T H.262 – Technologies de l'information – Codage générique des images animées et du son associé: données vidéo) et MPEG1 [ISO/CEI, 1992], disponibles sur www.mpeg.org/MPEG/MSSG. Dans la mise en oeuvre actuelle de l'algorithme CPqD-IES, ces routines fonctionnent en intra mode avec quantification fixe par paliers de 16. Il importe de noter que les codecs CD420 et CDCIF n'introduisent pas de différence de décalage ou de gain par rapport à ***O***.

FIGURE 17

Description générale de l'algorithme CPqD-IES



L'indice de qualité vidéo VQR_f de la trame f est obtenu par combinaison linéaire des niveaux de dégradation contextuelle L_i, $i = 1, 2, ..., 9$, comme suit:

$$VQR_f = \sum_{i=1}^{9} W_i \, . \, L_i \tag{83}$$

où W_i est le facteur de pondération du niveau de dégradation L_i pour cette scène naturelle particulière, lequel est calculé comme indiqué au § 7.

Maintenant, la séquence de valeurs VQR_1, VQR_2, ... VQR_n est transformée par un filtre médian de taille 3 en une autre séquence VQR'_1, VQR'_2, ... VQR'_n en ne calculant pas la valeur médiane dans le voisinage immédiat de VQR_1 et VQR_n. Pendant le filtrage médian, l'algorithme évite la répétition de deux valeurs médianes consécutives, c'est-à-dire que si la valeur médiane VQR'_{f-1} calculée à 1 unité près de VQR_f est égale à la valeur médiane VQR'_{f-2} calculée dans le voisinage 1×1 de VQR'_{f-1}, l'algorithme choisit VQR'_{f-1} comme valeur minimale calculée dans le voisinage 1×1 de VQR_f.

Cet algorithme peut être décrit comme suit:

1) Pour chaque f compris entre 2 et $n - 1$,

2) Calculer med, la valeur médiane entre VQR_{f-1}, VQR_f, VQR_{f+1};

3) Si $med = VQR'_{f-2}$ alors

4) Calculer VQR'_{f-1} comme la valeur minimale entre VQR_{f-1}, VQR_f, VQR_{f+1};

5) Sinon

6) $VQR'_{f-1} \leftarrow med$.

L'indice de qualité vidéo final VQR est alors la moyenne des valeurs VQR'_f.

$$VQR = \frac{1}{n-2} \cdot \sum_{f=1}^{n-2} VQR'_f \tag{84}$$

Les équations (82), (83) et l'algorithme ci-dessus décrivent le processus permettant d'estimer la valeur VQR à partir des modèles de dégradation contextuelle $\{F_i, G_i, W_i\}$ et des mesures objectives m_i, $i = 1, 2, ..., 9$. Les paragraphes suivants terminent la description de la méthode en présentant les détails à l'intérieur des blocs restants de la Fig. 17.

3 Correction du décalage et du gain

3.1 Décalage temporel

Le décalage temporel, dt, est un entier compris entre –2 et 2. Les scènes d'entrée présentant des décalages temporaires situés en dehors de cette fourchette ne sont pas prises en considération. Supposons que $\boldsymbol{I}_{dt}$ est la scène dégradée $\boldsymbol{I}$ avec un déplacement de f trames. Un coefficient de dissemblance entre $\boldsymbol{O}$ et chaque scène déplacée de $\boldsymbol{I}_{dt}$ est calculé. Le déplacement avec un coefficient de dissemblance le plus faible est utilisé comme décalage temporel et le résultat $\boldsymbol{I}_{dt}$ est ensuite $\boldsymbol{I}$ déplacé de ce décalage en vue du prochain calcul. Le coefficient de dissemblance entre la scène $\boldsymbol{O}$ et la scène $\boldsymbol{I}_{dt}$ est obtenu comme suit, où n est le nombre de trames situées dans l'intersection temporelle entre elles:

1) $\xi_T \leftarrow 0$

2) Pour chaque f de 1 à n,

3) Calculer S_b;

4) Calculer S'_b;

5) Calculer D_b;

6) Calculer μ, valeur moyenne des pixels en Db

7) $\xi_T \leftarrow \xi_T + (\mu/n)$

8) Retour ξ_T (coefficient de dissemblance entre $\boldsymbol{O}$ et $\boldsymbol{I}_{dt}$).

Où:

S_b: amplitude du gradient de Sobel de la composante Y de la $f^{\text{ième}}$ trame de $\boldsymbol{O}$

S'_b: amplitude du gradient de Sobel de la composante Y de la $f^{\text{ième}}$ trame de $\boldsymbol{I}_{dt}$

D_b: différence absolue au niveau des pixels entre S_b et S'_b.

3.2 Décalage spatial

Le décalage spatial (d_x, d_y) est l'un des déplacements entiers horizontaux et verticaux suivants $d_x = -6, -5, \ldots 6$ et $d_y = -6, -5, \ldots, 6$. Soit $\boldsymbol{I}_{dx,dy}$ scène dégradée $\boldsymbol{I}_{dt}$ avec toutes les trames déplacées de (d_x, d_y) pixels. On calcule un coefficient de dissemblance entre $\boldsymbol{O}$ et $\boldsymbol{I}_{dx,dy}$. Le déplacement spatial présentant la plus faible dissemblance est utilisé comme décalage spatial et le résultat $\boldsymbol{I}_{dx,dy}$ est alors $\boldsymbol{I}_{dt}$ déplacé de ce décalage, en vue du prochain calcul.

La dissemblance entre $\boldsymbol{O}$ et $\boldsymbol{I}_{dx,dy}$ est décrite ci-après:

1) $\xi_S \leftarrow 0$; $c \leftarrow 0$

2) Pour chaque f de 1 à n

3) Pour x de $x0$ à ($x0 + w/4$)

4) Pour y de $y0$ à ($y0 + h/4$)

5) $\xi_S \leftarrow \xi_S + |Y(4x,4y) - Y'(4x + dx, 4y + dy)| +$
$+ |Cb(4x,4y) - Cb'(4x + dx, 4y + dy)| +$
$+ |Cr(4x,4y) - Cr'(4x + dx, 4y + dy)|$

6) $c \leftarrow c + 3$

7) $\xi_S \leftarrow \xi_S / c$

8) Retour ξ_S (coefficient de dissemblance entre $\boldsymbol{O}$ et $\boldsymbol{I}_{dx,dy}$)

Où:

$w \times h$: dimensions de la zone d'intersection entre $\boldsymbol{O}$ et $\boldsymbol{I}_{dx,dy}$;

$Y(x, y)$, $Cb(x, y)$, $Cr(x, y)$: valeurs dans les composantes d'image d'une trame f de $\boldsymbol{O}$ pour un pixel (x, y);

$Y'(x + dx, y + dy)$
$Cb'(x + dx, y + dy)$
$Cr'(x + dx, y + dy)$: valeurs dans les composantes d'image d'une trame f de $\boldsymbol{I}_{dx,dy}$ pour un pixel $(x + dx, y + dy)$.

3.3 Gain

Le gain d'amplitude entre $\boldsymbol{O}$ et $\boldsymbol{I}_{dx,dy}$ est calculé séparément pour chaque composante d'image Y, C_B et C_R. L'algorithme calcule la moyenne des gains sur l'ensemble des n trames et corrige chaque composante d'image en conséquence. Le résultat $\boldsymbol{I}'$ est la scène dégradée qui est utilisée pour tous les calculs ultérieurs. Le gain d'amplitude entre une composante d'image C de la trame f de $\boldsymbol{I}_{dx, dy}$ par rapport à la même composante C de la trame f de $\boldsymbol{O}$ est obtenu en floutant les deux images C' et C au moyen d'un filtre gaussien [Gonzalez et Woods, 1992] de noyau:

$$\begin{pmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{pmatrix}$$

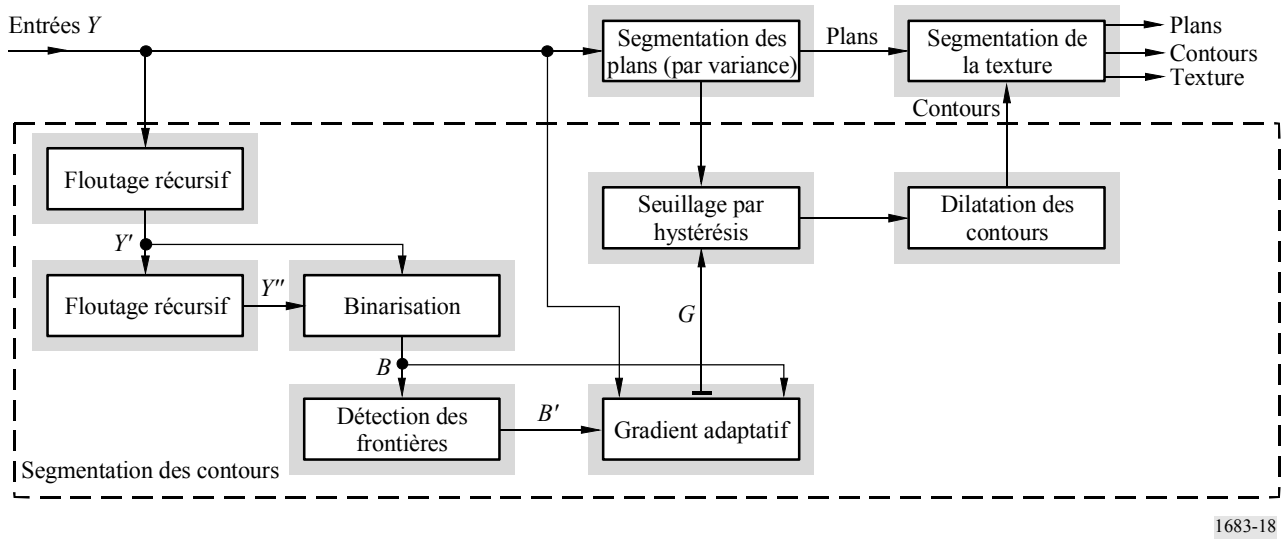
et en calculant le rapport entre la somme de leurs valeurs de pixels dans les images floutées. Un seul de chacun des 16 pixels est pris en considération (en balayant les images à composante floutées par incréments horizontal et vertical de 4 pixels, comme dans l'algorithme de calcul de ξ_S présenté au § 3.2).

4 Segmentation de l'image

Au départ, l'algorithme de segmentation classe chaque pixel de la composante *Y* d'une trame donnée *f* de la scène originale ***O*** dans la région du plan et/ou une autre région. L'algorithme applique également à *Y* un détecteur de contours et la région des contours est définie par les contours qui sont situés dans les limites de la région du plan. La région de texture est composée par les pixels restants de l'image *Y* (voir la Fig. 18).

FIGURE 18

Schéma fonctionnel du processus de segmentation



La segmentation est calculée sur chaque trame de la composante *Y* à partir de la scène originale d'entrée ***O***. Pour les composantes C_B et C_R, les régions sont mises en évidence par la position des pixels dans la composante *Y*, après échantillonnage ascendant dans C_B et C_R.

4.1 Régions du plan

La variance de la brillance de chaque pixel de la composante *Y* est calculée dans un voisinage de 5×5 pixels de part et d'autre du pixel considéré. Une opération de seuillage est appliquée à la variance de l'image de sorte que les pixels présentant une valeur de variance inférieure à 25^2 sont classés comme appartenant à la région du plan. Il résulte de ce processus que de petites composantes de pixels sont classées, à tort, dans la zone de texture. Un filtre médian de 3×3 est appliqué pour supprimer ces petites composantes. Enfin, l'image binaire des régions du plan est soumise à une dilatation morphologique en utilisant un élément structurant circulaire d'un diamètre de 11 pixels [Gonzales et Woods, 1992].

4.2 Régions des contours

Un filtrage récursif est appliqué à *Y*, créant une première image floue *Y′*, puis à *Y′* pour créer une deuxième image floue *Y″*. Chaque filtrage récursif se compose de quatre grilles appliquées à l'image d'entrée. Cet algorithme est décrit ci-après pour la composante d'image *Y* de la seule trame de la scène d'entrée ***O***.

1) Pour y variant de 0 à $(h - 1)$

2) Pour x variant de 0 à $(w - 2)$

3) $Y(x + 1, y) \leftarrow Y(x, y) + 0{,}7\,(Y(x + 1, y) - Y(x, y)])$

4) Pour y variant de 0 à $(h - 1)$

5) Pour x variant de $(w - 1)$ à 1

6) $Y(x - 1, y) \leftarrow Y(x, y) + 0{,}7\,[Y(x - 1, y) - Y(x, y)]$

7) Pour x variant de 0 à $(w - 1)$

8) Pour y variant de 0 à $(h - 2)$

9) $Y(x, y + 1) \leftarrow Y(x, y) + 0{,}7\,[Y(x, y + 1) - Y(x, y)]$

10) Pour x variant de 0 à $(w - 1)$

11) Pour y variant de $(h - 1)$ à 1

12) $Y(x, y + 1) \leftarrow Y(x, y) + 0{,}7\,[Y(-1) - Y(x, y)]$

13) Sauvegarder image Y en image Y'.

Où:

$Y(x, y)$: brillance du pixel (x, y)

h: nombre de lignes de Y

w: nombre de colonnes de Y.

La seconde application de l'algorithme créera Y''. Une image binaire B est créée à partir de Y' et Y'':

$$B(x, y) = \begin{cases} 1 & \text{si } Y'(x, y) \geq Y''(x, y) \\ 0 & \text{sinon} \end{cases} \tag{85}$$

Après cela, l'algorithme identifie les pixels frontières des zones de B ayant une valeur de pixel de 1 en créant une seconde image binaire B':

$$B'(x, y) = \begin{cases} 1 & \text{si } B(x,y) = 1 \text{ et } B(x', y') = 0 \text{ pour tout pixel } (x', y') \in N_8(x, y) \\ 0 & \text{sinon} \end{cases} \tag{86}$$

où $N_8(x, y)$ est l'ensemble des pixels (x', y') dans le voisinage 3 × 3 de (x, y) (c'est-à-dire ses 8 voisins immédiats).

Un filtre à gradient adaptatif est appliqué à Y limité aux pixels où $B'(x, y) = 1$:

$$G(x, y) = \begin{cases} |\mu_1 - \mu_0| & \text{si } B'(x, y) = 1 \\ 0 & \text{sinon} \end{cases} \tag{87}$$

où:

μ_1: valeur moyenne de $Y(x', y')$, pour tout $(x', y') \in N_8(x, y)$ de sorte que $B(x', y') = 1$

μ_0: valeur moyenne de $Y(x', y')$, pour tout $(x', y') \in N_8(x, y)$ de sorte $B(x', y') = 0$.

A noter que l'algorithme utilise B en lieu et place de B' pour calculer les valeurs moyennes μ_1 et μ_0.

Un seuillage par hystérésis [-Trucco et Verri, 1998] est appliqué à G limité aux pixels dont on a établi au § 4.1 qu'ils appartiennent à la zone des plans. Le seuil inférieur est de 30 et le seuil supérieur de 40. L'algorithme identifie tout d'abord les pixels de G, de sorte que $G(x, y) > 40$, puis applique un algorithme à zone croissante le long des lignes de G en utilisant ces pixels comme éléments de départ et en restreignant la croissante aux pixels appartenant à la même ligne pour lesquels $G(x, y) > 30$. Toutes les composantes 4 connexes comportant moins de 6 pixels sont éliminées de ce résultat. L'image binaire finale est dilatée par un élément structurant circulaire d'un diamètre de 5 pixels qui ignore la restriction à la zone des plans. Les pixels de valeur 1 dans cette dilatation sont classés comme appartenant à la zone des contours.

4.3 Région de texture

La région de texture se compose des pixels de *Y* qui ont été classés comme n'appartenant ni à la région des contours ni à la région du plan.

5 Mesures objectives

Soit S_b l'image d'amplitude du gradient de Sobel calculée pour une composante donnée (*Y, C_B* ou *C_R*) d'une trame donnée *f* de la scène originale $\boldsymbol{O}$, et S'_b, l'image d'amplitude du gradient de Sobel pour la même composante de la trame *f* de la scène dégradée $\boldsymbol{I'}$. L'image D_b de la différence absolue au niveau des pixels entre S_b et S'_b est calculée et la zone $\Re$ de pixels de l'image D_b qui appartient à un contexte donné (plan, contours ou texture) est prise en considération. La différence absolue de Sobel (ASD, *absolute Sobel difference*) pour cette composante d'image et ce contexte est définie comme étant la moyenne des valeurs de pixels de l'image D_b restreinte à $\Re$.

Cette procédure donne un ensemble de neuf mesures objectives $\{m_1, m_2, ..., m_9\}$ pour chaque trame d'image *f*, $f = 1, 2, ..., n$, tenant compte de l'ensemble des trois contextes et des trois composantes d'image.

Le même processus est appliqué pour créer des mesures objectives $\{m_1^{(420)}, m_2^{(420)}, ..., m_9^{(420)}\}$ et $\{m_1^{(CIF)}, m_2^{(CIF)}, ..., m_9^{(CIF)}\}$, pour la trame *f*, avec un fonctionnement des MPEG-2 4:2:0 et MPEG-1 CIF CODEC sur la scène $\boldsymbol{O}$ (voir la Fig. 17). Ces mesures servent de références avec les attributs spatial *S* et temporel *T* pour déterminer le modèle de dégradation contextuelle pour $\boldsymbol{I'}$ (§ 7). L'attribut temporel *T* est la valeur moyenne de la différence absolue au niveau des pixels entre les segmentations des trames *f* et *f*–1, normalisée dans l'intervalle [0,1]. L'attribut spatial *S* est défini comme étant le rapport $m_7^{(CIF)}/m_7^{(420)}$, normalisé dans l'intervalle [0,1], où $m_7^{(CIF)}$ et $m_7^{(420)}$ sont les différences ASD correspondantes pour la région de texture de la composante *Y* de la trame *f*.

6 Base de données des modèles de dégradation

Le système IES utilise une base de données de modèles de dégradation pour des scènes différentes de la scène de référence $\boldsymbol{O}$ pour évaluer l'indice de qualité vidéo de $\boldsymbol{I'}$. Cette base de données regroupe des informations sur douze scènes à 60 Hz illustrant divers degrés de mouvement (scènes dynamiques ou statiques), de nature (scènes réelles ou scènes synthétiques), et de contexte (quantité de pixels de texture, de plan ou de contour). Cette base de données a été créée comme suit.

Les valeurs moyennes des mesures objectives $\{\overline{m}_{1,j}, \overline{m}_{2,j}, ..., \overline{m}_{9,j}\}$, $\{\overline{m}_{1,j}^{(420)}, \overline{m}_{2,j}^{(420)}, ..., \overline{m}_{9,j}^{(420)}\}$, et $\{\overline{m}_{1,j}^{(CIF)}, \overline{m}_{2,j}^{(CIF)}, ..., \overline{m}_{9,j}^{(CIF)}\}$ ont été calculées pour les trames de chaque scène *j*, $j = 1, 2, ..., 12$. Les valeurs de S_j et T_j ont été calculées comme étant la moyenne de l'attribut spatial et de l'attribut temporel (voir le § 5) sur les trames de chaque scène *j*. Toutes les scènes dégradées de la base de données ont elles aussi fait l'objet d'une évaluation subjective, ce qui donne un niveau de dégradation subjective SL_j, normalisé dans l'intervalle entre 0% et 100% pour chaque scène *j*.

Selon l'équation (82), chaque mesure objective $m_{i,j}$, $i = 1, 2, ..., 9$ et $j = 1, 2, ..., 12$, est rattachée à un niveau de dégradation contextuelle $L_{i,j}$. Les valeurs de $F_{i,j}$ et $G_{i,j}$ dans l'équation (82) ont été calculées pour chaque scène *j* en minimisant l'espérance de l'erreur quadratique moyenne $E[(\overline{SL}_j - \overline{L}_{i,j})^2]$. On a par ailleurs calculé les valeurs de $W_{i,j}$ dans l'équation (83) pour minimiser l'espérance de l'erreur quadratique moyenne:

$$E\left[\left(\overline{SL}_j - \sum_{i=1}^{9} \overline{W}_{i,j}\overline{L}_{i,j}\right)^2\right] \tag{88}$$

A l'issue du processus, la base de données des modèles de dégradation comprend 9 ensembles $\{\overline{F}_{i,j}, \overline{G}_{i,j}, \overline{W}_{i,j}, \overline{S}_j, \overline{T}_j\}$, $i = 1, 2, ..., 9$ de paramètres pour chaque scène j, $j = 1, 2, ..., 12$. Le Tableau 9 contient les valeurs de $\overline{S}_j$ et $\overline{T}_j$ à utiliser pour calculer les attributs $\{\overline{F}_{i,j}, \overline{G}_{i,j}, \overline{W}_{i,j}\}$.

TABLEAU 9

Attribut temporel *T* et attribut spatial *S*

Scène *j*	*T* (temporel)	S_Y (spatial *Y*)	S_{Cb} (spatial C_B)	S_{Cr} (spatial C_R)
1	27,01	36,79	25,20	38,01
2	25,33	26,08	5,93	67,99
3	45,54	60,97	10,28	28,75
4	36,40	30,47	6,46	63,07
5	32,02	72,50	11,72	15,78
6	12,63	84,22	2,85	12,94
7	28,38	61,53	11,08	27,39
8	10,19	46,08	5,45	48,47
9	0,01	5,89	5,07	89,03
10	7,26	4,75	2,00	93,25
11	7,60	69,16	9,41	21,43
12	14,27	69,61	3,89	26,50

7 Estimation des modèles de dégradation

Les modèles de dégradation contextuels pour une trame f de $\boldsymbol{I'}$ se composent des paramètres $\{F_i, G_i, W_i\}$ des équations (82) et (83), $i = 1, 2, ..., 9$. Le présent paragraphe décrit comment calculer ces paramètres en utilisant les scènes dégradées $I^{(420)}$ et $I^{(CIF)}$ comme référence.

7.1 Calcul de W_i

Les distances locales contextuelles $D_{i,j}$ entre une trame f des scènes dégradées $I^{(420)}$ et $I^{(CIF)}$, et chaque scène j de la base de données sont définies comme suit:

$$\begin{aligned}
\overline{L}_{i,j}^{(420)} &= 100 / \left[1 + \left(\overline{F}_{i,j} / \overline{m}_i^{(420)} \right)^{\overline{G}_{i,j}} \right] \\
\overline{L}_{i,j}^{(CIF)} &= 100 / \left[1 + \left(\overline{F}_{i,j} / \overline{m}_i^{(CIF)} \right)^{\overline{G}_{i,j}} \right] \\
L_{i,j}^{(420)} &= 100 / \left[1 + \left(\overline{F}_{i,j} / m_i^{(420)} \right)^{\overline{G}_{i,j}} \right] \\
L_{i,j}^{(CIF)} &= 100 / \left[1 + \left(\overline{F}_{i,j} / m_i^{(CIF)} \right)^{\overline{G}_{i,j}} \right]
\end{aligned} \tag{89}$$

$L_{i,j}^{(420)}$ et $L_{i,j}^{(CIF)}$ sont le niveau de dégradation estimé de la scène d'entrée ***O***, qui sont calculés avec les paramètres $\overline{F}_{i,j}$ et $\overline{G}_{i,j}$, dans le contexte *i*, des scènes *j* de la base de données.

$$D_{i,j} = \frac{1}{2}\left(\left|L_{i,j}^{(420)} - \overline{L}_{i,j}^{(420)}\right| + \left|L_{i,j}^{(CIF)} - \overline{L}_{i,j}^{(420)}\right|\right) \tag{90}$$

L'algorithme trouve l'ensemble Ω des six scènes les plus proches de la base de données sur la base de la distance $D_{i,j}$ et définit $W_{i,j}$ comme:

$$a_k = \begin{cases} 1 & \text{si (scene } k) \in \Omega \\ 0 & \text{sinon} \end{cases} \tag{91}$$

$$W_{i,j} = \frac{a_j D_{i,j}^{-1}}{\sum_{k=1}^{12} a_k D_{i,j}^{-1}} \tag{92}$$

Soit $i = \{1, 2, ..., 9\} \equiv \{$(*plane, Y*), (*plane,* C_B), (*plane,* C_R), (*edge, Y*), (*edge,* C_B), (*edge,* C_R), (*texture, Y*), (*texture,* C_B), (*texture,* C_R)$\}$, où (*edge, C*), (*plane, C*) et (*texture, C*) représentent les régions contours, plan et texture de la composante d'image *C*, $C = Y, C_B, C_R$.

Soit *u* = *texture*, *edge*, *plane* et $v = Y, C_B, C_R$, les valeurs W_i, $i = 1, 2, ..., 9$, sont calculées comme suit:

$$E_i = \sum_{j=1}^{12} D_{i,j}\, W_{i,j}$$

$$\kappa_{u,v} = \begin{cases} 1 & \text{si } v = Y_i \\ \dfrac{1}{2} & \text{sinon} \end{cases}$$

$$\tau = \sum_u \left[\frac{1}{E_{u,Y}} + \frac{1}{2}\left(\frac{1}{E_{u,C_B}} + \frac{1}{E_{u,C_R}}\right)\right]$$

$$W_i = \frac{\kappa_i}{\tau} \cdot \frac{1}{E_i} \tag{93}$$

7.2 Calcul de F_i et G_i

Les niveaux de dégradation contextuels $L_i^{(420)}$ et $L_i^{(CIF)}$ de la trame *f* pour CD420 et CDCIF sont calculés comme suit:

$$L_i^{(420)} = \frac{1}{\gamma} \cdot \sum_{j=1}^{12} W_{i,j}\; L_{i,j}^{(420)} \tag{94}$$

$$L_i^{(CIF)} = \frac{1}{\gamma} \cdot \sum_{j=1}^{12} W_{i,j}\; L_{i,j}^{(CIF)} \tag{95}$$

où γ est un facteur limité à [1/2, 2], qui est calculé à partir des distances vectorielles D_j entre les attributs spatial et temporel (voir le § 5), (S_j, T_j) et $(\overline{S}_j, \overline{T}_j)$, de la scène d'entrée et de chaque scène de la base de données, respectivement.

$$D_j = \left(S - \overline{S}_j\right)^2 + (T - \overline{T}_j)^2 \tag{96}$$

$$w_j = \frac{D_j^{-1}}{\sum_{k=1}^{12} D_k^{-1}}$$

$$a = \sum_{j=1}^{12} w_j \left[\frac{\overline{S}_j \cdot \overline{T}_j}{2} + (1 - \overline{T}_j^2) \cdot \left(1 - \frac{\overline{S}_j^2}{2} \right) \right] \tag{97}$$

$$b = \frac{ST}{2} + (1 - T^2) \cdot \left(1 - \frac{S^2}{2} \right)$$

$$\gamma = 1 + a - b$$

Les paramètres F_i et G_i sont enfin obtenus en résolvant le système d'équations ci-après:

$$L_i^{(420)} = 100 / \left[1 + \left(\frac{F_i}{m_i^{(420)}} \right)^{G_i} \right] \tag{98}$$

$$L_i^{(CIF)} = 100 / \left[1 + \left(\frac{F_i}{m_i^{(CIF)}} \right)^{G_i} \right] \tag{99}$$

8 Références bibliographiques

GONZALEZ, R. C. et WOODS, R. E. [1992] *Digital Image Processing*. Addison-Wesley.

ISO/CEI [1992] Norme ISO/IEC 11172 – Information technology – Coding of moving pictures and associated audio for digital storage media up to about 1,5 Mbit/s.

TRUCCO, E. et VERRI, A. [1998] *Introductory Techniques for 3-D Computer Vision*. Prentice-Hall.

Annexe 4a

Résultats objectifs des essais, Phase II du VQEG

TABLEAU 10

Matrice de données objectives brutes 625/60

SRC	HRC									
	1	2	3	4	5	6	7	8	9	10
1		0,6343	0,5083	0,287		0,2461		0,1951		0,1548
2		0,5483	0,5966	0,3649		0,3185		0,2668		0,1597
3		0,5998	0,6299	0,4551		0,3927		0,3428		0,2553
4		0,6055	0,8159	0,5684		0,5397		0,4158		0,309
5		0,6483	0,7268	0,4358		0,418		0,2874		0,1898
6		0,6146	0,4908	0,3671		0,3139		0,2562		0,2107
7				0,5865		0,5536			0,4841	0,3917
8				0,5023		0,457			0,3949	0,3158
9				0,4563		0,3927			0,3399	0,2667
10				0,7036		0,6511			0,6025	0,5083
11	0,8124				0,6374		0,3205			0,3221
12	0,7015				0,547		0,4997			0,3922
13	0,709	0,5098					0,4199			0,3298

TABLEAU 11

Matrice de données objectives brutes 525/60

SRC	HRC													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	0,5472	0,3698	0,3429	0,1918										
2	0,5075	0,226	0,1028	0,0789										
3	0,3549	0,127	0,058	0,0339										
4					0,6062	0,419	0,36	0,3108						
5					0,4444	0,2957	0,2152	0,1635						
6					0,6098[1]	0,3462	0,2546	0,1967						
7					0,2404	0,135	0,0864	0,0609						
8									0,8666	0,7554	0,6944	0,7048	0,6685	0,494
9									0,8896	0,7134	0,6204	0,6504	0,6246	0,2326
10									0,8776	0,6419	0,4788	0,6392	0,6237	0,1571
11									0,8623	0,7207	0,5719	0,5619	0,5796	0,3012
12									0,8262	0,6193	0,5139	0,5391	0,4946	0,1992
13									0,8223	0,5609	0,3454	0,437	0,4246	0,215

(1) La valeur SRC = 6, HRC = 5 a été tirée de l'analyse car elle dépassait les critères d'alignement temporel du plan d'essai VQEG.

Annexe 5

Modèle 4

La présente Annexe contient une description fonctionnelle complète du modèle VQM de la NTIA et des techniques d'étalonnage qui lui sont associées.

Les algorithmes d'étalonnage décrits dans la présente Annexe sont suffisants pour garantir un fonctionnement correct du dispositif d'évaluation de la qualité vidéo de la NTIA. Ils présentent généralement une précision d'alignement spatial de plus ou moins 1/2 pixel et une précision d'alignement temporel de plus ou moins une trame entrelacée.

TABLE DES MATIÈRES

1 Introduction

La présente Annexe contient une description technique complète du modèle général de la NTIA et des techniques d'étalonnage qui lui sont associées (par exemple évaluation et correction de l'alignement spatial, de l'alignement temporel et des erreurs de gain/décalage). Le modèle général correspond au modèle H dans les essais de télévision avec image de référence complète de Phase II du VQEG. Il était conçu pour être un modèle VQM universel pour des systèmes vidéo avec une très large plage de niveaux de qualité et de débits binaires. De nombreux essais subjectifs et objectifs ont été effectués afin de vérifier les performances du modèle général avant de le soumettre aux essais de Phase II du VQEG, lesquels ont uniquement porté sur l'évaluation des performances du modèle général pour des systèmes vidéo MPEG-2 et la Recommandation UIT-T H.263. Mais le modèle général devrait fonctionner correctement pour de nombreux autres types de systèmes de codage et de transmission.

Les algorithmes d'étalonnage décrits dans la présente Annexe sont suffisants pour garantir un fonctionnement correct du dispositif d'évaluation de la qualité vidéo. Ils présentent généralement une précision d'alignement spatial de plus ou moins 1/2 pixel et une précision d'alignement temporel de plus ou moins une trame entrelacée.

Le modèle général et les techniques d'étalonnage automatique qui lui sont associées ont été entièrement mis en œuvre sous forme de logiciel convivial. Toutes les parties intéressées peuvent accéder à ce logiciel sous réserve de l'acceptation d'un accord de licence d'évaluation gratuite (pour plus d'informations, on se reportera à l'adresse suivante: www.its.bldrdoc.gov/n3/video/vqmsoftware.htm).

2 Références normatives

Recommandation UIT-R BT.601 – Paramètres de codage en studio de la télévision numérique pour des formats standards d'image 4:3 (normalisé) et 16:9 (écran panoramique).

3 Définitions

4:2:2 – Format d'échantillonnage d'image Y, C_b, C_r pour lequel les plans de chrominance (C_b et C_r) sont échantillonnés horizontalement à une fréquence qui vaut la moitié de la fréquence d'échantillonnage du plan de luminance (Y). Voir la Recommandation UIT-R BT.601 (voir le § 2).

Alignement spatial: Processus utilisé pour évaluer et corriger les décalages spatiaux de la séquence vidéo traitée par rapport à la séquence vidéo d'origine.

Alignement temporel: Processus utilisé pour évaluer et corriger le décalage temporel (c'est-à-dire le retard vidéo) de la séquence vidéo traitée par rapport à la séquence vidéo d'origine (voir le § 6.4.1).

Big *YUV*: Format de fichier binaire utilisé pour stocker les clips qui ont été échantillonnés conformément à la Recommandation UIT-R BT.601. Dans ce format, toutes les images vidéo d'une scène sont stockées dans un seul grand fichier binaire, dans lequel chaque image est échantillonnée conformément à la Recommandation UIT-R BT.601. Y représente l'information de canal de luminance, U représente le canal de différence de couleur bleue (c'est-à-dire C_B dans la Recommandation UIT-R BT.601) et V représente le canal de différence de couleur rouge (c'est-à-dire C_R dans la Recommandation UIT-R BT.601). L'ordre des pixels dans le fichier binaire est le même que celui qui est spécifié dans le document 125M de la SMPTE [SMPTE, 1995]. La spécification complète du format de fichier Big *YUV* figure au § 5 et les routines logicielles permettant de lire et d'afficher des fichiers au format Big *YUV* sont données dans le document [Pinson et Wolf, 2002].

Caractéristique: Grandeur associée à – ou extraite d' – une sous-région spatio-temporelle d'un flux vidéo (d'origine ou traité).

Chrominance (*C*, C_B, C_R): Partie du signal vidéo qui achemine avant tout l'information de couleur (*C*), qui peut de plus être séparée en un signal de différence de couleur bleue (C_B) et un signal de différence de couleur rouge (C_R).

Circuit fictif de référence (HRC, *hypothetical reference circuit*): Système vidéo testé, par exemple un codec ou un système de transmission vidéonumérique.

Clip: Représentation numérique d'une scène qui est stockée sur support informatique.

Codec: Abréviation pour codeur/décodeur ou compresseur/décompresseur.

Coordonnées de rectangle: Sous-région d'image de forme rectangulaire qui est entièrement contenue dans le format de production et qui est spécifiée par quatre coordonnées (haut, gauche, bas, droite). La numérotation, qui commence à zéro, est telle que le coin (haut, gauche) de l'image échantillonnée a pour coordonnées (0,0). Voir le § 5.3.

Décalage ou décalage de niveau: Facteur additif appliqué par le HRC à tous les pixels d'un plan d'image donné (par exemple luminance, chrominance). Le décalage du signal de luminance est généralement appelé brillance.

Format de production: Grille d'image qui représente le format maximal possible de l'image pour un système standard donné. Le format de production représente le format souhaitable pour l'acquisition, la génération et le traitement de l'image, avant suppression. Pour les séquences vidéo échantillonnées selon la Recommandation UIT-R BT.601, le format de production est de 720 pixels × 486 lignes pour les systèmes à 525 lignes et de 720 pixels × 576 lignes pour les systèmes à 625 lignes [SMPTE, 1995b].

Format intermédiaire commun (CIF, *common intermediate format*): Structure d'échantillonnage vidéo utilisée en visioconférence, pour laquelle le canal de luminance est échantillonné à 352 pixels par 288 lignes (Recommandation UIT-T H.261 – Codec vidéo pour services audiovisuels à $p \times 64$ kbit/s).

Gain: Facteur multiplicatif appliqué par le circuit fictif de référence (HRC, *hypothetical reference circuit*) à tous les pixels d'un plan d'image donné (par exemple luminance, chrominance). Le gain du signal de luminance est généralement appelé contraste.

Groupe d'experts en qualité vidéo (VQEG, Video Quality Experts Group): Groupe d'experts internationaux en qualité vidéo qui réalisent des essais de validation de méthodes objectives de mesure de la qualité vidéo. Les résultats du VQEG sont transmis à l'Union internationale des télécommunications (UIT) et peuvent servir de base à des recommandations internationales sur la mesure de la qualité vidéo.

Groupe d'experts pour les images animées (MPEG, Moving Picture Experts Group): Groupe de travail de l'ISO/CEI chargé d'élaborer des normes pour la représentation codée des séquences audio et vidéo numériques (par exemple MPEG-1, MPEG-2, MPEG-4).

H.261: Désigne la Recommandation UIT-T H.261.

Incertitude (*U, uncertainty*): Evaluation de l'erreur d'alignement temporel (plus ou moins), compte tenu de la valeur la plus probable du retard vidéo dû au circuit fictif de référence. Voir le § 6.4.

Information temporelle absolue (ATI, *absolute temporal information*): Caractéristique déduite de la valeur absolue des images d'information temporelle qui sont calculées comme étant la différence entre deux images successives d'un clip vidéo. La caractéristique ATI quantifie la quantité de mouvement présente dans une scène vidéo. Le § 7.5 contient la définition mathématique précise.

Image: Une image de télévision complète.

Images par seconde (FPS, *frames per second*): Nombre d'images d'origine par seconde transmises par le système vidéo testé. Par exemple, un système vidéo NTSC transmet environ 30 fps.

Information spatiale (SI, *spatial information*): Caractéristique fondée sur des statistiques qui sont extraites des gradients spatiaux (c'est-à-dire des contours) d'une image ou d'une scène vidéo. La Recommandation UIT-T P.910 – Méthodes subjectives d'évaluation de la qualité vidéographique pour les applications multimédias contient une définition de SI fondée sur des statistiques extraites d'images auxquelles on a appliqué des filtres de Sobel 3×3 [Jain, 1989] tandis que le § 7.2 de la présente Annexe contient une définition de SI fondée sur des statistiques extraites d'images auxquelles on a appliqué des filtres de souligné des contours de taille beaucoup plus grande (13×13) (voir la Fig. 29).

Information temporelle (TI, *temporal information*): Caractéristique fondée sur des statistiques qui sont extraites des gradients temporels (c'est-à-dire du mouvement) d'une scène vidéo. La Recommandation UIT-T P.910 et le § 7.5 de la présente Annexe contiennent des définitions de l'information temporelle fondée sur des statistiques extraites de simples différences entre images.

Luminance (*Y*): Partie du signal vidéo qui achemine avant tout l'information de luminance (c'est-à-dire la partie en noir et blanc de l'image).

Mesure de la qualité vidéo, modèle de mesure de la qualité vidéo, qualité VQM (VQM, *video quality metric, model, or measurement*): Mesure globale de la dégradation de la qualité vidéo (voir qualité VQM d'un clip, modèle général). La qualité VQM est un nombre unique dont la plage nominale est comprise entre zéro et un, zéro correspondant à aucune dégradation perçue et un à la dégradation maximale perçue.

Modèle général: Modèle de mesure de la qualité vidéo, ou modèle VQM, qui fait l'objet de la présente Annexe 5. Ce modèle a été soumis aux essais de Phase II réalisés par le Groupe d'experts en qualité vidéo (VQEG). Le rapport final du VQEG sur la Phase II décrit les performances du modèle général (voir le modèle H^1).

Note moyenne d'opinion (MOS, *mean opinion score*): Appréciation subjective moyenne de la qualité d'un clip vidéo traité attribuée par un groupe d'observateurs.

Paramètre: Mesure de la distorsion vidéo résultant de la comparaison de deux flux parallèles de caractéristiques, l'un des flux provenant de la séquence vidéo d'origine et l'autre étant le flux correspondant provenant de la séquence vidéo traitée.

Qualité VQM d'un clip: Qualité VQM d'un seul clip vidéo traité.

Quart de format intermédiaire commun (QCIF, *quarter common intermediate format*): Structure d'échantillonnage vidéo utilisée en visioconférence, pour laquelle le canal de luminance est échantillonné à 176 pixels par 144 lignes (Recommandation UIT-T H.261).

Recommandation UIT-R BT.601: Norme (voir le § 2) commune d'échantillonnage vidéo sur 8 bits selon laquelle le canal de luminance (Y) est échantillonné à 13,5 MHz et les canaux de différence de couleur bleue et rouge (C_B et C_R) sont échantillonnés à 6,75 MHz. Pour plus d'informations, on se reportera au § 5.

Référence réduite: Méthode de mesure de la qualité vidéo qui utilise des caractéristiques de faible largeur de bande extraites des flux vidéo d'origine et traité, par opposition à une méthode fondée sur l'image de référence complète pour laquelle il faut connaître entièrement les flux vidéo d'origine et traité (Recommandation UIT-T J.143 – Prescriptions de l'utilisateur relatives aux mesures objectives de la qualité vidéo perçue en télévision numérique par câble). Les méthodes fondées sur une référence réduite présentent des avantages quant à la surveillance de qualité de bout en bout en service étant donné que les informations de référence réduite sont transmises facilement sur les réseaux de télécommunications du monde entier.

Région d'intérêt (ROI, *region of interest*): Grille d'image (spécifiée en coordonnées de rectangle) utilisée pour désigner une sous-région particulière d'une trame ou d'une image vidéo. Voir aussi SROI.

Région d'intérêt d'origine (OROI, *original region of interest*): Région d'intérêt (ROI) extraite de la séquence vidéo d'origine, spécifiée en coordonnées de rectangle.

Région d'intérêt spatiale (SROI, *spatial region of interest*): Grille d'image particulière (spécifiée en coordonnées de rectangle) utilisée pour calculer la qualité VQM d'un clip vidéo. La région SROI est un sous-ensemble rectangulaire entièrement compris dans la région valable traitée. Pour les séquences vidéo échantillonnées selon la Recommandation UIT-R BT.601, la région SROI recommandée est de 672 pixels × 448 lignes pour les systèmes à 525 lignes et de 672 pixels × 544 lignes pour les systèmes à 625 lignes, centrée à l'intérieur du format de production. Cette région SROI recommandée correspond approximativement à la partie de l'image vidéo que l'on peut voir sur un écran, à l'exclusion de la zone de surbalayage. Voir aussi ROI.

Région d'intérêt temporelle (TROI, *temporal region of interest*): Segment temporel, séquence ou sous-ensemble particulier d'images qui est utilisé pour calculer la qualité VQM d'un clip. La région TROI est un segment contigu d'images qui est entièrement contenu dans la région valable temporelle. La région TROI maximale correspond au segment temporel entièrement aligné et contient toutes les images alignées temporellement de la région TVR. Si une resynchronisation de trame est requise, elle s'applique toujours au clip traité, mais pas au clip d'origine.

Région d'intérêt traitée (PROI, *processed region of interest*): Région d'intérêt (ROI) extraite de la séquence vidéo traitée et dont les décalages spatiaux dus au circuit fictif de référence ont été corrigés, spécifiée en coordonnées de rectangle.

Région valable (VR, *valid region*): Partie rectangulaire d'une grille d'image (spécifiée en coordonnées de rectangle) qui n'est ni supprimée ni altérée par le traitement. La région valable est un sous-ensemble du format de production du système vidéo standard considéré et n'inclut que les pixels d'image qui contiennent une information d'image qui n'a été ni supprimée ni altérée. Voir région valable d'origine et région valable traitée.

Région valable d'origine (OVR, *original valid region*): Région valable d'un clip vidéo d'origine, spécifiée en coordonnées de rectangle.

Région valable temporelle (TVR, *temporal valid region*): Segment temporel, séquence ou sous-ensemble maximal d'images vidéo pouvant être utilisé pour l'étalonnage et le calcul de la qualité VQM. Les images situées en dehors de ce segment temporel seront toujours considérées comme non valables.

Région valable traitée (PVR, *processed valid region*): Région valable d'un clip vidéo traité provenant d'un HRC, spécifiée en coordonnées de rectangle. La région PVR est toujours spécifiée par rapport à la séquence vidéo d'origine, il faut donc corriger les décalages spatiaux de la séquence vidéo dus au HRC avant de calculer la région PVR. Ainsi, la région PVR est toujours contenue dans l'OVR. La région comprise entre la région PVR et la région OVR est la partie de la séquence vidéo qui a été supprimée ou altérée par le HRC.

Resynchronisation de trame: Processus consistant à réordonner, dans une image vidéo, deux trames entrelacées échantillonnées consécutivement d'une séquence vidéo traitée. La resynchronisation de trame est nécessaire lorsque des HRC ne conservent pas l'ordre standard des trames entrelacées (par exemple une trame NTSC une sort sous forme de trame NTSC deux et inversement). Voir le § 6.1.2.

Scène: Séquence d'images vidéo.

Séquence vidéo d'entrée: Séquence vidéo avant traitement ou distorsion par un HRC (voir la Fig. 19). On parle aussi de séquence vidéo d'origine.

Séquence vidéo d'origine: Séquence vidéo avant traitement ou distorsion par un HRC (voir la Fig. 19). On parle aussi de séquence vidéo d'entrée puisque c'est la séquence vidéo qui entre dans le système de transmission vidéo numérique.

Séquence vidéo de sortie: Séquence vidéo qui a été traitée ou distordue par un HRC (voir la Fig. 19). On parle aussi de séquence vidéo traitée.

Séquence vidéo traitée: Séquence vidéo qui a été traitée ou distordue par un HRC (voir la Fig. 19). On parle aussi de séquence vidéo de sortie puisque c'est la séquence de sortie du système de transmission vidéo numérique.

Société des ingénieurs en images animées et télévision (SMPTE, Society of Motion Picture and Television Engineers): Importante pour les industriels travaillant dans le domaine des images animées et de la télévision, cette société se charge de développer la théorie et les applications dans le domaine des images animées, y compris les films, la télévision, la vidéo, l'imagerie sur ordinateur et les télécommunications. Les industriels attendent de la SMPTE qu'elle élabore des normes, des lignes directrices en matière d'ingénierie et des pratiques recommandées qui doivent ensuite être suivies par les professionnels respectifs sur le terrain.

Sous-région spatio-temporelle (S-T): Bloc de pixels d'image d'un flux vidéo d'origine ou traité qui inclut une dimension verticale (nombre de lignes), une dimension horizontale (nombre de colonnes) et une dimension temporelle (nombre d'images). Voir la Fig. 27.

Surbalayage: Partie du flux vidéo qu'on ne peut généralement pas voir sur un écran de télévision standard.

Système NTSC (National Television Systems Committee): Système couleur de vidéo composite analogique à 525 lignes [SMPTE, 1999].

Système PAL (*phase-alternate line*): Système couleur de vidéo composite analogique à 625 lignes.

Trame: La moitié d'une image, contenant toutes les lignes impaires ou toutes les lignes paires.

Unité IRE (Institute for Radio Engineers): Unité de tension couramment utilisée pour mesurer les signaux vidéo. Une IRE vaut 1/140 de volt.

Union internationale des télécommunications (UIT): Organisation internationale du système des Nations Unies où le secteur public et le secteur privé coordonnent les réseaux et services mondiaux de télécommunications. L'UIT inclut le Secteur des radiocommunications (UIT-R) et le Secteur de la normalisation des télécommunications (UIT-T).

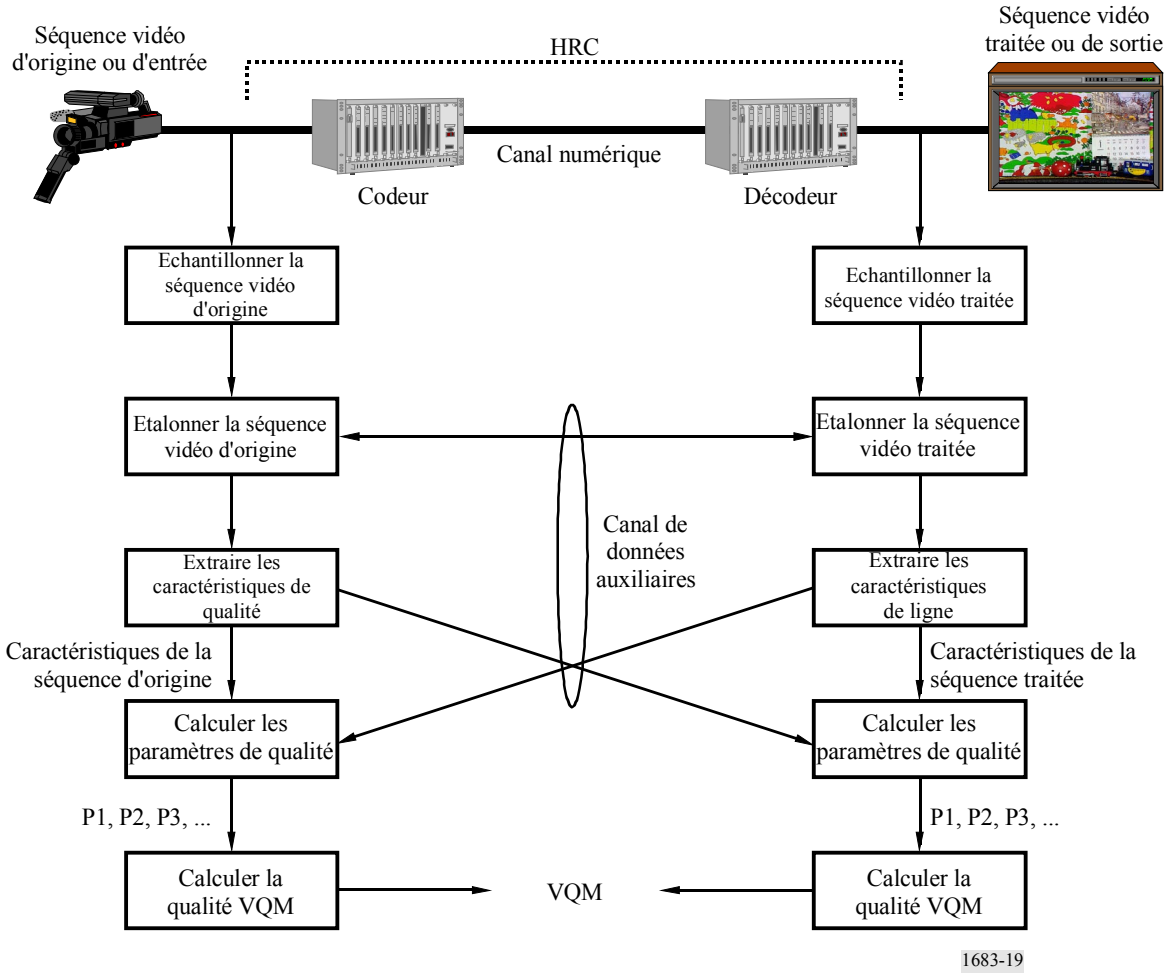
4 Aperçu du calcul de la qualité VQM

La présente Annexe contient une description complète du modèle général et des algorithmes d'étalonnage qui lui sont associés. La méthode de mesure objective automatisée considérée ici donne des résultats proches des appréciations globales (notes moyennes d'opinion) de la qualité vidéo numérique attribuées par des groupes d'observateurs (voir la Recommandation UIT-R BT.500). La Fig. 19 donne un diagramme d'ensemble des processus requis pour calculer la qualité VQM selon le modèle général. Ces processus comprennent l'échantillonnage des flux vidéo d'origine et traité (voir le § 5), l'étalonnage de ces flux (voir le § 6), l'extraction de caractéristiques fondées sur la perception (voir le § 7), le calcul de paramètres de qualité vidéo (voir le § 8) et le calcul du modèle général (voir le § 9). Le modèle général mesure les modifications perçues de la qualité résultant de distorsions dues à n'importe quel composant du système de transmission vidéo numérique (par exemple codeur, canal numérique, décodeur).

La méthode de mesure décrite ici utilise des paramètres de référence réduite de largeur de bande élevée (Recommandation UIT-T J.143). Ces paramètres sont fondés sur des caractéristiques extraites de régions spatio-temporelles (S-T) de la séquence vidéo (voir le § 7.1.1). La méthode de mesure présentée ici peut donc aussi être utilisée pour surveiller la qualité vidéo en service lorsqu'un canal de données auxiliaires est disponible pour transmettre les caractéristiques extraites entre la source et la destination d'un HRC (voir la Fig. 19).

FIGURE 19

Etapes nécessaires pour calculer la qualité VQM



5 Echantillonnage

Pour les algorithmes informatiques exposés dans la présente Annexe, on suppose que les flux vidéo d'origine et traité sont disponibles sous forme de représentations numériques stockées sur support informatique (on parle de clip dans la présente Annexe). Si le flux vidéo est en format analogique, l'une des normes d'échantillonnage numérique les plus largement utilisées est la Recommandation UIT-R BT.601 (§ 2). Un flux vidéo composite (par exemple NTSC ou PAL) doit d'abord être converti en flux vidéo en composantes contenant les trois signaux suivants: luminance (Y), différence de couleur bleue, C_B, et différence de couleur rouge, C_R,. L'échantillonnage selon la Recommandation UIT-R BT.601 est souvent appelé échantillonnage 4:2:2 car la fréquence d'échantillonnage du canal Y est le double de la fréquence d'échantillonnage des canaux C_B et C_R. La Recommandation UIT-R BT.601 spécifie une fréquence d'échantillonnage de 13,5 MHz pour le canal Y, qui produit 720 échantillons Y par ligne vidéo. Etant donné que dans le système NTSC à

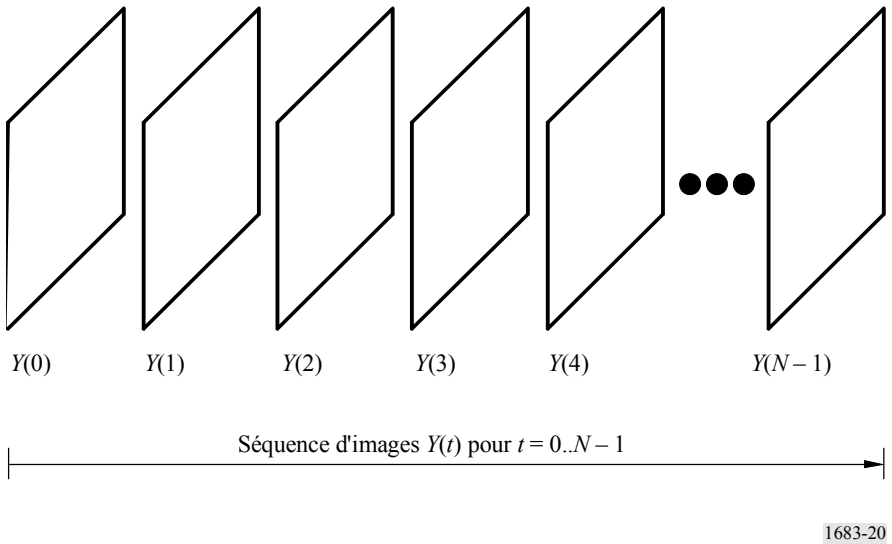
525 lignes, les informations d'image sont contenues dans 486 lignes, l'image vidéo *Y* complète échantillonnée selon la Recommandation UIT-R BT.601 sera de 720 pixels par 486 lignes. De même, lorsqu'un flux vidéo PAL à 625 lignes est échantillonné selon la Recommandation UIT-R BT.601, l'image vidéo *Y* sera de 720 pixels par 576 lignes. Si on utilise 8 bits pour échantillonner de manière uniforme le signal *Y*, la Recommandation UIT-R BT.601 spécifie que la valeur d'échantillonnage du noir de référence (c'est-à-dire 7,5 unités IRE) est «16» et que celle du blanc de référence (c'est-à-dire 100 unités IRE) est «235». Ainsi, une marge de travail est prévue pour les signaux vidéo qui dépassent les niveaux du noir et du blanc de référence avant écrêtage par le convertisseur analogique-numérique. Chacun des canaux de chrominance (C_B et C_R) est échantillonné à 6,75 MHz et le premier couple d'échantillons de chrominance (C_B, C_R) est associé au premier échantillon de luminance *Y*, le deuxième couple d'échantillons de chrominance est associé au troisième échantillon de luminance, etc. Comme les canaux de chrominance sont bipolaires, la valeur d'échantillonnage du signal nul est «128».

5.1 Indexation temporelle des images figurant dans les fichiers vidéo d'origine et traité

Une image de luminance de flux vidéo échantillonnée selon la Recommandation UIT-R BT.601 sera désignée par $Y(t)$. La variable t est utilisée ici comme indice pour les images échantillonnées figurant dans les fichiers Big *YUV* d'origine et traité; elle ne désigne pas le temps véritable. Si le fichier Big *YUV* contient N images, comme indiqué sur la Fig. 20, $t = 0$ désigne la première image qui a été échantillonnée et $t = (N - 1)$ désigne la dernière image qui a été échantillonnée.

FIGURE 20

Indexation temporelle des images figurant dans les fichiers Big *YUV*



Tous les algorithmes décrits ici fonctionnent sur la base de couples de fichiers échantillonnés, chaque couple comprenant un fichier pour la séquence vidéo d'origine et un fichier pour la séquence vidéo traitée associée. Pour éviter toute confusion, on suppose que les deux fichiers d'un couple ont la même longueur. Par ailleurs, on suppose au départ que la première image du fichier d'origine est alignée temporellement avec la première image du fichier traité, avec plus ou moins une certaine incertitude temporelle.

Pour les implémentations en service et en temps réel, cette hypothèse d'incertitude bilatérale peut être remplacée par une hypothèse d'incertitude unilatérale, découlant de la causalité. Par exemple, une image traitée apparaissant à l'instant $t = n$ doit provenir d'images d'origine apparues à l'instant $t = n$ ou antérieurement.

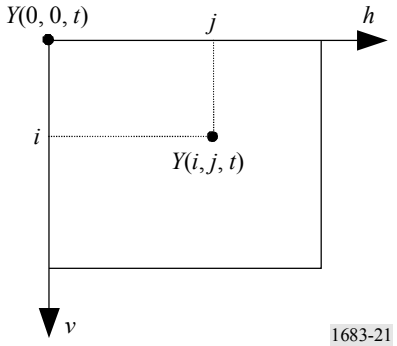
L'hypothèse susmentionnée concernant les fichiers vidéo d'origine et traité (à savoir que les premières images sont alignées) équivaut à choisir la valeur la plus probable du retard dû au HRC présenté sur la Fig. 19. Par conséquent, l'incertitude restante quant à l'évaluation du retard vidéo sera de plus ou moins U.

5.2 Indexation spatiale des images des flux vidéo d'origine et traité

Le système de coordonnées utilisé pour les images de luminance échantillonnées est présenté sur la Fig. 21. Les coordonnées horizontale et verticale du coin en haut à gauche des images de luminance sont définies comme valant ($v = 0$, $h = 0$), où la valeur de la coordonnée sur l'axe horizontal, h, croît vers la droite et la valeur de la coordonnée sur l'axe vertical v croît vers le bas. La coordonnée sur l'axe horizontal est comprise entre 0 et le nombre de pixels d'une ligne moins un. La coordonnée sur l'axe vertical est comprise entre 0 et le nombre de lignes moins un, le nombre de lignes étant le nombre de lignes d'une image pour les systèmes à balayage progressif et soit le nombre de lignes d'une trame soit le nombre de lignes d'une image pour les systèmes à balayage avec entrelacement. L'amplitude du pixel de $Y(t)$ échantillonné correspondant à la ligne i ($v = i$) et à la colonne j ($h = j$) et à l'instant t est désignée par $Y(i, j, t)$.

FIGURE 21

Système de coordonnées utilisé pour les images *Y* de luminance échantillonnées



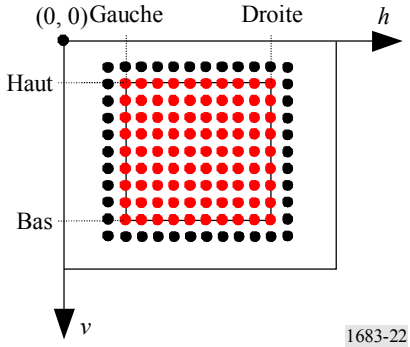
Un clip vidéo échantillonné selon la Recommandation UIT-R BT.601 est stocké dans un fichier de format «Big *YUV*», *Y* désignant l'information de luminance selon la Recommandation UIT-R BT.601, *U* l'information de différence de couleur bleue (c'est-à-dire C_B dans la Recommandation UIT-R BT.601) et *V* l'information de différence de couleur rouge (c'est-à-dire C_R dans la Recommandation UIT-R BT.601). Avec le format de fichier Big *YUV*, toutes les images sont stockées séquentiellement dans un seul grand fichier binaire continu. Les pixels d'image sont stockés séquentiellement par ligne de balayage vidéo sous forme d'octets dans l'ordre suivant: C_{B0}, Y_0, C_{R0}, Y_1, C_{B2}, Y_2, C_{R2}, Y_3, etc., l'indice numérique désignant le numéro du pixel (on doit procéder à une duplication de pixel ou à une interpolation entre pixels pour déterminer les échantillons de chrominance C_B et C_R associés à Y_1, Y_3, …). Cet ordre des octets est équivalent à celui qui est spécifié dans la norme SMPTE 125M [SMPTE, 1995a].

5.3 Spécification de sous-régions rectangulaires

On utilise des sous-régions rectangulaires d'une image échantillonnée pour contrôler le calcul de la qualité VQM. Par exemple, on peut calculer la qualité VQM sur la région valable de l'image échantillonnée ou sur une région d'intérêt spatiale spécifiée par l'utilisateur qui est plus petite que la région valable. Pour spécifier des sous-régions rectangulaires, on utilise les coordonnées de rectangle définies par les quatre grandeurs suivantes: haut, gauche, bas et droite. La Fig. 22 illustre la spécification d'une sous-région rectangulaire d'une image vidéo échantillonnée. Les pixels rouges de l'image sont inclus dans la sous-région mais les pixels noirs de l'image en sont exclus. Pour le calcul de la qualité VQM, une image est souvent subdivisée en un grand nombre de sous-régions plus petites contiguës. La définition d'une sous-région rectangulaire présentée sur la Fig. 22 permet de définir la grille utilisée pour afficher ces sous-régions contiguës et les fonctions mathématiques utilisées pour extraire les caractéristiques de chacune de ces sous-régions.

FIGURE 22

Coordonnées de rectangle pour la spécification de sous-régions d'une image



5.4 Considérations relatives aux séquences vidéo de plus de 10 s

Pour les mesures de la qualité vidéo dont il est question dans la présente Annexe, on s'est fondé sur les résultats d'essais subjectifs relatifs à des clips vidéo de 8 à 10 s. Lorsque la séquence est plus longue, il convient de la subdiviser en segments vidéo plus courts, chaque segment étant supposé avoir ses propres attributs d'étalonnage et de qualité. La méthode consistant à subdiviser le flux vidéo en segments se chevauchant et à traiter chaque segment indépendamment des autres permet d'émuler des évaluations continues de la qualité pour les longues séquences vidéo au moyen des techniques de mesure VQM présentées ici.

6 Etalonnage

Quatre étapes sont nécessaires pour étalonner correctement les séquences vidéo échantillonnées en vue de l'extraction des caractéristiques. Ces étapes sont les suivantes:

Etape 1: évaluation de l'alignement spatial et correction;

Etape 2: évaluation de la région valable afin de limiter l'extraction des caractéristiques aux pixels qui contiennent l'information d'image;

Etape 3: évaluation du gain et du décalage de niveau (généralement appelés contraste et brillance) et correction et

Etape 4: évaluation de l'alignement temporel et correction.

L'Etape 2 doit être appliquée aux flux vidéo d'origine et traité. Les Etapes 1, 3 et 4 doivent être appliquées au flux vidéo traité. Généralement, l'alignement spatial, le gain et le décalage de niveau sont constants pour un système vidéo donné et ces grandeurs n'ont donc à être calculées qu'une seule fois. Toutefois, il est courant que la région valable et l'alignement temporel changent en fonction du contenu de la scène. Par exemple, une scène au format plein écran et une scène au format boîte aux lettres auront des régions valables différentes; les systèmes de visioconférence présentent souvent des retards vidéo variables qui dépendent du contenu de la scène (par exemple une scène dans laquelle on voit la tête d'une personne qui parle et une scène d'une épreuve sportive). En plus des techniques d'étalonnage présentées ici, le lecteur souhaitera peut-être aussi examiner d'autres méthodes d'alignement spatial et temporel (voir la Recommandation UIT-T P.931 – Mesure du temps de transmission, de la synchronisation et du débit de trames dans les communications multimédias).

Le fait de procéder à un étalonnage avant l'extraction des caractéristiques implique que les décalages horizontal et vertical de l'image, les décalages temporels du flux vidéo résultant de retards vidéo non nuls et les modifications du contraste et de la brillance d'image comprises dans la plage dynamique de l'unité d'échantillonnage vidéo n'auront pas d'incidence sur la qualité VQM. Ces grandeurs liées à l'étalonnage peuvent avoir une grande incidence sur la qualité globale perçue (par exemple des images à faible contraste issues d'un système vidéo avec un gain de 0,3), mais la philosophie adoptée ici consiste à séparer les informations liées à l'étalonnage de la qualité VQM. De bonnes pratiques techniques permettent généralement d'ajuster les décalages spatiaux, les régions valables, les gains et les décalages de niveau; les décalages temporels fournissent des informations importantes sur la qualité lors de l'évaluation de systèmes vidéo bidirectionnels ou interactifs.

Pour toutes les caractéristiques et tous les paramètres de qualité vidéo (voir les § 7 et 8), on suppose qu'un seul retard vidéo est supprimé pour l'alignement temporel de la séquence vidéo traitée (retard vidéo constant). Certains systèmes vidéo ou HRC appliquent un retard différent à chaque image traitée (retard vidéo variable). Dans la présente Annexe, on considère que tous les systèmes vidéo ont un retard vidéo constant. Les variations par rapport à ce retard sont considérées comme des dégradations qui sont mesurées par les caractéristiques et les paramètres. Cette approche semble conduire à de meilleures corrélations avec la note subjective que les mesures de la qualité vidéo fondées sur des séquences vidéo traitées dont le retard vidéo variable a été supprimé. Lorsqu'une séquence vidéo est longue (voir le § 5.4), il convient de la subdiviser en segments vidéo plus courts, chaque segment ayant son propre retard vidéo constant, ce qui autorise une certaine variation du retard en fonction du temps. Il est possible d'obtenir une évaluation plus continue des variations du retard en subdivisant la séquence en segments temporels se chevauchant.

Si le HRC testé réduit ou agrandit la taille de l'image (par exemple zoom), il faudrait inclure, dans le processus d'étalonnage, une étape additionnelle visant à évaluer et à supprimer cette réduction ou cet agrandissement spatial. Cette étape n'entre pas dans le cadre de la présente Annexe.

6.1 Alignement spatial

6.1.1 Aperçu

Le processus d'alignement spatial détermine les décalages spatiaux horizontal et vertical d'une image vidéo traitée par rapport à l'image vidéo d'origine. Un décalage horizontal positif correspond à une image traitée qui a été déplacée vers la droite par un certain nombre de pixels. Un décalage vertical positif correspond à une image traitée qui a été déplacée vers le bas par un certain nombre de lignes. Ainsi, pour l'alignement spatial d'une image vidéo avec balayage à entrelacement, il faut

tenir compte de trois grandeurs: le décalage horizontal en nombre de pixels, le décalage vertical de la trame une en nombre de lignes de trame et le décalage vertical de la trame deux en nombre de lignes de trame. Pour l'alignement spatial d'une image vidéo avec balayage progressif, il faut tenir compte de deux grandeurs: le décalage horizontal et le décalage vertical en nombre de lignes d'image. L'algorithme d'alignement spatial est précis au pixel près pour les décalages horizontaux et à la ligne près pour les décalages verticaux. Une fois que l'alignement spatial a été calculé, le décalage spatial est supprimé du flux vidéo traité (par exemple une image traitée qui a été décalée vers le bas est redécalée vers le haut). En cas de balayage à entrelacement, le processus peut inclure une resynchronisation de trame du flux vidéo traité découlant de la comparaison des décalages verticaux des trames une et deux.

Dans le cas du balayage à entrelacement, toutes les opérations s'appliquent à chaque trame séparément; dans le cas du balayage progressif, toutes les opérations s'appliquent à l'image entière. Dans un souci de simplicité, l'algorithme d'alignement spatial sera d'abord entièrement décrit dans le cas du balayage à entrelacement, car c'est le cas le plus compliqué. Les modifications à apporter dans le cas du balayage progressif sont présentées au § 6.1.6.

L'alignement spatial doit être déterminé avant la région PVR, le gain et le décalage de niveau ainsi que l'alignement temporel. Plus précisément, pour calculer chacune de ces grandeurs, il faut comparer le contenu vidéo d'origine et le contenu vidéo traité qui a été aligné spatialement. Si le flux vidéo traité a été décalé spatialement par rapport au flux vidéo d'origine et que ce décalage spatial n'a pas été corrigé, les évaluations seraient mauvaises car elles seraient fondées sur des contenus vidéo non analogues. Malheureusement, on ne peut pas déterminer correctement l'alignement spatial si on ne connaît pas la PVR, le gain et le décalage de niveau ainsi que l'alignement temporel. L'interdépendance de ces grandeurs cause un problème de mesure du type "de la poule et de l'œuf". Pour pouvoir calculer l'alignement spatial d'une trame traitée, il faut connaître la PVR, le gain et le décalage de niveau ainsi que la trame d'origine lui correspondant le mieux. Toutefois, il est impossible de déterminer ces grandeurs si le décalage spatial n'est pas connu. Une recherche entièrement exhaustive couvrant toutes les variables nécessiterait un nombre considérable de calculs en cas de grosses incertitudes concernant les grandeurs ci-dessus.

La solution présentée ici consiste à procéder à une recherche itérative afin de trouver la trame d'origine correspondant le mieux à chaque trame traitée. Cette recherche inclut une mise à jour itérative des évaluations de PVR, de gain et de décalage de niveau ainsi que d'alignement temporel. Toutefois, pour certaines trames traitées, l'algorithme d'alignement spatial peut échouer. Généralement, lorsque l'alignement spatial n'est pas évalué correctement pour une trame traitée, l'ambiguïté est due aux caractéristiques de la scène. Considérons, par exemple, une scène avec balayage à entrelacement créée numériquement contenant un panoramique vers la gauche. Comme le panoramique a été généré par ordinateur, cette scène pourrait comporter un panoramique horizontal d'exactement un pixel à chaque trame. Du point de vue de l'algorithme de recherche de l'alignement spatial, il serait impossible de faire une différence entre l'alignement spatial correct calculé par rapport à la trame d'origine correspondante, et un décalage horizontal de deux pixels calculé par rapport à la trame qui précède de deux trames la trame d'origine correspondante. Considérons un autre exemple dans lequel une image est entièrement constituée de lignes verticales noires et blanches numériquement parfaites. Comme l'image ne contient pas de ligne horizontale, le décalage vertical est complètement ambigu. Comme le motif de lignes verticales se répète, le décalage horizontal est ambigu, deux décalages horizontaux ou davantage étant tout aussi probables.

Par conséquent, il convient d'appliquer l'algorithme de recherche itérative à une séquence de trames traitées. Les évaluations individuelles de décalage spatial de plusieurs trames traitées peuvent alors servir à produire une évaluation plus robuste. Les évaluations de décalage spatial de plusieurs

séquences ou scènes peuvent ensuite être combinées afin de produire une évaluation encore plus robuste pour le HRC testé, dans l'hypothèse où le décalage spatial est constant pour toutes les scènes qui passent par ce circuit.

6.1.2 Questions relatives à l'entrelacement

L'alignement spatial vertical est plus complexe pour un flux vidéo avec balayage à entrelacement que pour un flux vidéo avec balayage progressif, car le processus d'alignement spatial doit faire la différence entre la trame une et la trame deux. Trois conditions de décalage vertical doivent être différenciées afin d'obtenir l'alignement vertical correct pour les systèmes avec balayage à entrelacement: le décalage vertical de la trame une est égal au décalage vertical de la trame deux, le décalage vertical de la trame une est inférieur de un au décalage vertical de la trame deux, le décalage vertical est tout autre.

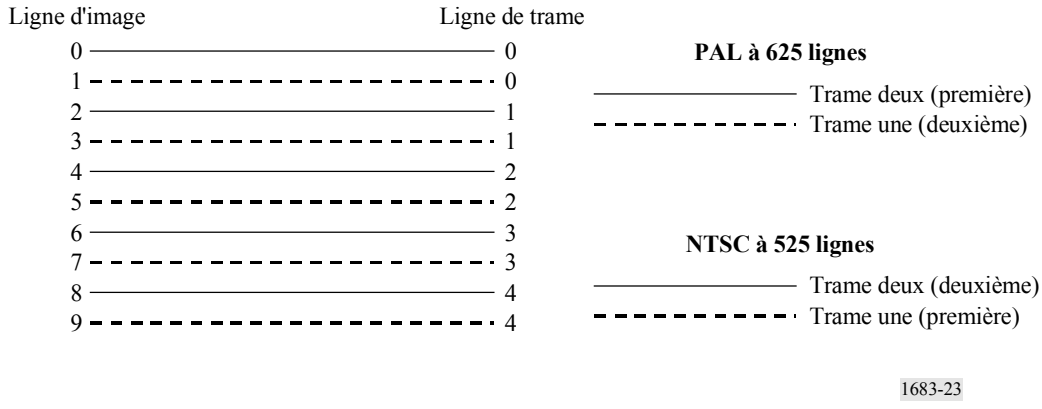
Certains HRC décalent de manière identique la trame une et la trame deux; dans ce cas, le décalage vertical de la trame une est égal au décalage vertical de la trame deux. Pour les HRC qui ne répètent pas les trames ou les images (c'est-à-dire les HRC qui émettent au plein débit d'images du système vidéo), cette condition signifie que ce qui était une trame une dans le flux vidéo d'origine est également une trame une dans le flux vidéo traité et ce qui était une trame deux dans le flux d'origine est également une trame deux dans le flux traité.

D'autres HRC procèdent à une resynchronisation de trame du flux vidéo, décalant l'image échantillonnée par un nombre impair de lignes d'image. La trame une de la séquence d'origine devient la trame deux de la séquence traitée et la trame deux de la séquence d'origine devient la trame une de l'image suivante. Visuellement, le flux vidéo affiché semble correct car l'être humain ne peut pas percevoir un décalage d'image du flux vidéo correspondant à une ligne.

Comme indiqué sur la Fig. 23, la trame une commence à la ligne d'image une et contient toutes les lignes d'image impaires. La trame deux commence à la ligne d'image zéro (ligne d'image la plus en haut) et contient toutes les lignes d'image paires. Pour les systèmes NTSC, la trame une est la première dans le temps et la trame deux la deuxième. Pour les systèmes PAL, la trame deux est la première dans le temps et la trame une la deuxième.

FIGURE 23

Diagramme illustrant le numérotage des trames entrelacées et des lignes d'image/de trame



Une resynchronisation de trame a lieu lorsque la première trame devient la deuxième et la deuxième devient la première de l'image suivante (retard d'une trame) ou lorsque la deuxième trame devient la première et la première de l'image suivante devient la deuxième de l'image en cours (avance d'une trame). Par exemple, lorsque la trame d'origine NTSC deux devient la trame une de l'image NTSC

suivante, la ligne du haut de la trame qui était la ligne d'image 0 de la trame d'origine deux devient la ligne d'image 1 de la trame traitée une. Selon le numérotage des lignes de trame, la ligne du haut reste la ligne de trame 0; ainsi, la trame traitée une présente un décalage vertical nul (car les décalages verticaux sont mesurés pour chaque trame au moyen des lignes de trame). Lorsque la trame NTSC d'origine une devient la trame deux de la même image, la ligne du haut de la trame qui était la ligne d'image 1 de la trame d'origine une devient la ligne d'image 2 de la trame traitée deux. Selon le numérotage des lignes de trame, la ligne du haut qui était la ligne de trame 0 devient la ligne de trame 1; ainsi, la trame traitée deux présente un décalage vertical d'une ligne de trame. La règle générale applicable à la fois au système NTSC et au système PAL est la suivante: lorsque le décalage vertical de la trame deux (en nombre de lignes de trame) est supérieur de un au décalage vertical de la trame une (en nombre de lignes de trame), une resynchronisation de trame a eu lieu.

Si le décalage vertical de la trame deux est différent de celui de la trame une et ne vaut pas non plus un de plus que celui de la trame une, le HRCa altéré l'échantillonnage spatial des deux trames entrelacées et la scène vidéo résultante apparaîtra comme montant et descendant brusquement. Une telle dégradation est évidente et gênante pour l'observateur et, de fait, se produit rarement dans la pratique car le concepteur de HRC découvre et corrige l'erreur. Par conséquent, l'alignement spatial repose la plupart du temps sur deux schémas courants. Dans les systèmes sans resynchronisation de trame, le décalage vertical de la trame une est égal au décalage vertical de la trame deux; dans les systèmes avec resynchronisation de trame, le décalage vertical de la trame une plus un est égal au décalage vertical de la trame deux.

En outre, il est à noter que l'alignement spatial inclut certaines informations d'alignement temporel, notamment la question de savoir s'il y a eu resynchronisation de trame ou non. Le processus d'alignement temporel peut ne pas être capable de détecter une resynchronisation de trame, mais même s'il le peut, la resynchronisation de trame est inhérente au processus d'alignement spatial. L'alignement spatial doit donc être capable de déterminer si la trame traitée considérée correspond le mieux à une trame d'origine une ou deux. L'alignement spatial pour chaque trame ne peut être calculé correctement que lorsque la trame traitée est comparée avec la trame d'origine dont elle est issue. Mise à part la question de la resynchronisation de trame, l'utilisation de la mauvaise trame d'origine (trame une/trame deux) peut entraîner des imprécisions quant à l'alignement spatial en raison des différences intrinsèques de contenu spatial dans les deux trames entrelacées.

6.1.3 Variables d'entrée requises par l'algorithme d'alignement spatial

Le présent paragraphe contient la liste des variables d'entrée requises par l'algorithme d'alignement spatial. Ce sont notamment la plage des décalages spatiaux et la plage des trames d'origine sur lesquelles la recherche doit porter. Si ces plages sont trop grandes, la vitesse de convergence de l'algorithme de recherche itérative utilisé pour trouver le décalage spatial risque d'être lente et la probabilité pour que l'alignement spatial pour des scènes au contenu répétitif soit erroné sera élevée (par exemple quelqu'un qui fait un signe de la main). Inversement, si ces plages sont trop petites, l'algorithme de recherche se heurtera aux limites des plages de recherche et les repoussera lentement au cours des itérations successives. Cette intelligence de recherche intégrée est utile si l'utilisateur fait une faible erreur d'évaluation des incertitudes de la recherche, mais risque d'augmenter considérablement le temps d'exécution si l'utilisateur fait une forte erreur d'évaluation. Par ailleurs, l'algorithme de recherche risque de ne pas trouver le décalage spatial correct dans ce cas.

6.1.3.1 Plage prévue des décalages spatiaux

La plage prévue des décalages spatiaux pour des flux vidéo à 525 lignes et à 625 lignes échantillonnés conformément à la Recommandation UIT-R BT.601 est de ±20 pixels horizontalement et de ±12 lignes de trame verticalement. Elle a été déterminée empiriquement sur la base du traitement de données vidéo issues de centaines de HRC. La plage prévue des décalages

spatiaux pour des flux vidéo échantillonnés conformément à d'autres formats plus petits que ceux de la Recommandation UIT-R BT.601 (par exemple CIF) est supposée être moitié moins grande que la plage observée pour les systèmes à 525 lignes et à 625 lignes. L'algorithme de recherche devrait fonctionner correctement - quoiqu'un peu plus lentement - lorsque la trame traitée présente des décalages spatiaux non compris dans la plage prévue des décalages spatiaux. Cela est dû au fait que l'algorithme de recherche élargira la recherche au-delà de la plage prévue des décalages spatiaux lorsque c'est justifié. Toutefois, le résultat de la détermination de l'alignement spatial correct risque d'être signalé comme étant un échec si les excursions dépassent 50% de la plage prévue.

6.1.3.2 Incertitude temporelle

L'utilisateur doit aussi spécifier l'incertitude quant à l'alignement temporel, c'est-à-dire la plage des trames d'origine à examiner pour chaque trame traitée. Cette incertitude temporelle est exprimée sous la forme d'un certain nombre de trames avant et après l'alignement temporel par défaut. Si les séquences vidéo d'origine et traitée sont stockées sous forme de fichiers, un alignement temporel par défaut raisonnable consiste à supposer que la première trame d'un fichier est alignée avec la première trame de l'autre fichier. L'incertitude temporelle qui est spécifiée devrait être suffisamment grande pour inclure l'alignement temporel réel. Une incertitude de plus ou moins une seconde (30 images dans le cas NTSC à 525 lignes; 25 images dans le cas PAL à 625 lignes) devrait suffire pour la plupart des systèmes vidéo. Une incertitude temporelle plus grande pourra être nécessaire pour les HRC présentant de longs retards vidéo. L'algorithme de recherche pourra envisager des alignements temporels qui sortent de la plage d'incertitude spécifiée lorsque c'est justifié (par exemple lorsque la trame d'origine la plus éloignée est choisie comme correspondant au meilleur alignement temporel).

6.1.3.3 Evaluation de la région PVR

L'évaluation de la région PVR consiste à spécifier la partie de l'image traitée qui n'a été ni supprimée ni altérée par le traitement, en supposant qu'il n'y a pas eu de décalage spatial (car le décalage spatial n'a pas encore été mesuré). L'évaluation de la PVR peut être déterminée empiriquement, mais une évaluation de la PVR qui est spécifiée par l'utilisateur et qui exclut la zone de surbalayage constitue un bon choix. Dans la plupart des cas, cela permet de ne pas utiliser les parties vidéo non valables dans l'algorithme d'alignement spatial. Concernant les flux vidéo NTSC à 525 lignes échantillonnés conformément à la Recommandation UIT-R BT.601, la zone de surbalayage couvre environ 18 lignes d'image en haut et en bas de l'image et 22 pixels à gauche et à droite de l'image. Concernant les flux vidéo PAL à 625 lignes échantillonnés conformément à la Recommandation UIT-R BT.601, la zone de surbalayage couvre environ 14 lignes d'image en haut et en bas de l'image et 22 pixels à gauche et à droite de l'image. Pour les autres formats d'image (par exemple CIF), il convient de choisir une PVR par défaut raisonnable.

6.1.4 Sous-algorithmes utilisés par l'algorithme d'alignement spatial

L'algorithme d'alignement spatial utilise un certain nombre de sous-algorithmes – notamment pour évaluer le gain et le décalage de niveau – et des formules permettant de déterminer la trame d'origine qui correspond le mieux à une trame traitée donnée. Ces sous-algorithmes ont été conçus pour être efficaces sur le plan du calcul, étant donné qu'ils doivent être exécutés de nombreuses fois dans le cadre de l'algorithme de recherche itérative.

6.1.4.1 Région ROI utilisée par tous les calculs

Toutes les comparaisons de trame opérées par l'algorithme se font entre des versions décalées spatialement d'une ROI extraite du flux vidéo traité (afin de compenser les décalages spatiaux introduits par le HRC) et la ROI correspondante extraite du flux vidéo d'origine. Toute ROI extraite du flux vidéo traité et décalée spatialement sera appelée PROI (ROI traitée) et la ROI

correspondante extraite du flux vidéo d'origine sera appelée OROI (ROI d'origine). Les coordonnées de rectangle qui spécifient la OROI sont fixes tout au long de l'algorithme et sont choisies de manière à avoir la plus grande OROI possible qui satisfait aux deux conditions suivantes:

– La OROI doit correspondre à une PROI qui est située dans la région PVR pour tous les décalages spatiaux possibles qui sont examinés.

– La OROI est centrée dans l'image d'origine.

6.1.4.2 Gain et décalage de niveau

L'algorithme qui suit sert à évaluer le gain du flux vidéo traité. On corrige le décalage spatial de la trame traitée examinée en utilisant l'évaluation courante du décalage spatial. Après cette correction, on choisit une PROI qui correspond à la OROI fixe déterminée au § 6.1.4.1. On calcule ensuite l'écart type des valeurs des pixels de luminance (Y) de cette PROI et l'écart type des valeurs des pixels de luminance (Y) de la OROI. On évalue alors le gain comme étant l'écart type associé à la PROI divisé par l'écart type associé à la OROI.

A mesure que l'on se rapproche du décalage spatio-temporel correct au cours des itérations successives de l'algorithme, la fiabilité de cette évaluation du gain est renforcée. On peut utiliser un gain de 1,0 (c'est-à-dire aucune correction du gain) pendant les premiers cycles d'itération. Le calcul de gain décrit ci-dessus est sensible aux dégradations présentes dans le flux vidéo traité (par exemple flou). Toutefois, pour l'alignement spatial, cette évaluation du gain est utile car elle permet au flux vidéo traité de ressembler le plus possible au flux vidéo d'origine. Pour supprimer le gain de la trame traitée, la valeur de chaque pixel de luminance de la trame traitée est divisée par le gain.

Il n'est pas nécessaire de déterminer ou de corriger le décalage de niveau, car les décalages de niveau n'ont pas d'incidence sur les critères de recherche de l'algorithme d'alignement spatial (voir le § 6.1.4.3).

6.1.4.3 Formules utilisées pour comparer la PROI avec la OROI

Après avoir corrigé le gain[3] dans la PROI (§ 6.1.4.2), on utilise l'écart type de l'image de différence (OROI-PROI) pour choisir un décalage spatial et un décalage temporel parmi différentes valeurs. On utilise l'évaluation de gain associée à la meilleure correspondance précédente pour corriger le gain de la PROI. Pour déterminer un décalage spatial parmi plusieurs valeurs (le décalage temporel étant maintenu constant), on calcule l'écart type de l'image de différence (OROI-PROI) pour plusieurs PROI générées avec différents décalages spatiaux. Pour une trame traitée donnée, on choisit la combinaison de décalages spatial et temporel qui produit l'écart type le plus petit (c'est-à-dire la plus grande annulation par rapport à la trame d'origine) comme correspondant à la meilleure correspondance.

6.1.5 Alignement spatial utilisant des scènes arbitraires

Pour l'alignement spatial d'une trame traitée extraite d'une scène, il faut examiner plusieurs trames d'origine et décalages spatiaux car le décalage temporel (c'est-à-dire le retard vidéo) et le décalage spatial sont tous deux inconnus. Il s'ensuit que l'algorithme de recherche est complexe et nécessite beaucoup de calculs. Par ailleurs, comme le contenu de la scène est arbitraire, il est possible que

---

[3] Afin de réduire la complexité des calculs, la compensation du gain peut parfois être omise. Toutefois, l'omission de la correction du gain n'est recommandée qu'au cours des premières étapes de l'algorithme de recherche itérative, dont l'objectif est de déterminer un alignement spatial approximatif (voir par exemple les § 6.1.5.2 et 6.1.5.3).

l'algorithme détermine un alignement spatial incorrect (voir le § 6.1.1). Il est donc prudent de calculer l'alignement spatial de plusieurs trames traitées extraites de plusieurs scènes différentes qui sont toutes passées par le même HRC et de combiner les résultats afin d'obtenir une évaluation robuste du décalage spatial. Un HRC donné devrait avoir un seul alignement spatial constant. Si ce n'est pas le cas, des décalages spatiaux variables dans le temps seraient perçus comme une dégradation (par exemple le flux vidéo rebondirait de haut en bas et de bas en haut ainsi que d'un côté à l'autre). Le présent paragraphe décrit l'algorithme d'alignement spatial dans le cas haut-bas; pour cela, on décrit d'abord les principaux composants de l'algorithme puis leur application pour des scènes et des HRC.

6.1.5.1 Meilleure correspondance de trame d'origine dans le temps

Pour déterminer l'alignement spatial à partir du contenu d'une scène, l'algorithme doit déterminer la trame d'origine qui correspond le mieux à la trame traitée courante. Malheureusement, il se peut que cette trame d'origine n'existe pas. Par exemple, une trame traitée peut contenir des parties de deux trames d'origine différentes car elle a pu être interpolée à partir d'autres trames traitées. L'évaluation courante de la meilleure correspondance de trame d'origine (c'est-à-dire la trame d'origine qui correspond le mieux à la trame traitée courante) est conservée à toutes les étapes de l'algorithme de recherche.

On suppose au départ que la première trame du fichier Big *YUV* traité est alignée avec la première trame du fichier Big *YUV* d'origine, avec plus ou moins une certaine incertitude temporelle en nombre d'images (appelée *U*). Pour chaque trame traitée qui est examinée par l'algorithme, il faut un tampon de *U* images d'origine avant et après cette trame. L'algorithme commence donc à examiner les trames traitées se trouvant à *U* images après le début du fichier, examine toutes les trames qui suivent correspondant à une certaine fréquence (appelée *F*), et s'arrête *U* images avant la fin du fichier.

Les résultats finals de la recherche pour la trame traitée précédente (gain, décalage vertical, décalage horizontal, décalage temporel) sont utilisés pour initialiser la recherche pour la trame traitée courante. Pour calculer la meilleure correspondance de trame d'origine pour la trame traitée courante, on suppose que le retard vidéo est constant. Par exemple, s'il a été déterminé que la meilleure correspondance pour la trame traitée *N* est la trame d'origine *M* dans les fichiers Big *YUV*, on suppose, au début de la recherche, que la meilleure correspondance pour la trame traitée $N + F$ est la trame d'origine $M + F$.

6.1.5.2 Recherche large du décalage temporel

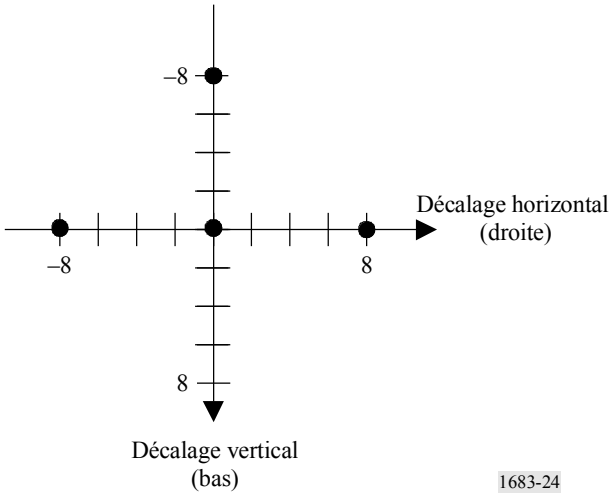
Une recherche complète parmi tous les décalages spatiaux possibles dans toute la plage d'incertitude temporelle pour chaque trame traitée nécessiterait un grand nombre de calculs. A la place, on utilise une recherche en plusieurs étapes, la première étape étant une recherche large du décalage temporel sur un ensemble très limité de décalages spatiaux, dont le but est de se rapprocher de la correspondance correcte de trame d'origine.

Dans le cadre de cette recherche large pour l'image traitée considérée, on examine la trame une de cette image (voir la Fig. 23) et on ne considère que les trames d'origine qui sont des trames unes et qui sont espacées de deux images (c'est-à-dire qui sont espacées de quatre trames) dans toute la plage correspondant à plus ou moins l'incertitude d'alignement temporel. Dans le cadre de cette recherche large, on considère les quatre décalages spatiaux suivants du flux vidéo traité: pas de décalage, huit pixels vers la gauche, huit pixels vers la droite et huit lignes de trame vers le haut (voir la Fig. 24). Sur la Fig. 24, les décalages positifs correspondent aux décalages vers le bas et

vers la droite du flux vidéo traité par rapport au flux vidéo d'origine. Le décalage de «huit lignes de trame vers le bas» n'est pas envisagé car des observations empiriques ont montré que très peu de systèmes vidéo déplacent l'image vers le bas. La meilleure évaluation précédente du décalage spatial (c'est-à-dire associé à une trame traitée précédemment) est également incluse comme cinquième décalage possible lorsqu'elle est disponible. Pour déterminer la trame d'origine correspondant le mieux à la trame traitée considérée, on utilise la technique de comparaison décrite au § 6.1.4.3. Le décalage temporel associé à la meilleure correspondance de trame d'origine devient le point de départ de l'étape suivante de l'algorithme, à savoir une recherche large du décalage spatial (§ 6.1.5.3). Conformément au système de coordonnées de la Fig. 21, un décalage temporel positif signifie que le flux vidéo traité a été décalé dans le sens temporel positif (c'est-à-dire que le flux vidéo traité est retardé par rapport au flux vidéo d'origine). En ce qui concerne les fichiers Big *YUV* d'origine et traité, un décalage temporel positif signifie donc que des trames doivent être éliminées au début du fichier Big *YUV* traité alors qu'un décalage temporel négatif signifie que des trames doivent être éliminées au début du fichier Big *YUV* d'origine.

FIGURE 24

Décalages spatiaux envisagés dans le cadre de la recherche large du décalage temporel

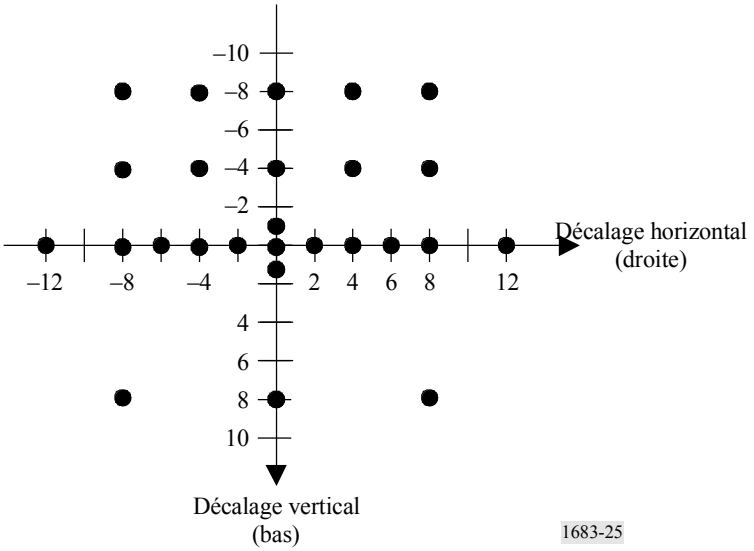


6.1.5.3 Recherche large du décalage spatial

Compte tenu de l'alignement temporel déterminé par la recherche large du décalage temporel (voir le § 6.1.5.2), on procède alors à une recherche large du décalage spatial sur une plage plus limitée de trames d'origine. La plage des trames d'origine qui sont considérées pour cette recherche comprend la trame d'origine de meilleure correspondance qui est une trame une (voir le § 6.1.5.2) et les quatre trames d'origine les plus proches qui sont également des trames unes (trames unes des deux images qui précèdent et des deux images qui suivent la trame d'origine de meilleure correspondance). La recherche large du décalage spatial couvre la plage des décalages spatiaux donnée à la Fig. 25. Il est à noter que l'on envisage un moins grand nombre de décalages vers le bas (comme au § 6.1.5.2), car ceux-ci sont moins fréquents dans la pratique. On applique alors la technique de comparaison décrite au § 6.1.4.3 à l'ensemble de ces décalages spatiaux et de ces trames d'origine. Les meilleurs décalages temporel et spatial résultants servent alors d'évaluations améliorées pour l'étape suivante de l'algorithme décrite au § 6.1.5.4.

FIGURE 25

Décalages spatiaux envisagés dans le cadre de la recherche large du décalage spatial



6.1.5.4 Recherche fine du décalage spatio-temporel

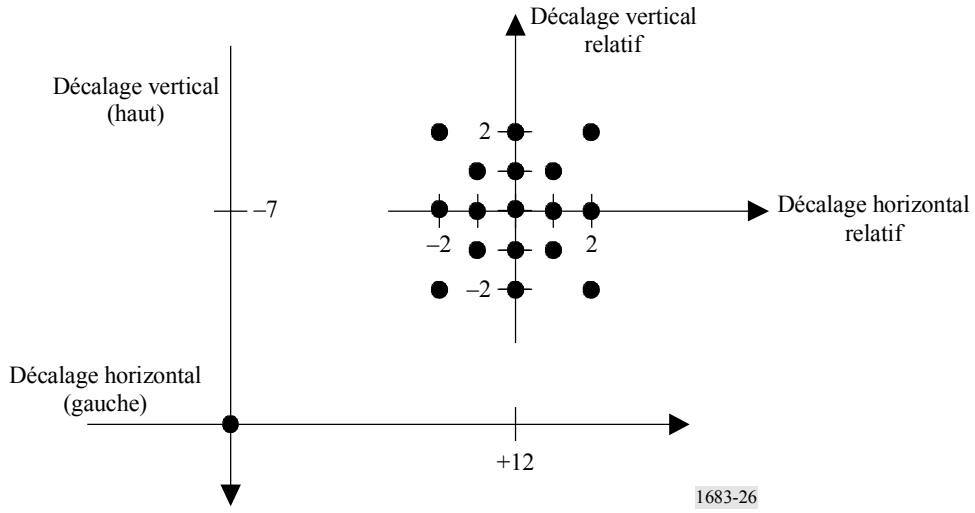
Pour la recherche fine, on utilise un ensemble beaucoup plus petit de décalages spatiaux centrés autour de l'évaluation courante de l'alignement spatial et uniquement cinq trames centrées autour de la trame d'origine de meilleure correspondance. Ainsi, si cette trame est une trame une, on inclut dans la recherche trois trames unes et deux trames deux. Les décalages spatiaux qui sont envisagés comprennent l'évaluation courante du décalage, les huit décalages d'un pixel et/ou d'une ligne par rapport à l'évaluation courante, les huit décalages de deux pixels et/ou de deux lignes par rapport à l'évaluation courante du décalage, et le décalage nul (voir la Fig. 26). Dans l'exemple présenté sur la Fig. 26, l'évaluation courante du décalage spatial pour le flux vidéo traité est un décalage de 7 lignes de trame vers le haut et de 12 pixels vers la droite par rapport au flux vidéo d'origine. L'ensemble des décalages spatiaux présenté sur la Fig. 26 constitue un ensemble local presque complet d'alignements spatiaux proches de l'évaluation courante de l'alignement spatial. Le décalage nul est inclus comme condition de sécurité afin d'empêcher l'algorithme d'errer et de converger vers un minimum local. On applique alors avec soin la technique de comparaison décrite au § 6.1.4.3 à l'ensemble de ces décalages spatiaux et de ces trames d'origine. Les meilleurs décalages temporel et spatial résultants servent alors d'évaluations améliorées pour l'étape suivante de l'algorithme décrite au § 6.1.5.5.

6.1.5.5 Recherches fines répétées

Lorsqu'on procède à une itération de la recherche fine décrite au § 6.1.5.4, l'évaluation courante du décalage spatial se rapproche du décalage spatial réel ou (plus rarement) d'un faux minimum. De même, lorsqu'on procède à une telle itération, l'évaluation courante de la trame d'origine de meilleure correspondance se rapproche de la trame d'origine de meilleure correspondance réelle ou (plus rarement) d'un faux minimum. Ainsi, chaque recherche fine rapproche ces évaluations d'une valeur stable. Comme les recherches fines portent sur une zone très limitée spatialement et temporellement, elles doivent être répétées afin de s'assurer que la convergence a été atteinte. En cas d'utilisation de la compensation de gain, le gain de la trame traitée est réévalué à chaque recherche fine (voir le § 6.1.4.2).

FIGURE 26

Décalages spatiaux envisagés dans le cadre de la recherche fine du décalage spatial



Les recherches fines portant sur la trame traitée (voir le § 6.1.5.4) sont répétées jusqu'à ce que le meilleur décalage spatial et la trame d'origine associée à ce décalage spatial restent inchangés d'une recherche à la suivante. On cesse de répéter les recherches fines si l'algorithme alterne entre deux décalages spatiaux (par exemple un décalage horizontal de 3 puis un décalage horizontal de 4, toutes les autres grandeurs gardant les mêmes valeurs). Cette alternance apparaît lorsque la meilleure évaluation courante du décalage spatial et la trame d'origine associée à ce décalage spatial sont identiques à celles qui ont été déterminées deux itérations avant.

Parfois, les recherches répétées ne parviennent pas à converger. En l'absence de convergence au bout d'un certain nombre maximal d'itérations demandées, l'algorithme est arrêté et une condition «d'échec de la détermination du décalage» est signalée pour cette trame traitée. Ce cas particulier ne pose généralement pas de problème car de multiples trames traitées sont examinées pour chaque scène (voir le § 6.1.5.6) et de multiples scènes sont examinées pour chaque HRC (voir le § 6.1.5.7).

6.1.5.6 Algorithme pour une scène donnée

On commence par calculer une évaluation de base (de départ) du décalage vertical, du décalage horizontal et de l'alignement temporel sans compensation de gain comme suit. On saute les premières images du fichier Big *YUV* traité correspondant à l'incertitude temporelle, *U*. Une recherche large du décalage temporel est appliquée à la trame traitée suivante qui est une trame une (voir le § 6.1.5.2). Il est à noter que cette recherche large porte sur les $U \cdot 2 + 1$ premières images de la séquence vidéo d'origine afin de trouver la trame une de meilleure correspondance. On procède alors à une recherche large du décalage spatial, centrée sur cette trame d'origine de meilleure correspondance (voir le § 6.1.5.3). On procède ensuite à un maximum de cinq recherches fines afin d'affiner les évaluations du décalage spatial et du décalage temporel (voir les § 6.1.5.4 et 6.1.5.5). Si ces recherches fines répétées n'aboutissent pas à un résultat stable, on élimine cette trame traitée de l'ensemble des trames considérées. On répète la procédure ci-dessus pour chaque image correspondant à une certaine fréquence, *F*, jusqu'à ce qu'on trouve une trame d'origine qui soit une trame une et qui produise des résultats stables. L'évaluation de base sera mise à jour régulièrement, comme décrit ci-dessous.

Les évaluations du décalage spatial sont calculées pour les deux trames d'une image du fichier Big *YUV* traité comme suit. En utilisant l'évaluation de base comme point de départ, on applique un maximum de trois recherches fines à la première trame traitée qui est une trame une. Si l'évaluation de base est correcte ou pratiquement correcte, les recherches fines répétées conduiront à un résultat stable. Si c'est le cas, le décalage spatial et le décalage temporel pour cette trame traitée sont stockés dans une matrice réservée au stockage des résultats relatifs aux trames unes. Si aucun résultat stable n'est trouvé, il est très probable que le décalage spatial est correct mais que l'évaluation du décalage temporel est aberrante (c'est-à-dire qu'elle est éloignée de plus de deux images du décalage temporel réel). On procède alors à une recherche large du décalage temporel qui inclut la meilleure évaluation courante du décalage spatial. Cette recherche large permet généralement de corriger l'évaluation du décalage temporel. Lorsque cette recherche est terminée, son résultat est utilisé comme point de départ et on procède à un maximum de cinq recherches fines répétées. Si cette deuxième série de recherches fines n'aboutit pas à un résultat stable, on signale alors un échec d'alignement spatial pour l'image considérée (c'est-à-dire à la fois pour la trame une et pour la trame deux). Si cette deuxième série aboutit à un résultat stable, le décalage spatial et le décalage temporel pour cette trame sont stockés dans la matrice des trames unes. Par ailleurs, le décalage spatial et le décalage temporel utilisés comme point de départ pour la trame traitée suivante qui est une trame une sont mis à jour (autrement dit, on utilise les résultats de base pour la première trame traitée et, ensuite, on utilise le dernier résultat stable). Une fois que le décalage spatial a été évalué pour la première trame traitée qui est une trame une, on évalue le décalage spatial pour la première trame traitée qui est une trame deux. En utilisant les résultats spatiaux de la trame une comme point de départ, on applique les mêmes étapes pour trouver le décalage spatial de la trame deux (c'est-à-dire les trois recherches fines et, si nécessaire, une recherche large du décalage temporel suivie par cinq recherches fines répétées). Si un résultat stable est trouvé pour la trame deux, on stocke le décalage vertical et le décalage horizontal de la trame deux dans une matrice différente qui est réservée au stockage des résultats pour les trames deux.

On applique la procédure décrite dans le paragraphe ci-dessus pour évaluer le décalage spatial des deux trames de chaque image correspondant à la fréquence *F* du fichier Big *YUV* qui contient la séquence vidéo traitée. On saute les premières images du fichier Big *YUV* traité correspondant à l'incertitude temporelle, *U*. On utilise alors cette séquence d'évaluations pour calculer une évaluation robuste du décalage spatial pour chaque type de trame de la scène considérée. On trie les résultats de décalage vertical de la trame une de chaque image et on retient la valeur du $50^{ème}$ percentile comme valeur globale du décalage vertical pour les trames unes. De même, on trie les résultats de décalage vertical de la trame deux de chaque image et on retient la valeur du $50^{ème}$ percentile comme valeur globale du décalage vertical pour les trames deux. On trie les résultats de décalage horizontal de la trame une de chaque image et on retient la valeur du $50^{ème}$ percentile comme valeur globale du décalage horizontal. Toute différence entre le décalage horizontal des trames unes et celui des trames deux est très probablement due à un décalage horizontal sous-pixel (par exemple un décalage horizontal de 0,5 pixel). Les décalages horizontaux sous-pixel conduisent à des évaluations qui incluent les deux décalages les plus proches. L'utilisation de la valeur du $50^{ème}$ percentile permet de choisir le décalage horizontal le plus probable, conduisant à une précision de l'alignement spatial à 0,5 pixel près[4].

[4] Un alignement spatial à 0,5 pixel près est suffisant pour les mesures de la qualité vidéo décrites dans la présente Annexe. Les techniques d'alignement spatial sous-pixel sortent du cadre de la présente Annexe.

6.1.5.7 Algorithme pour un HRC donné

Si plusieurs scènes sont passées par le même HRC, les résultats de l'alignement spatial pour chaque scène devraient être identiques. Ainsi, le filtrage des résultats obtenus pour de multiples scènes permet d'augmenter la robustesse et la précision des mesures du décalage spatial. On peut alors utiliser les résultats globaux d'alignement spatial obtenus pour le HRC considéré pour procéder à une compensation pour toutes les séquences vidéo traitées par ce HRC.

6.1.5.8 Commentaires concernant l'algorithme

Certaines scènes vidéo ne conviennent pas vraiment pour l'évaluation de l'alignement spatial. L'algorithme décrit aura parfois pour résultat un faux minimum. D'autres fois, il errera entre plusieurs solutions et ne donnera jamais de résultat stable. C'est pourquoi il est conseillé d'examiner de multiples images d'une même scène et de déterminer la valeur médiane (c'est-à-dire de trier les résultats de la valeur la plus faible à la valeur la plus élevée et de choisir la valeur du 50ème percentile) de ces résultats sur plusieurs scènes. L'algorithme d'alignement spatial fondé sur des scènes est un algorithme heuristique utilisant les décalages spatiaux qui ont été observés pour un échantillon de systèmes vidéo. Ces hypothèses peuvent être incorrectes pour certains systèmes, auquel cas l'algorithme détermine un décalage spatial incorrect. Toutefois, lorsque l'algorithme donne des résultats incorrects, il a tendance à produire des décalages spatiaux qui sont incohérents d'une image à l'autre et d'une scène à l'autre (autrement dit, lorsque l'algorithme donne des résultats incorrects, il produit généralement des résultats épars). Lorsque l'algorithme a pour résultat le même décalage spatial ou des décalages spatiaux très semblables pour chaque scène, cela indique un niveau de confiance élevé. En cas de résultats épars pour les trames d'une scène donnée, cela indique un niveau de confiance faible.

6.1.6 Alignement spatial d'un flux vidéo avec balayage progressif

L'alignement spatial d'un flux vidéo avec balayage progressif suit le même algorithme que dans le cas d'un flux vidéo avec balayage à entrelacement, avec quelques légères modifications. L'algorithme dans le cas du balayage avec entrelacement s'applique séparément à la trame une et à la trame deux, alors que l'algorithme dans le cas du balayage progressif s'applique à l'image entière. Ainsi, il faut ignorer toutes les mentions de trame deux et, à l'exception des recherches fines, il faut doubler la plage des décalages verticaux.

La modification de la plage des décalages verticaux est particulièrement importante pour la recherche large du décalage spatial. Pour une telle recherche (voir le § 6.1.5.3), il faut doubler les nombres sur l'axe vertical de la Fig. 25 (par exemple +8 devient +16 et –4 devient –8)[5]. Par ailleurs, dans le cas des images CIF et QCIF à balayage progressif, les plages de décalage horizontal et de décalage vertical utilisées pour les recherches larges sont réduites de moitié car les décalages observés avec ces formats d'image sont généralement plus petits. Par exemple, dans le cas d'images CIF, l'axe horizontal de la Fig. 25 irait de –6 à +6 pixels et l'axe vertical irait de –8 à +8 lignes d'image.

La plage utilisée pour la recherche du décalage temporel, spécifiée en nombre d'images, reste essentiellement la même. Pour la recherche large du décalage temporel décrite au § 6.1.5.2, au lieu de comparer une trame traitée une avec une trame d'origine une sur deux, l'algorithme dans le cas du balayage progressif compare une image traitée avec une image d'origine sur deux. Concernant l'algorithme pour la mire chromatique, la recherche examine les décalages spatiaux entre une seule image traitée et une seule image d'origine (autrement dit il n'y a pas de recherche de décalage temporel).

---

[5] Il existe une exception possible à ce doublement: le décalage spatial de zéro pixel horizontalement et de plus ou moins une ligne de trame verticalement peut être laissé à plus ou moins une ligne d'image verticalement. Les décalages spatiaux très proches de (zéro, zéro) sont fréquents.

La seule étape qui nécessite des modifications plus complexes est l'étape de recherche fine du § 6.1.5.4. Dans cette étape, les décalages verticaux restent inchangés, compris entre –2 lignes d'image et +2 lignes d'image. Ainsi, les nombres représentés sur l'axe vertical de la Fig. 26 sont interprétés comme étant des nombres de lignes d'image. On peut définir la plage des décalages temporels pour cette recherche fine comme comprenant les cinq images d'origine centrées sur l'image d'origine courante, au lieu des trois images d'origine susmentionnées. Une plage de cinq images peut améliorer la vitesse et l'efficacité de la recherche fine par rapport à l'algorithme dans le cas du balayage à entrelacement, car les HRC à balayage progressif ont davantage tendance à engendrer des retards vidéo plutôt que des décalages spatiaux non nuls.

Lorsqu'on examine les modifications à apporter à l'algorithme utilisé pour les systèmes vidéo à balayage progressif, il est possible de modifier de nombreux paramètres utilisés pour la recherche du décalage spatial sans compromettre l'intégrité de l'algorithme. Considérons, à titre d'exemple, les décalages spatiaux autres que zéro pixel et zéro ligne utilisés pour la recherche large du décalage temporel. Le décalage spatial de zéro pixel horizontalement et de 8 lignes de trame verticalement utilisé pour les systèmes à balayage à entrelacement peut être porté à 16 lignes d'image pour les systèmes à balayage progressif, comme recommandé plus haut, ou fixé à 8 lignes d'image, si on suppose qu'il est peu probable que des séquences vidéo à balayage progressif contiennent un décalage vertical de 16 lignes d'image. De même, un décalage spatial de zéro ligne verticalement et de 8 pixels horizontalement peut être porté à 9 ou 10 pixels horizontalement sans effets préjudiciables. Autre exemple: le nombre exact de répétitions de la recherche fine peut être augmenté ou diminué pour des applications particulières. Les valeurs exactes recommandées ici sont nettement moins élevées que dans la structure réelle de l'algorithme de recherche.

6.2 Région valable

Les séquences vidéo NTSC (525 lignes) et PAL (625 lignes) échantillonnées conformément à la Recommandation UIT-R BT.601 sont susceptibles d'avoir une bordure de pixels et de lignes qui ne contient pas d'information d'image. Il est possible que la séquence vidéo d'origine saisie par la caméra ne remplisse qu'une partie de l'image telle qu'elle est définie dans la Recommandation UIT-R BT.601. Un système vidéo numérique qui utilise une compression risque de réduire encore la zone de l'image afin de réduire le nombre de bits transmis. Si les pixels et les lignes qui ne sont pas transmis se trouvent dans la zone de surbalayage de l'image de télévision, l'utilisateur final ne devrait pas remarquer qu'il manque des lignes et des pixels. Si les pixels et les lignes qui ne sont pas transmis dépassent la zone de surbalayage, l'observateur pourra remarquer une bordure noire tout autour de l'image, car le système insérera généralement du noir dans cette zone d'image non transmise. Les systèmes vidéo (notamment ceux qui procèdent à un filtrage passe-bas) risquent de causer une avancée de la bordure noire dans la zone d'image. La plupart du temps, ces effets transitoires ont lieu à gauche et à droite de l'image mais ils peuvent aussi avoir lieu en haut ou en bas. Par ailleurs, la séquence vidéo traitée peut parfois contenir plusieurs lignes de données vidéo altérées en haut ou en bas de l'image que l'observateur ne verra pas nécessairement (les magnétoscopes VHS altèrent plusieurs lignes en bas de l'image dans la zone de surbalayage). Afin d'éviter que les zones ne contenant pas d'information d'image aient une incidence sur les mesures de la qualité VQM, il convient d'exclure ces zones de ces mesures. L'algorithme automatisé de la région valable présenté ici évalue la région valable du flux vidéo d'origine et du flux vidéo traité de sorte que, pour les calculs suivants, on ne tienne pas compte des lignes altérées en haut et en bas de l'image telle qu'elle est définie dans la Recommandation UIT-R BT.601, des pixels de la bordure noire ou des effets transitoires où la bordure noire avance dans la zone d'image.

6.2.1 Algorithme principal de la région valable

Le présent paragraphe décrit l'algorithme principal de la région valable qui est appliqué à une seule image d'origine ou traitée. Cet algorithme nécessite trois arguments d'entrée: une image, une région valable maximale et l'évaluation de la région valable courante.

– *Image*: l'algorithme principal utilise l'image de luminance définie dans la Recommandation UIT-R BT.601 associée à une seule image vidéo. Pour la mesure de la région valable d'une séquence vidéo traitée, tout décalage spatial imposé par le système vidéo doit avoir été supprimé de l'image de luminance avant que l'algorithme principal ne soit appliqué (voir le § 6.1).

– *Région valable maximale*: l'algorithme principal ne tiendra pas compte des pixels et des lignes qui se trouvent en dehors d'une région vidéo valable maximale. Cela permet à l'utilisateur de spécifier une région valable maximale qui est plus petite que la zone entière de l'image si des informations a priori indiquent que des pixels ou des lignes de l'image échantillonnée ont été altérés (voir le § 6.2).

– *Région valable courante*: la région valable courante est une évaluation de la région valable qui est entièrement comprise dans la région valable maximale. Tous les pixels de la région valable courante contiennent une information vidéo valable; les pixels qui sont situés en dehors de cette région contiennent une information vidéo qui peut être soit valable soit non valable. Au départ, on prend, comme région valable courante, la plus petite zone possible située exactement au centre de l'image.

L'algorithme principal examine la zone vidéo comprise entre la région valable maximale et la région valable courante. Si certains de ces pixels contiennent une information vidéo valable, la région valable courante est élargie. L'algorithme est alors décrit en détail pour la partie gauche de l'image.

Etape 1: Calculer le niveau moyen de la colonne de pixels la plus à gauche de la région valable maximale. Cette colonne est désignée par J-1 et la moyenne est représentée par $M_{J\text{-}1}$.

Etape 2: Calculer le niveau moyen de la colonne de pixels suivante, M_J.

Etape 3: La colonne J est déclarée comme contenant des informations vidéo non valables si elle est noire ($M_J < 20$) ou si le niveau moyen des pixels pour des colonnes successives indique une avancée de la bordure noire dans l'image valable ($M_J - 2 > M_{J\text{-}1}$). Si l'une de ces conditions est remplie, incrémenter J et répéter les Etapes 2 et 3. Dans les autres cas, aller à l'Etape 4.

Etape 4: Si la colonne finale J se trouve dans la région valable courante, aucune nouvelle information n'a été obtenue. Dans le cas contraire, mettre à jour la région valable courante avec J comme coordonnée de gauche.

L'algorithme permettant de déterminer le haut de l'image est analogue à celui qui est présenté ci-dessus pour la partie gauche. Pour le bas et la partie droite, J est décrémenté au lieu d'être incrémenté; à cette exception près, l'algorithme est le même. Les valeurs obtenues pour le haut, la gauche, le bas et la droite désignent le dernier pixel ou la dernière ligne valable.

Le contenu de la scène peut être tel que l'une des conditions spécifiées à l'Etape 3 est remplie alors qu'elle ne devrait pas l'être. Par exemple, dans le cas d'une image qui contient du noir intentionnel dans la partie gauche (autrement dit du noir qui fait partie de la scène), l'algorithme principal conclura que la colonne vidéo valable la plus à gauche est beaucoup plus proche du milieu de l'image qu'elle ne devrait être. C'est pourquoi l'algorithme principal est appliqué à de multiples images issues d'une séquence vidéo, ce qui permet d'accroître la précision de l'évaluation de la région valable.

6.2.2 Application de l'algorithme principal de la région valable à une séquence vidéo

6.2.2.1 Séquence vidéo d'origine

L'algorithme principal est d'abord appliqué à la séquence d'images d'origine. Concernant les séquences vidéo NTSC échantillonnées conformément à la Recommandation UIT-R BT.601 (voir le § 5), il est recommandé que la région valable maximale soit telle que haut = 6, gauche = 6, bas = 482, droite = 714. Concernant les séquences vidéo PAL échantillonnées conformément à la Recommandation UIT-R BT.601, il est recommandé que la région valable maximale soit telle que haut = 6, gauche = 16, bas = 570, droite = 704. L'algorithme principal est appliqué à la première image de la séquence vidéo et à chaque image ultérieure correspondant à une certaine fréquence. Par exemple, si la fréquence spécifiée vaut 15, l'algorithme principal examine les images de la séquence numéros 0, 15, 30, 45, etc. Une fois que toutes les images de la séquence ont été examinées, la région valable courante contient la plus grande zone valable parmi toutes les images examinées dans la séquence vidéo. Les pixels et les lignes qui sont compris entre cette région valable courante finale et la région valable maximale sont considérés comme contenant du noir ou une avancée transitoire du noir.

La région valable finale doit contenir un nombre pair de lignes et un nombre pair de pixels. Si la coordonnée du haut est impaire, elle est incrémentée de un. De même, si la coordonnée de gauche est impaire, elle est incrémentée de un. Ensuite, si la région contient un nombre impair de lignes, on décrémente la coordonnée du bas; de même, si la région contient un nombre impair de pixels (horizontalement), on décrémente la coordonnée de droite. Cela permet de simplifier le traitement chromatique des séquences vidéo échantillonnées conformément à la Recommandation UIT-R BT.601, car la fréquence d'échantillonnage des canaux de couleur vaut la moitié de la fréquence d'échantillonnage du canal de luminance. Par ailleurs, chaque trame vidéo entrelacée contient le même nombre de lignes vidéo. Cela permet de garantir que les sous-régions spatio-temporelles (à partir desquelles les caractéristiques sont extraites) contiennent toujours des informations vidéo valables avec des contributions égales des deux trames entrelacées. La région valable résultante est retournée comme étant la région valable d'origine.

6.2.2.2 Séquence vidéo traitée

Pour le calcul de la région valable de la séquence vidéo traitée, on considère d'abord que la région valable maximale pour l'algorithme principal est égale à la région valable d'origine correspondante déterminée pour cette scène. On réduit ensuite la taille de cette région valable maximale en supprimant les pixels et les lignes considérés comme non valables par suite de l'alignement spatial des images vidéo traitées. L'algorithme principal est alors appliqué à la première image de la séquence vidéo traitée et à chaque image ultérieure correspondant à une certaine fréquence (si la fréquence vaut F, on utilise les images $Y(0)$, $Y(F)$, $Y(2F)$, $Y(3F)$, etc.).

Une fois que l'algorithme principal a été appliqué à la séquence vidéo traitée, la région valable déterminée par l'algorithme principal est réduite vers l'intérieur par une marge de sécurité. La marge de sécurité recommandée est de une ligne en haut et en bas et de cinq pixels à gauche et à droite. Les valeurs élevées à gauche et à droite permettent de garantir que toute avancée et tout recul du noir sont exclus de la région valable traitée.

La région valable traitée finale doit contenir un nombre pair de lignes et un nombre pair de pixels. Si la coordonnée du haut est impaire, elle est incrémentée de un. De même, si la coordonnée de gauche est impaire, elle est incrémentée de un. Ensuite, si la région contient un nombre impair de lignes, on décrémente la coordonnée du bas; de même, si la région contient un nombre impair de pixels (horizontalement), on décrémente la coordonnée de droite. La région valable résultante est retournée comme étant la région valable traitée.

6.2.3 Commentaires concernant l'algorithme de la région valable

Cet algorithme automatisé permet d'évaluer correctement la région valable de la plupart des scènes. En raison du très grand nombre de possibilités de contenu pour une scène, l'algorithme décrit ici est fondé sur une approche prudente de l'évaluation de la région valable. Un examen manuel de la région valable conduirait certainement au choix d'une région plus grande. Les évaluations prudentes de la région valable conviennent mieux pour un système automatisé de mesure de la qualité vidéo, car l'élimination d'une faible quantité de contenu vidéo aura une faible incidence sur l'évaluation de la qualité et, de toute manière, ce contenu vidéo éliminé se trouvait généralement dans la zone de surbalayage de la séquence vidéo. En revanche, la prise en considération d'un contenu vidéo altéré dans les calculs de la qualité vidéo risque d'avoir une forte incidence sur l'évaluation de la qualité.

Cet algorithme ne contient pas une intelligence artificielle suffisante pour faire la distinction entre des pixels et des lignes altérés au bord d'une image et un véritable contenu de la scène. A la place, on utilise une règle empirique, selon laquelle un tel contenu vidéo non valable est généralement situé aux bords extrêmes de l'image. La spécification d'une région vidéo valable maximale prudente définissable par l'utilisateur (c'est-à-dire le point de départ de l'algorithme automatisé) permet de ne pas prendre en considération ces bords d'image éventuellement altérés.

Lorsque l'algorithme de la région valable est appliqué à une séquence vidéo qui n'est pas échantillonnée conformément à la Recommandation UIT-R BT.601 (par exemple le format intermédiaire commun, ou CIF, utilisé par la Recommandation UIT-T H.261), il est recommandé de prendre l'image entière comme région valable maximale lors de l'examen de la séquence vidéo d'origine. Dans ces cas, la séquence vidéo échantillonnée ne contient généralement pas de zone de surbalayage altérée, il est donc inutile de prendre une région valable maximale plus petite que l'image entière.

6.3 Gain et décalage

6.3.1 Algorithme principal du gain et du décalage de niveau

Le présent paragraphe expose la méthode à utiliser pour étalonner le gain et le décalage de niveau. Pour pouvoir appliquer cet algorithme, l'image d'origine et l'image traitée doivent être alignées spatialement (voir le § 6.1). Elles doivent aussi être alignées temporellement (voir plus loin le § 6.4). L'étalonnage du gain et du décalage de niveau peut être appliqué aux trames ou aux images, selon le cas.

Dans la méthode présentée ici, on suppose que chacun des signaux Y, C_B et C_R de la Recommandation UIT-R BT.601 a un gain et un décalage de niveau indépendants. Cette hypothèse sera généralement suffisante pour l'étalonnage dans le cas des systèmes vidéo en composantes (par exemple Y, R-Y, B-Y). Toutefois, dans le cas des systèmes composites ou *S*-vidéo, il est possible d'avoir une rotation de phase des informations de chrominance car les deux composantes de chrominance sont multiplexées dans un vecteur de signal complexe comprenant une amplitude et une phase. L'algorithme présenté ici ne permet pas de procéder à un étalonnage correct dans le cas des systèmes vidéo qui introduisent une rotation de phase des informations de chrominance (par exemple l'ajustement de teinte sur un poste de télévision).

– Comme indiqué précédemment, on suppose dans ce modèle d'étalonnage qu'il n'existe aucun couplage croisé entre les trois composantes vidéo. Cela étant, l'algorithme principal d'étalonnage est appliqué de manière indépendante à chacun des trois canaux: Y, C_B et C_R.

– La région valable du plan d'image d'origine et celle du plan d'image traitée sont d'abord subdivisées en N sous-régions. Pour chacune des sous-régions, on calcule la valeur moyenne *origine* et la valeur moyenne *traité* (moyenne dans l'espace). On représente ensuite ces valeurs sous la forme des vecteurs colonnes à N éléments $\underline{O}$ et $\underline{P}$, respectivement:

$$\underline{O}_{N\times 1} = \begin{bmatrix} origine_1 \\ . \\ . \\ . \\ origine_N \end{bmatrix}, \quad \underline{P}_{N\times 1} = \begin{bmatrix} traité_1 \\ . \\ . \\ . \\ traité_N \end{bmatrix}$$

Pour l'étalonnage, il faut calculer le gain, g, et le décalage de niveau, l, conformément au modèle suivant:

$$\underline{P} = g\underline{O} + l$$

Comme il n'y a que deux inconnues (g et l) mais N équations (N sous-régions), il faut résoudre le système d'équations linéaires surdéterminé donné par:

$$\underline{\hat{P}} = A \begin{bmatrix} l \\ g \end{bmatrix}$$

où A est une matrice $N \times 2$ donnée par $A_{N\times 2} = [\underline{1} \quad \underline{O}]$, et $\underline{1}$ est un vecteur colonne à N éléments valant «1» donné par:

$$\underline{1}_{N\times 1} = \begin{bmatrix} 1_1 \\ . \\ . \\ . \\ 1_N \end{bmatrix}$$

$\hat{P}$ est l'évaluation des échantillons traités découlant de l'application du gain et du décalage de niveau aux échantillons d'origine. La solution donnée par les moindres carrés à ce problème surdéterminé (à condition que $N > 2$) est donnée par:

$$\begin{bmatrix} l \\ g \end{bmatrix} = \left(A^{\mathrm{T}} A\right)^{-1} A^{\mathrm{T}} P$$

où l'exposant, T, désigne la transposée de la matrice et l'exposant, –1, désigne l'inverse de la matrice.

Lorsque l'algorithme principal du gain et du décalage de niveau est appliqué de manière indépendante à chacun des trois canaux, six grandeurs sont évaluées: gain Y, décalage Y, gain C_B, décalage C_B, gain C_R et décalage C_R.

6.3.2 Utilisation de scènes

L'algorithme de base donné au § 6.3.1 peut être appliqué à des flux vidéo d'origine et traité sous réserve qu'ils aient été alignés spatialement et temporellement. Cette technique fondée sur les scènes subdivise l'image en blocs contigus de niveau d'intensité inconnu. Une taille de sous-région de 16 lignes × 16 pixels est recommandée pour les images (c'est-à-dire 8 lignes × 16 pixels pour une trame NTSC ou PAL Y; 8 lignes × 8 pixels pour C_B et C_R en raison du sous-échantillonnage des plans de couleur). La moyenne dans l'espace des échantillons [Y, C_B, C_R] est calculée pour chaque sous-région ou bloc d'origine et sous-région ou bloc traité correspondant, afin de former une image sous-échantillonnée spatialement. Tous les blocs choisis doivent se trouver dans la région PVR.

6.3.2.1 Alignement des images traitées

Dans un souci de simplicité, on suppose que le meilleur alignement spatial a déjà été déterminé au moyen de l'une des techniques présentées au § 6.1. Pour pouvoir évaluer le gain et le décalage de niveau, chaque image traitée doit être alignée temporellement. L'image d'origine qui correspond le mieux à l'image traitée doit être utilisée pour le calcul du gain et du décalage de niveau. Si le retard vidéo est variable, cet alignement temporel doit être opéré pour chaque image traitée. Si le retard vidéo est constant pour la scène, il n'est nécessaire d'opérer l'alignement temporel qu'une seule fois.

Pour aligner temporellement une image traitée, on commence par créer les trames d'origine et traitée sous-échantillonnées spatialement (ou les images dans le cas du balayage progressif) comme spécifié au § 6.3.2, après avoir corrigé le décalage spatial du flux vidéo traité. En utilisant les images *Y* sous-échantillonnées, on applique la fonction de recherche donnée au § 6.1.4.3, à l'exception d'effectuer cette recherche en utilisant toutes les images d'origine correspondant à l'incertitude d'alignement temporel, *U*. On utilise le meilleur alignement temporel résultant pour les trois plans d'image, *Y*, C_B et C_R.

6.3.2.2 Gain et décalage de niveau des images alignées

On utilise une solution itérative donnée par les moindres carrés avec une fonction de coût afin de réduire au minimum le poids des valeurs aberrantes dans l'ajustement. En effet, les valeurs aberrantes sont généralement dues à des distorsions et non à de simples modifications du décalage de niveau et du gain, de sorte que l'attribution d'un poids égal à ces valeurs aberrantes conduirait à une distorsion de l'ajustement.

L'algorithme suivant est appliqué séparément aux *N* pixels d'origine et traités correspondants issus de chacune des trois images sous-échantillonnées spatialement [*Y*, C_B, C_R].

Etape 1: Utiliser la solution normale donnée par les moindres carrés (voir le § 6.3.1) pour générer l'évaluation initiale du décalage de niveau et du gain: $\begin{bmatrix} l \\ g \end{bmatrix} = \left(A^{\mathrm{T}} A\right)^{-1} A^{\mathrm{T}} \underline{P}$.

Etape 2: Générer un vecteur d'erreur, $\underline{E}$, qui est égal à la valeur absolue de la différence entre les échantillons traités réels et les échantillons traités ajustés: $\underline{E} = \left|\underline{P} - \underline{\hat{P}}\right|$.

Etape 3: Générer un vecteur de coût, $\underline{C}$, dont chaque élément est le réciproque de l'élément correspondant du vecteur d'erreur, *E*, plus un petit epsilon, ε: $\underline{C} = \dfrac{1}{\underline{E} + \varepsilon}$. ε permet d'éviter la division par zéro et définit le poids relatif d'un point qui est sur la courbe ajustée par rapport au poids d'un point qui est en dehors de cette courbe. Il est recommandé d'utiliser une valeur de 0,1 pour ε.

Etape 4: Normaliser le vecteur de coût *C* (autrement dit, on divise chaque élément de *C* par la racine carrée de la somme des carrés de tous les éléments de *C*).

Etape 5: Générer le vecteur de coût *C* 2 dont chaque élément est le carré de l'élément correspondant du vecteur de coût *C* issu de l'Etape 4.

Etape 6: Générer une matrice de coût diagonale $N \times N$, *C* 2, qui contient les éléments du vecteur de coût, *C* 2, sur la diagonale et des zéros partout ailleurs.

Etape 7: En utilisant la matrice de coût diagonale, *C* 2, issue de l'Etape 6, procéder à un ajustement par les moindres carrés avec pondération par le coût pour déterminer l'évaluation suivante du décalage de niveau et du gain: $\begin{bmatrix} l \\ g \end{bmatrix} = \left(A^{\mathrm{T}} C^2 A\right)^{-1} A^{\mathrm{T}} C^2 \underline{P}$.

Etape 8: Répéter les Etapes 2 à 7 jusqu'à ce que les évaluations du décalage de niveau et du gain convergent à la quatrième décimale près.

Ces étapes sont appliquées séparément à la trame traitée une et à la trame traitée deux, ce qui donne deux évaluations de *g* et deux évaluations de *l*. Il faut examiner séparément la trame une et la trame deux, car les trames d'origine alignées temporellement ne correspondent pas nécessairement à une même image dans la séquence vidéo d'origine. Dans le cas des systèmes vidéo à balayage progressif, les étapes ci-dessus sont appliquées à l'image traitée tout entière.

6.3.2.3 Evaluation du gain et du décalage de niveau pour une séquence vidéo et un HRC

L'algorithme décrit ci-dessus est appliqué à plusieurs couples trame d'origine-trame traitée correspondante répartis tout au long de la scène avec une certaine fréquence (dans le cas des systèmes vidéo à balayage progressif, on utilise des couples image d'origine-image traitée). On détermine alors la valeur médiane de chacun des six historiques temporels de décalages de niveau et de gains pour produire des évaluations moyennes pour la scène.

Si plusieurs scènes passent par le même HRC, le décalage de niveau et le gain pour chaque scène seront considérés comme identiques. Ainsi, les valeurs médianes obtenues à partir de plusieurs scènes permettent d'augmenter la robustesse et la précision des mesures de décalage de niveau et de gain. On peut alors utiliser les résultats globaux de décalage de niveau et de gain obtenus pour le HRC considéré pour procéder à une compensation pour tous les flux vidéo traités par ce circuit.

6.3.3 Application des corrections de gain et de décalage de niveau

Pour les algorithmes d'alignement temporel (voir le § 6.4) et pour l'extraction de la plupart des caractéristiques de qualité (voir le § 7), il convient de supprimer le gain calculé ici. Pour supprimer le gain et le décalage de niveau du plan Y, on applique la formule suivante à chaque pixel traité:

$$\text{Nouveau } Y(i,j,t) = [Y(i,j,t) - 1]/g$$

Le gain et le décalage de niveau des plans de couleur (C_B et C_R) ne sont pas corrigés. A la place, on mesure les erreurs de chrominance perçues. Le gain et le décalage de niveau des plans d'image C_B et C_R peuvent être corrigés à des fins d'affichage.

6.4 Alignement temporel

Les systèmes de communication vidéo numériques modernes ont généralement besoin de plusieurs dixièmes de seconde pour traiter et transmettre le flux vidéo de la caméra au dispositif de visualisation. Des retards vidéo excessifs empêchent d'avoir une communication bidirectionnelle efficace. Les méthodes de mesure objective du retard de bout en bout pour les communications vidéo sont donc importantes pour les utilisateurs finals afin de pouvoir spécifier et comparer les services ainsi que pour les fournisseurs d'équipements/de services afin de pouvoir optimiser et mettre à jour leurs offres de produits. Le retard vidéo peut dépendre des attributs dynamiques de la scène d'origine (par exemple détail spatial, mouvement) et du système vidéo (par exemple débit binaire). A titre d'exemple, le retard vidéo risque d'être plus grand pour des scènes comportant beaucoup de mouvements que pour des scènes en comportant peu. Les mesures du retard vidéo devraient donc être faites en service afin d'être vraiment représentatives et précises. Il est nécessaire d'évaluer le retard vidéo pour pouvoir aligner temporellement les caractéristiques vidéo du flux d'origine et du flux traité (voir la Fig. 19) avant de procéder aux mesures de la qualité.

Certains systèmes de transmission vidéo peuvent fournir des informations de synchronisation temporelle (les images d'origine et traitées peuvent par exemple être étiquetées au moyen d'un certain type de système de numérotation d'image). Toutefois, la synchronisation temporelle entre le flux vidéo d'origine et le flux vidéo traité doit généralement être mesurée. Le présent paragraphe

expose une technique permettant d'évaluer le retard vidéo sur la base des images vidéo d'origine et des images vidéo traitées. La technique est «fondée sur les images» en ce sens qu'elle consiste à corréler des images à plus faible résolution, sous-échantillonnées dans l'espace et extraites des flux vidéo d'origine et traité. Cette technique fondée sur les images évalue le retard de chaque image ou de chaque trame (dans le cas des systèmes vidéo avec balayage à entrelacement). On combine ces différentes évaluations pour évaluer le retard moyen pour la séquence vidéo.

6.4.1 Algorithme fondé sur les images pour évaluer les décalages temporels variables entre une séquence vidéo d'origine et une séquence vidéo traitée

Le présent paragraphe décrit un algorithme d'alignement temporel fondé sur les images. Pour réduire l'influence des distorsions sur l'alignement temporel, les images sont sous-échantillonnées spatialement et normalisées de manière à avoir une variance unitaire. Cet algorithme permet d'aligner temporellement chaque image traitée séparément, en localisant l'image d'origine la plus analogue. Certaines de ces différentes mesures d'alignement temporel peuvent être incorrectes mais les erreurs ont tendance à être distribuées aléatoirement. Lorsqu'on attribue les mesures du retard issues d'une série d'images au moyen d'un système de vote, on obtient une évaluation globale du retard moyen d'une séquence vidéo relativement précise. Cet algorithme d'alignement temporel n'utilise pas les parties fixes ou pratiquement sans mouvement de la scène, car les images d'origine sont pratiquement identiques les unes aux autres.

6.4.1.1 Constantes utilisées par l'algorithme

BELOW_WARN: Seuil utilisé lors de l'examen des corrélations afin de décider si un maximum de corrélation secondaire est suffisamment grand pour indiquer un alignement temporel ambigu. Il est recommandé d'utiliser une valeur de 0,9 pour BELOW_WARN.

BLOCK_SIZE: Facteur de sous-échantillonnage, spécifié en nombre de lignes d'image verticalement et en nombre de pixels horizontalement. Il est recommandé d'utiliser une valeur de 16 pour BLOCK_SIZE.

DELTA: Les maximums secondaires de la courbe de corrélation qui sont éloignés de moins de DELTA de la (meilleure) corrélation maximale sont ignorés. Il est recommandé d'utiliser une valeur de 4 pour DELTA.

HFW: La moitié de la largeur du filtre utilisé pour lisser l'histogramme des valeurs d'alignement temporel associées à chaque image. Il est recommandé d'utiliser une valeur de 3 pour HFW.

STILL_THRESHOLD: Seuil utilisé pour détecter les scènes vidéo fixes (l'alignement temporel fondé sur les images ne peut pas être utilisé pour des scènes vidéo fixes). Il est recommandé d'utiliser une valeur de 0,002 pour STILL_THRESHOLD.

6.4.1.2 Variables d'entrée de l'algorithme

Une séquence de N images de luminance du flux vidéo d'origine: $\boldsymbol{Y_O}(t)$, $0 \le t < N$ [6].

Une séquence de N images de luminance du flux vidéo traité: $\boldsymbol{Y_P}(t)$, $0 \le t < N$.

Facteurs de gain et de décalage de niveau pour les images de luminance traitées.

---

[6] Lorsqu'un flux vidéo avec balayage à entrelacement nécessite une resynchronisation de trame, la longueur des séquences d'origine et traitée doit être réduite de un afin de tenir compte de la resynchronisation de trame. La longueur du fichier sera donc ramenée à $N-1$ images vidéo (voir la Fig. 20).

Informations d'alignement spatial: décalage horizontal et décalage vertical. Dans le cas des systèmes vidéo avec balayage à entrelacement, le décalage vertical pour chaque trame permet de déterminer si le flux vidéo traité nécessite une resynchronisation de trame.

Région valable de la séquence vidéo traitée (PVR).

Incertitude (U): nombre indiquant la précision de l'alignement temporel initial. On suppose au départ que le véritable alignement temporel pour $\boldsymbol{Y_P}(t)$ est compris entre plus ou moins (U – HFW) de $\boldsymbol{Y_O}(t)$, pour $0 \le t < N$.

6.4.1.3 Images ou trames

L'algorithme d'alignement temporel fondé sur les images fonctionne à la fois pour les systèmes vidéo avec balayage à entrelacement et pour les systèmes vidéo avec balayage progressif. En cas de séquence vidéo avec balayage progressif, l'algorithme aligne des images. En cas de séquence vidéo avec balayage à entrelacement, l'algorithme aligne des trames. Lors de l'alignement de séquences vidéo avec balayage à entrelacement, soit des alignements d'image soit des alignements de trames resynchronisées sont considérés, mais pas les deux. Lorsque des alignements d'image sont considérés, la trame une de l'image vidéo traitée est comparée avec la trame une de l'image vidéo d'origine et la trame deux de l'image vidéo traitée est comparée avec la trame deux de l'image vidéo d'origine. Lorsque des alignements de trames resynchronisées sont considérés, la trame une de l'image vidéo traitée est comparée avec la trame deux de l'image vidéo d'origine et la trame deux de l'image vidéo traitée est comparée avec la trame une de l'image vidéo d'origine. Les valeurs d'alignement spatial fournies comme valeurs d'entrée de l'algorithme déterminent si ce sont des alignements d'image ou des alignements de trames resynchronisées qui sont considérés. Pour détecter la présence d'une resynchronisation de trame, on examine l'alignement spatial vertical pour chaque trame. Si le décalage vertical de la trame une est égal au décalage vertical de la trame deux, la séquence vidéo traitée n'a pas été soumise à une resynchronisation de trame; seuls des alignements d'image sont considérés. Si le décalage vertical de la trame deux vaut un de plus que le décalage vertical de la trame une, seuls des alignements de trames resynchronisées sont considérés. Toutes les autres combinaisons de décalages verticaux témoignent de l'existence de problèmes qu'il convient de régler avant l'alignement temporel.

6.4.1.4 Description de l'algorithme

Etape 1: *Etalonner les séquences vidéo*

Il convient de corriger la séquence vidéo traitée, $\boldsymbol{Y_P}(t)$, en utilisant les informations d'alignement spatial et de gain-décalage données comme informations d'entrée de l'algorithme.

Etape 2: *Choisir la sous-région vidéo à utiliser*

La sous-région d'intérêt à utiliser par l'algorithme doit être un multiple de BLOCK_SIZE et doit être comprise dans la PVR. Il convient de choisir la plus grande sous-région qui remplit ces deux conditions et qui est la plus proche du centre de l'image. L'ensemble du traitement ultérieur portera uniquement sur les informations vidéo présentes dans cette sous-région d'intérêt choisie.

Etape 3: *Sous-échantillonner spatialement les images d'origine et traitées*

Il convient de sous-échantillonner spatialement la région d'intérêt de $\boldsymbol{Y_O}(t)$ et $\boldsymbol{Y_P}(t)$ par un facteur BLOCK_SIZE en calculant la moyenne de chaque bloc. Pour les images d'une séquence vidéo à balayage progressif, le sous-échantillonnage sera de BLOCK_SIZE horizontalement et verticalement, alors que pour les trames d'une séquence vidéo à balayage à entrelacement, le sous-échantillonnage sera de BLOCK_SIZE horizontalement et de BLOCK_SIZE/2 verticalement. A titre d'exemple, le sous-échantillonnage d'une séquence vidéo à balayage progressif par un BLOCK_SIZE de 16 prendra la moyenne de chaque bloc de 16 pixels par 16 lignes d'image, alors

que le sous-échantillonnage d'une séquence vidéo à balayage à entrelacement par un BLOCK_SIZE de 16 prendra la moyenne de chaque bloc de 16 pixels par 8 lignes de trame. Ce sous-échantillonnage permet de réduire l'incidence des dégradations sur le processus d'alignement temporel.

Etape 4: *Normaliser les images sous-échantillonnées*

Il convient de normaliser chaque image sous-échantillonnée par l'écart type de cette image. On sautera cette normalisation pour toute image pour laquelle l'écart type est inférieur à un (par exemple pour les images contenant une trame de couleur uniforme)[7]. Cette normalisation permet de réduire au minimum l'influence des fluctuations du contraste et de l'énergie de chaque image sur les résultats de l'alignement temporel. Après cette étape, la séquence vidéo d'origine et la séquence vidéo traitée sont respectivement désignées par $\boldsymbol{S_O}(t)$ et $\boldsymbol{S_P}(t)$, afin d'indiquer que les images ont été sous-échantillonnées et normalisées.

Etape 5: *Comparer les images traitées avec les images d'origine*

Il convient de comparer chaque image traitée $\boldsymbol{S_P}(t)$ avec les images d'origine $\boldsymbol{S_O}(t + d)$, où les valeurs valables de d sont les suivantes: $(-U \le d \le +U)$ et les valeurs valables de t sont les suivantes: $(U \le t < N - U)$. La comparaison entre une image traitée t et une image d'origine $t + d$, désignée par $\boldsymbol{C}_{td}$, est calculée comme étant l'écart type dans l'espace de l'image formée par la différence entre l'image d'origine $t + d$ et l'image traitée t: $\boldsymbol{C}_{td} = std_{space}(\boldsymbol{S_O}(t + d) - \boldsymbol{S_P}(t))$. Les comparaisons $\boldsymbol{C}_{td}$ permettent de corréler la $t^{\text{ième}}$ image traitée avec chaque image d'origine comprise dans une certaine plage d'incertitude d'alignement. Plus la valeur de $\boldsymbol{C}_{td}$ est faible, plus l'image traitée ressemble à l'image d'origine, étant donné qu'une plus grande partie de la variance d'image est annulée. La plage de t, $U \delta t < N - U$, couvre l'ensemble des images traitées pour lesquelles les images d'origine sont disponibles pour toute la plage d'incertitude d'alignement temporel.

Etape 6: *Vérifier globalement le degré de mouvement de la séquence vidéo*

Pour déterminer si la séquence contient suffisamment de mouvement, il convient de calculer la moyenne de $\boldsymbol{C}_{td}$ sur l'indice temporel t pour chaque d:

$$A_d = \frac{1}{N - 2 * U} \cdot \sum_{t=U}^{N-U-1} C_{td} \tag{100}$$

Cette sommation (100) porte sur l'ensemble des images vidéo traitées t pour lesquelles toutes les images d'origine associées à l'incertitude considérée sont disponibles. A_d contient une valeur pour chaque décalage temporel d considéré. Si (maximum(A_d) – minimum(A_d) < STILL_THRESHOLD), la scène ne contient pas suffisamment de mouvement pour l'alignement temporel fondé sur les images. La scène entière est fixe ou quasiment fixe. Les résultats de corrélation pour les différents retards vidéo sont alors tellement analogues que toute différence sera due au hasard et non à des mesures fiables. En cas de détection d'une séquence vidéo fixe, l'utilisateur en est averti et l'algorithme prend alors fin.

Etape 7: *Aligner temporellement chaque image traitée*

Pour chaque image traitée t ($U \delta t < N - U$), il convient de déterminer la valeur de d dans la plage d'incertitude temporelle $(-U \le d \le +U)$ qui minimise $\boldsymbol{C}_{td}$. En d'autres termes, pour chaque image traitée t, il convient de déterminer $d_{min}(t)$ tel que $\boldsymbol{C}_{t\ dmin(t)} \le \boldsymbol{C}_{td}$, quel que soit d. Le meilleur alignement temporel de l'image traitée t est donné par $d_{min}(t)$. La plupart du temps, l'alignement

---

[7] On saute la normalisation lorsque l'écart type est inférieur à un afin d'éviter toute amplification du bruit et d'éviter une éventuelle division par zéro pour les images qui contiennent un niveau d'intensité uniforme.

temporel indiqué pour chaque image est correct ou très proche de l'alignement correct. Les cas où l'alignement temporel est incorrect peuvent s'expliquer par diverses raisons (distorsion d'image, erreurs, bruit, mouvement insuffisant, etc.).

Etape 8: *Vérifier le degré de mouvement pour chaque image traitée*

Si, pour une image traitée t et pour toutes les valeurs de d ($-U \le d \le U$), maximum($\boldsymbol{C}_{td}$) – minimum($\boldsymbol{C}_{td}$) < STILL_THRESHOLD, alors $d_{min}(t)$ est indéfini pour cette image traitée t. Plus précisément, le mouvement est insuffisant autour de l'image t pour que l'alignement temporel fondé sur les images puisse fonctionner correctement.

Etape 9: *Etablir un histogramme de tous les alignements temporels définis*

Il convient d'établir un histogramme en utilisant toutes les valeurs définies de $d_{min}(t)$ avec $2 * U + 1$ bâtons, chaque bâton représentant un retard vidéo différent (de $-U$ à $+U$). Les valeurs de $d_{min}(t)$ qui sont indéfinies (par exemple images fixes) sont ignorées dans l'établissement de l'histogramme. Cet histogramme, désigné par H_d, est l'histogramme des décalages temporels pour toutes les images traitées qui contenaient suffisamment de mouvement pour pouvoir effectuer un alignement temporel valable. Chaque bâton de l'histogramme contient le nombre d'images traitées présentant un certain retard vidéo d, où d varie de $-U$ à $+U$.

Etape 10: *Lisser l'histogramme*

On lisse l'histogramme H_d en procédant à sa convolution avec un filtre passe-bas de longueur $2 * \text{HFW} + 1$ et défini à l'indice k par:

$$F_k = \frac{0{,}5 + 0{,}5 * \cos\left[\pi * (k - \text{HFW})/(1 + \text{HFW})\right]}{\sum_{i=0}^{2*\text{HFW}} \left\{0{,}5 + 0{,}5 * \cos\left[\pi * (i - \text{HFW})/(1 + \text{HFW})\right]\right\}} \qquad \text{pour} \quad 0 \le k \le 2 * \text{HFW} \quad (101)$$

Concernant l'histogramme lissé (101) SH_d résultant de cette étape, les HFW bâtons à chaque extrémité de SH_d sont considérés comme indéfinis. Cela restreint les retards vidéo qui peuvent être évalués à plus ou moins (UNCERTAINTY-HFW). Le lissage de l'histogramme permet d'augmenter la robustesse des évaluations du retard vidéo.

Etape 11: *Examiner les informations de l'histogramme*

A partir de l'histogramme d'origine H_d et de l'histogramme lissé SH_d, on détermine les trois valeurs suivantes:

max_H_value: valeur maximale de H_d.
max_SH_offset: décalage d qui maximise SH_d.
max_SH_value: valeur maximale de SH_d (c'est-à-dire pour d = max_SH_offset).

On procède ensuite aux deux vérifications suivantes:

– La valeur de U était-elle suffisamment élevée? On rappelle que les HFW premiers et les HFW derniers bâtons de H_d sont ôtés dans SH_d. On examine les valeurs de H_d dans ces bâtons. Si (H_d > max_H_value * BELOW_WARN), l'incertitude d'alignement temporel est trop faible. Il faut refaire tourner l'algorithme avec une valeur de U plus grande. Les valeurs de d à examiner sont ($-U \le d < -U + \text{HFW}$) et ($U - \text{HFW} < d \le U$).

– Est-ce que SH_d a un retard bien défini? On examine SH_d, sauf pour les décalages situés à moins de DELTA de max_SH_offset. Si (SH_d > max_SH_value * BELOW_WARN) pour tout retard vidéo d tel que ($-U \le d <$ max_SH_offset – DELTA) ou (max_SH_offset + DELTA < $d \le U$), l'alignement temporel est ambigu.

Si la réponse aux deux vérifications ci-dessus est positive, on choisit le retard vidéo donné par max_SH_offset comme meilleur alignement temporel moyen pour la scène.

6.4.1.5 Observations et conclusions

L'algorithme de mesure du retard vidéo fondé sur les images utilise des séquences vidéo d'origine et traitée sous-échantillonnées. Il permet d'aligner des séquences vidéo dans un environnement hors service entièrement automatisé, avant qu'il ne soit procédé aux mesures de la qualité vidéo. Cet algorithme évalue l'alignement temporel pour chaque image, établit des histogrammes avec les différentes évaluations puis utilise le retard le plus couramment indiqué comme retard vidéo global – ou alignement temporel – pour la séquence d'images vidéo considérée.

Le retard indiqué à la dernière étape de l'algorithme (Etape 11 du § 6.4.1.4) peut être différent du retard qu'un observateur choisirait s'il alignait les scènes visuellement. Les observateurs ont tendance à se concentrer sur le mouvement, alignant les parties de la scène présentant beaucoup de mouvement, alors que l'algorithme fondé sur les images détermine le retard le plus fréquemment observé parmi toutes les images examinées. Les histogrammes globaux de retard peuvent servir à déterminer l'amplitude et des statistiques de tout retard vidéo variable dû au circuit fictif de référence.

6.4.2 Application de la correction d'alignement temporel

Pour toutes les caractéristiques de qualité, il faut que le décalage temporel calculé ici soit supprimé. Pour les décalages positifs, on supprime des images au début du fichier traité et à la fin du fichier d'origine. Pour les décalages négatifs, on supprime des images à la fin du fichier traité et au début du fichier d'origine. En cas de resynchronisation de trame de séquences vidéo à balayage à entrelacement, la séquence traitée est soumise à une resynchronisation de trame. Il convient donc de supprimer une trame au début et à la fin de la séquence vidéo traitée en plus des suppressions susmentionnées. Il faut simultanément supprimer une image au début du fichier vidéo d'origine (pour un retard de trame global de –1) ou à la fin de ce fichier (pour un retard de trame global de +1).

La correction de l'alignement temporel a pour effet de réduire le nombre d'images disponibles dans la séquence vidéo. Dans un souci de simplicité, tous les calculs ultérieurs sont fondés sur le nombre d'images vidéo disponibles une fois que toutes les corrections liées à l'étalonnage ont été appliquées.

7 Caractéristiques de qualité

7.1 Introduction

Une caractéristique de qualité est définie comme étant une grandeur associée à - ou extraite d' - une sous-région spatio-temporelle d'un flux vidéo (d'origine ou traité). Les flux de caractéristiques qui sont produits sont fonction de l'espace et du temps. En comparant les caractéristiques extraites d'une séquence vidéo traitée étalonnée avec les caractéristiques extraites de la séquence vidéo d'origine étalonnée, on peut calculer un ensemble de paramètres de qualité (§ 8) qui donnent une indication des modifications perçues de la qualité vidéo. Le présent paragraphe décrit un ensemble de caractéristiques de qualité qui caractérisent les modifications perçues des propriétés spatiales, temporelles et de chrominance des flux vidéo. Un filtre de perception est généralement appliqué au flux vidéo afin d'accentuer certaines propriétés de la qualité vidéo perçue, comme les informations de contour. Une fois ce filtrage opéré, les caractéristiques sont extraites des sous-régions spatio-temporelles (S-T) au moyen d'une fonction mathématique (par exemple un écart type). Enfin, un seuil de perceptibilité est appliqué aux caractéristiques extraites.

Dans ce qui suit, un flux de caractéristiques d'une séquence d'origine sera désigné par $f_o(s, t)$ et le flux de caractéristiques de la séquence traitée correspondante sera désigné par $f_p(s, t)$, où s et t sont des indices qui désignent respectivement la position spatiale et la position temporelle de la région S-T dans les flux vidéo d'origine et traité étalonnés. Pour nommer les caractéristiques, qui sont décrites dans les paragraphes qui suivent, on utilise des caractères en indice, ceux-ci étant choisis de manière à indiquer ce que la caractéristique mesure. Toutes les caractéristiques

concernent des images d'une séquence vidéo étalonnée (voir le § 6), les questions relatives à l'entrelacement étant abordées au moment de l'étalonnage. Toutes les caractéristiques sont indépendantes de la taille d'image (autrement dit la taille de la région S-T ne varie pas quand la taille d'image varie)[8].

En résumé, les étapes du calcul des caractéristiques sont les suivantes. Pour certaines caractéristiques, les étapes marquées comme étant une [option] ne seront peut-être pas nécessaires.

Etape 1: [option] Appliquer un filtre de perception.

Etape 2: Subdiviser le flux vidéo en régions S-T.

Etape 3: Extraire les caractéristiques, ou les statistiques récapitulatives, de chaque région S-T (par exemple la moyenne, l'écart type).

Etape 4: [option] Appliquer un seuil de perceptibilité.

Pour certaines caractéristiques, on peut utiliser deux filtres de perception différents ou plus.

7.1.1 Régions S-T

En général, les caractéristiques sont extraites des régions S-T localisées après application d'un ou de plusieurs filtres de perception aux flux vidéo d'origine et traité. Les positions des régions S-T sont telles que les flux vidéo sont subdivisés en régions S-T contiguës. Comme le flux vidéo traité a été étalonné, pour chaque région S-T de ce flux, il existe une région S-T du flux d'origine ayant la même position spatiale et la même position temporelle dans le flux vidéo. Pour extraire les caractéristiques de chaque région S-T, on calcule des statistiques récapitulatives ou on applique une certaine autre fonction mathématique sur la région d'intérêt S-T.

Chaque région S-T correspond à un bloc de pixels. La taille d'une région S-T est décrite par:

– le nombre de pixels horizontalement;
– le nombre de lignes d'image verticalement; et
– la dimension temporelle de la région, donnée en nombre équivalent d'images vidéo d'un système vidéo à 30 fps[9].

La Fig. 27 illustre une région S-T de 8 pixels horizontaux × 8 lignes verticales × 6 images vidéo NTSC, pour un total de 384 pixels. Dans le cas d'un système vidéo à 25 fps (PAL), cette même région S-T couvre 8 pixels horizontaux × 8 lignes verticales × 5 images vidéo, pour un total de 320 pixels.

Un cinquième de seconde est une dimension temporelle souhaitable, en raison de la facilité de conversion entre les fréquences d'images (un cinquième de seconde donne un nombre entier d'images vidéo pour les systèmes vidéo fonctionnant à 10, 15, 25 et 30 fps). La règle générale à appliquer pour la conversion entre fréquences d'images consiste à prendre la dimension de la région

---

[8] On suppose implicitement que le rapport entre la distance de visualisation et la hauteur d'image reste fixe (on utilise des distances de visualisation plus courtes lorsque les images sont plus petites). On trouvera au § 9 davantage d'observations sur la distance de visualisation supposée.

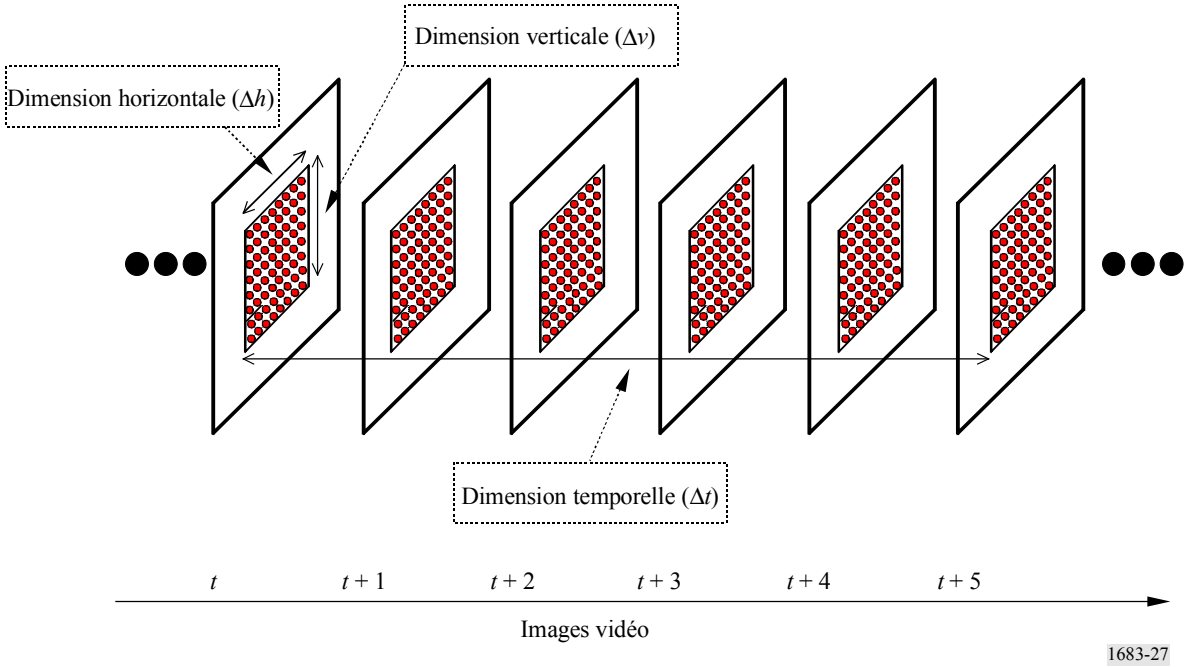
[9] Dans la présente Annexe, toutes les dimensions temporelles seront données en nombre équivalent d'images vidéo d'un système vidéo à 30 fps. Ainsi, une dimension temporelle de 6 images (*F*) représente à la fois 6 images d'un système NTSC (6/30) et 5 images d'un système PAL (5/25). Par ailleurs, on utilise 30 fps et 29,97 fps de manière interchangeable dans la présente Annexe, étant donné que cette légère différence de la fréquence d'images n'a pas d'incidence sur le calcul de la qualité VQM.

S-T en nombre d'images vidéo d'un système à 30 fps, à diviser par 30 puis à multiplier par la fréquence d'images du système vidéo testé. Les régions S-T qui contiennent une seule image vidéo sont supposées toujours contenir une seule image vidéo, indépendamment de la fréquence d'images.

La région d'intérêt spatial (SROI, voir § 3) englobant toutes les régions S-T est identique pour la séquence vidéo d'origine et la séquence vidéo traitée étalonnées. La SROI doit être entièrement comprise dans la PVR, éventuellement avec un tampon de pixels, comme cela est requis par les filtres de perception convolutifs. La dimension horizontale de la SROI doit être divisible par la dimension horizontale de la région S-T. De même, la dimension verticale de la région SROI doit être divisible par la dimension verticale de la région S-T. Un utilisateur peut ensuite contraindre la SROI à englober une région d'intérêt particulière, par exemple le centre de l'image vidéo.

FIGURE 27

Exemple de taille de région S-T pour l'extraction des caractéristiques



Temporellement, la séquence vidéo d'origine et la séquence vidéo traitée étalonnées sont subdivisées en un nombre identique de régions S-T, commençant à la première image vidéo alignée temporellement. Si le nombre d'images valables disponibles n'est pas divisible par la dimension temporelle de la région S-T, on ne tient pas compte des images se trouvant à la fin du clip.

Pour certaines caractéristiques, par exemple celles qui sont présentées au § 7.2, le bloc $8 \times 8\_6F$ permet d'obtenir une très bonne corrélation avec les évaluations subjectives. Il est toutefois à noter que la corrélation décroît lentement à mesure qu'on s'éloigne de la taille de région S-T optimale. Des dimensions horizontales et verticales allant jusqu'à 32 voire davantage et des dimensions temporelles allant jusqu'à 30 images donnent des résultats satisfaisants, ce qui laisse une grande liberté au concepteur du système de mesures objectives pour adapter les caractéristiques au volume de stockage ou à la largeur de bande de transmission disponible [Wolf et Pinson, 2001].

Une fois que le flux vidéo a été subdivisé en régions S-T, l'axe temporel de la caractéristique (t) ne correspond plus à des images individuelles. En revanche, il contient un nombre d'échantillons égal au nombre d'images valables de la séquence vidéo étalonnée divisé par la dimension temporelle de la région S-T.

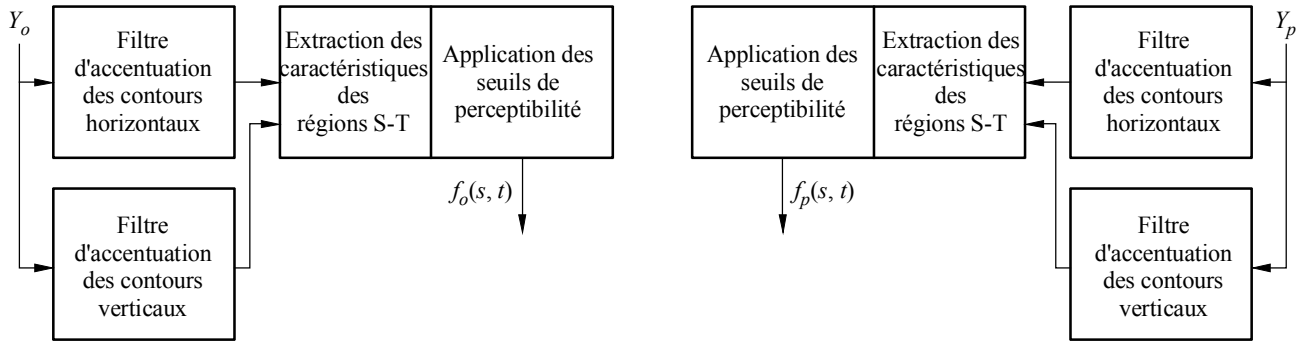
Lorsqu'on calcule simultanément deux caractéristiques ou plus, d'autres considérations deviennent importantes. Idéalement, toutes les caractéristiques devraient être calculées pour la même SROI.

7.2 Caractéristiques fondées sur les gradients spatiaux

Les caractéristiques déduites des gradients spatiaux peuvent servir à caractériser les distorsions perçues au niveau des contours. Par exemple, une perte générale d'informations relatives aux contours résulte d'un flou tandis qu'un excès d'informations relatives aux contours horizontaux et verticaux peut être lié à une distorsion due à une subdivision en blocs ou à un pavage. Les composantes *Y* du flux vidéo d'origine et du flux vidéo traité sont filtrées au moyen d'un filtre d'accentuation des contours horizontaux et d'un filtre d'accentuation des contours verticaux. Ces flux vidéo filtrés sont ensuite subdivisés en régions S-T à partir desquelles on extrait les caractéristiques, ou les statistiques récapitulatives, permettant de quantifier l'activité spatiale en fonction de l'angle d'orientation. Ces caractéristiques sont ensuite coupées à l'extrémité inférieure afin d'émuler les seuils de perceptibilité. Les filtres d'accentuation des contours, la taille de la région S-T et les seuils de perceptibilité ont été choisis sur la base de flux vidéo conformes à la Recommandation UIT-R BT.601 qui ont été évalués subjectivement à une distance de visualisation de six hauteurs d'image. La Fig. 28 présente un aperçu de l'algorithme utilisé pour extraire les caractéristiques fondées sur les gradients spatiaux.

FIGURE 28

Aperçu de l'algorithme utilisé pour extraire les caratéristiques fondées sur les gradients spatiaux

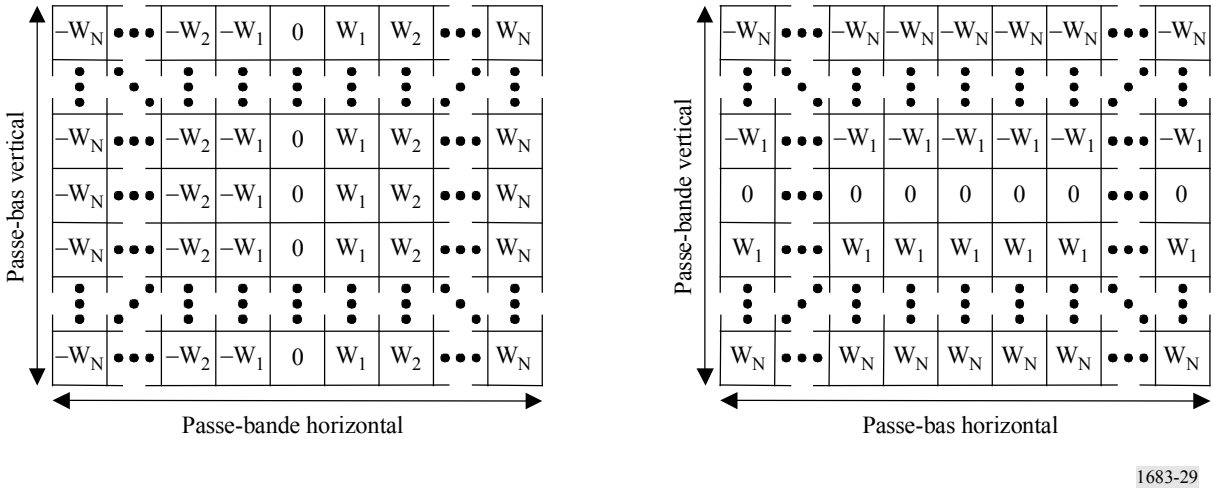


7.2.1 Filtres d'accentuation des contours

Les images *Y* (de luminance) du flux vidéo d'origine et du flux vidéo traité sont d'abord traitées par des filtres d'accentuation des contours horizontaux et verticaux qui accentuent les contours tout en réduisant le bruit. Les deux filtres présentés sur la Fig. 29 sont appliqués séparément. Le premier (filtre de gauche) accentue les différences entre pixels horizontaux tout en procédant à un lissage vertical et le second (filtre de droite) accentue les différences entre pixels verticaux tout en procédant à un lissage horizontal.

FIGURE 29

Filtres d'accentuation des contours



Les deux filtres sont transposés l'un de l'autre, ont une taille de 13 × 13, et ont pour coefficients de pondération:

$$w_x = k \cdot \left(\frac{x}{c}\right) \cdot \exp\left\{-\left(\frac{1}{2}\right)\left(\frac{x}{c}\right)^2\right\}$$

où:

x: déplacement en pixels par rapport au centre du filtre (0, 1, 2, …, N)

c: constante fixant la largeur du filtre passe-bande et

k: constante de normalisation choisie de manière que chaque filtre présente le même gain qu'un véritable filtre de Sobel [Jain, 1989].

L'expérience a montré que le filtrage optimal en passe-bande horizontal pour une distance de visualisation égale à six hauteurs d'image était réalisé pour un filtre avec c = 2 présentant une réponse crête d'environ 4,5 cycles/degré. Les coefficients de pondération utilisés pour le filtre passe-bande sont les suivants:

[–0,0052625; –0,0173446; –0,0427401; –0,0768961; –0,0957739; –0,0696751; 0; 0,0696751; 0,0957739; 0,0768961; 0,0427401; 0,0173446; 0,0052625]

Il est à noter que les filtres de la Fig. 29 présentent une réponse passe-bas uniforme. Cette réponse a généré la meilleure évaluation de qualité et présente de plus l'avantage d'être efficace sur le plan des calculs (dans le cas du filtre de gauche de la Fig. 29 par exemple, il suffit de sommer les pixels d'une colonne et de multiplier le résultat par le coefficient de pondération).

7.2.2 Description des caractéristiques f_{SI13} et f_{HV13}

Le présent paragraphe décrit l'extraction de deux caractéristiques d'activité spatiale des régions S-T de flux vidéo d'origine et traité aux contours accentués comme décrits au § 7.2.1. Ces caractéristiques servent pour la détection de dégradations spatiales telles que le flou et la subdivision en blocs. Le filtre présenté sur la Fig. 29 (à gauche) accentue les gradients spatiaux suivant la direction horizontale, H, alors que le transposé de ce filtre (à droite) accentue les gradients spatiaux suivant la direction verticale, V. On peut tracer pour chaque pixel la réponse de ces filtres H et V sur un diagramme à deux dimensions comme celui de la Fig. 30: la réponse du

filtre *H* correspond à l'abscisse et la réponse du filtre *V* correspond à l'ordonnée. Pour un pixel donné de l'image repéré par sa ligne *i*, sa colonne *j* et le temps *t*, les réponses des filtres *H* et *V* seront respectivement notées $H(i, j, t)$ et $V(i, j, t)$. On peut convertir ces réponses en coordonnées polaires (R, θ) en utilisant les relations suivantes:

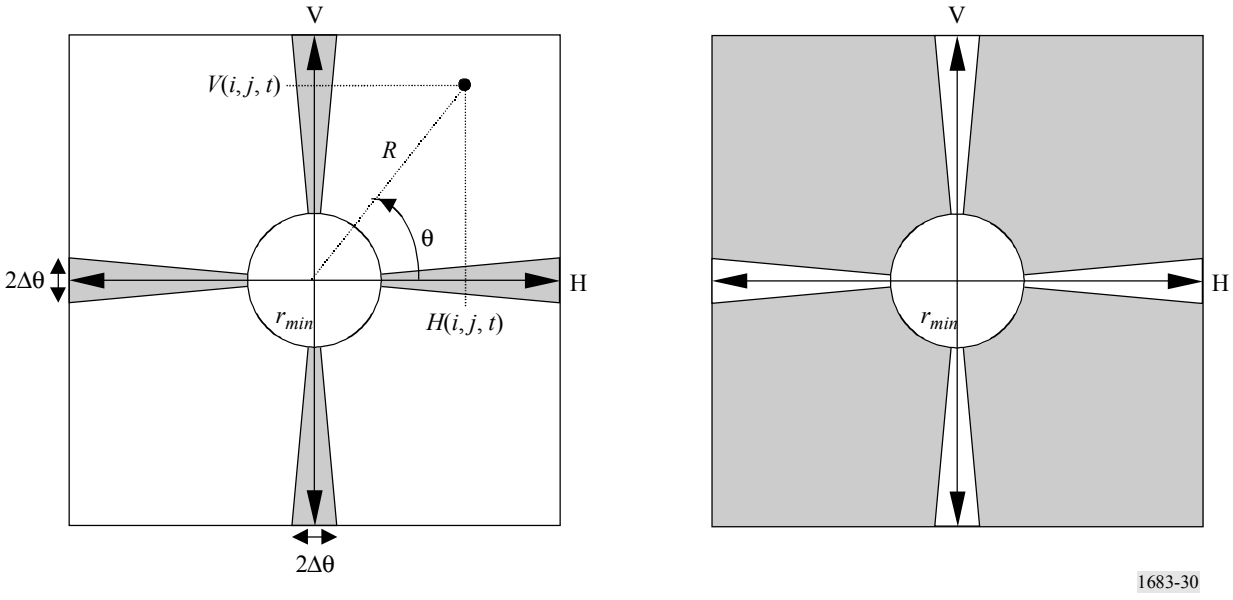
$$R(i,j,t) = \sqrt{H(i,j,t)^2 + V(i,j,t)^2}$$

et (102)

$$\theta(i,j,t) = tg^{-1}\left[\frac{V(i,j,t)}{H(i,j,t)}\right]$$

FIGURE 30

Subdivision de l'activité spatiale horizontale (H) et verticale (V) en distribution $\overline{HV}$ (à gauche) et HV (droite)



La première caractéristique, qui est une mesure de l'information SI globale, est désignée par f_{SI13} car les images ont été préalablement traitées par les filtres 13×13 présentés sur la Fig. 29. Cette caractéristique se calcule simplement comme l'écart type (std, *standard deviation*) sur la région S-T des échantillons $R(i, j, t)$. Elle est ensuite coupée au seuil de perceptibilité *P* (ce qui signifie que f_{SI13} est fixé à *P* si le calcul de std donne un résultat inférieur à *P*). On obtient ainsi:

$$f_{SI13} = \left\{\text{std}\left[R(i,j,t)\right]\right\}\Big|_P / i,j,t \in \left\{\text{Région S-T}\right\} \quad (103)$$

Cette caractéristique est sensible aux modifications affectant la quantité globale d'activité spatiale au sein d'une région S-T donnée. Par exemple, un flou localisé entraîne une diminution de la quantité d'activité spatiale alors qu'un bruit accroît cette dernière. Le seuil *P* recommandé pour cette caractéristique est de 12.

La seconde caractéristique, f_{HV13}, est sensible aux modifications de distribution angulaire (ou d'orientation) de l'activité spatiale. On calcule les images complémentaires avec les distributions de gradients spatiaux représentées en ombragé sur la Fig. 30. L'image avec les gradients horizontaux et verticaux, notée *HV*, contient les pixels $R(i, j, t)$ correspondant à des contours horizontaux ou

verticaux (les pixels correspondant à des contours diagonaux sont mis à zéro). L'image avec les gradients diagonaux, notée $\overline{HV}$, contient les pixels $R(i, j, t)$ correspondant à des contours diagonaux (les pixels correspondant à des contours horizontaux ou verticaux sont mis à zéro). Les amplitudes de gradient $R(i, j, t)$ inférieures à r_{min} sont mises à zéro dans les deux images pour garantir des calculs exacts de θ. On peut représenter mathématiquement les pixels de HV et $\overline{HV}$ de la façon suivante:

$$HV(i,j,t)=\begin{cases} R(i,j,t) & \text{si } R(i,j,t)\geq r_{min} \text{ et } m\frac{\pi}{2}-\Delta\theta<\theta(i,j,t)<m\frac{\pi}{2}+\Delta\theta \quad (m=0,1,2,3) \\ 0 & \text{sinon} \end{cases} \tag{104}$$

et:

$$\overline{HV}(i,j,t)=\begin{cases} R(i,j,t) & \text{si } R(i,j,t)\geq r_{min} \text{ et } m\frac{\pi}{2}+\Delta\theta\leq\theta(i,j,t)\leq(m+1)\frac{\pi}{2}-\Delta\theta \quad (m=0,1,2,3) \\ 0 & \text{sinon} \end{cases} \tag{105}$$

avec:

$$i,j,t\in\{\text{région S-T}\}$$

Pour les calculs de HV et $\overline{HV}$ ci-dessus, il est recommandé d'utiliser une valeur de 20 pour r_{min} et une valeur de 0,225 radians pour Δθ. La caractéristique f_{HV13} pour une région S-T donnée est ensuite obtenue comme étant le rapport entre la moyenne de HV et la moyenne de $\overline{HV}$, ces moyennes étant coupées à leur seuil de perceptibilité P. On obtient ainsi:

$$f_{HV13}=\frac{\{\text{mean}[HV(i,j,t)]\}|_P}{\{\text{mean}[\overline{HV}(i,j,t)]\}|_P} \tag{106}$$

Le seuil de perceptibilité P recommandé pour les moyennes de HV et $\overline{HV}$ est de 3. La caractéristique f_{HV13} est sensible aux modifications de distribution angulaire de l'activité spatiale au sein d'une région S-T donnée. Par exemple, si les contours horizontaux et verticaux sont plus flous que les contours diagonaux, la valeur de f_{HV13} du flux vidéo traité sera moins élevée que celle du flux vidéo d'origine. D'autre part, si des contours horizontaux ou verticaux erronés sont introduits (par exemple sous forme de distorsions liée à une subdivision en blocs ou à un pavage), la valeur de f_{HV13} du flux vidéo traité sera alors plus élevée que celle du flux vidéo d'origine. La caractéristique f_{HV13} fournit aussi un moyen simple pour tenir compte des variations de sensibilité du système visuel humain en fonction de l'angle d'orientation[10].

---

[10] Cet exposé de la caractéristique f_{HV13}, quoique globalement valable, est quelque peu simplifié. Par exemple, lorsqu'il rencontre certaines formes, le filtre f_{HV13} se comporte d'une manière qui peut être contraire à l'intuition (par exemple un coin formé par une ligne horizontale et une ligne verticale conduira à une énergie diagonale).

7.3 Caractéristiques fondées sur les informations de chrominance

Dans le présent paragraphe, on décrit une seule caractéristique qui peut être utilisée pour mesurer les distorsions des signaux de chrominance (C_B, C_R). Pour un pixel donné de l'image repéré par sa ligne i, sa colonne j et le temps t, désignons par C_B (i, j, t) et C_R (i, j, t) les valeurs des composantes C_B et C_R définies dans la Recommandation UIT-R BT.601[11]. Les composantes d'un vecteur de caractéristique de chrominance à deux dimensions, f_{COHER_COLOR}, sont calculées comme étant la moyenne, mean sur la région S-T des échantillons C_B (i, j, t) et C_R (i, j, t), respectivement, un poids de perception plus élevé étant affecté à la composante C_R:

$$\underline{f}_{COHER\_COLOR} = \left(\text{mean}\left[C_B(i,j,t)\right], W_R * \text{mean}\left[C_R(i,j,t)\right]\right) / i,j,t \in \{\text{région S-T}\}, \text{ et } W_R = 1{,}5 \quad (107)$$

L'équation (107) permet de procéder à une intégration cohérente (d'où le nom f_{COHER_COLOR}) car la relation de phase entre C_B et C_R est préservée. Pour ceux qui connaissent bien les vecteurscopes, l'utilité du vecteur de caractéristique de chrominance apparaît directement à l'examen des signaux de mire chromatique. Pour les scènes générales, on peut visualiser l'utilité du vecteur de caractéristique de chrominance concernant la mesure des distorsions de la chrominance pour des blocs vidéo qui couvrent une certaine plage spatio-temporelle. Toutefois, si la taille de la région S-T est trop grande, de nombreuses couleurs risquent d'être incluses dans le calcul et l'utilité de f_{COHER_COLOR} est alors réduite. Une taille de région S-T de 8 pixels horizontaux $\times$ 8 lignes verticales $\times$ (1 à 3) images vidéo permet de générer un vecteur de caractéristique de chrominance robuste (en fait 4 pixels C_B et C_R horizontaux, étant donné que ces signaux sont sous-échantillonnés par un facteur deux horizontalement pour ce qui est de l'échantillonnage selon la Recommandation UIT-R BT.601).

7.4 Caractéristiques fondées sur les informations de contraste

Les caractéristiques qui mesurent les informations de contraste localisé sont sensibles aux dégradations de la qualité telles que le flou (perte de contraste) et l'ajout de bruit (gain de contraste). On peut calculer facilement une caractéristique de contraste localisé, f_{CONT}, pour chaque région S-T à partir de l'image de luminance Y comme suit:

$$f_{CONT} = \left\{\text{std}\left[Y(i,j,t)\right]\right\}\Big|_P / i,j,t \in \{\text{région S-T}\} \quad (108)$$

Le seuil de perceptibilité P recommandé pour la caractéristique f_{CONT} est compris entre quatre et six.

7.5 Caractéristiques fondées sur l'ATI

Les caractéristiques qui mesurent les distorsions du flux de mouvement sont sensibles aux dégradations de la qualité telles que l'élimination ou la répétition d'images (perte de mouvement) et l'ajout de bruit (gain de mouvement). On calcule une caractéristique d'ATI, f_{ATI}, pour chaque région S-T. Pour cela, on commence par générer un flux de mouvement fondé sur la valeur absolue de la différence entre deux images vidéo consécutives aux instants t et $t-1$ puis on calcule l'écart type sur la région S-T. Mathématiquement, ce processus est représenté de la façon suivante:

$$f_{ATI} = \left\{\text{std}\left|Y(i,j,t) - Y(i,j,t)\right|\right\}\Big|_P / i,j,t \in \{\text{région S-T}\} \quad (109)$$

Le seuil de perceptibilité P recommandé pour la caractéristique f_{ATI} est compris entre un et trois.

---

[11] Les corrections de gain et de décalage ne sont pas appliquées aux plans d'image C_B et C_R. Voir le § 6.3.3.

L'utilisation d'une image précédente introduit des considérations qui vont au-delà de celles qui sont associées aux autres caractéristiques. Lors du calcul de f_{ATI} conjointement avec une autre caractéristique (par exemple $f_{CONTRAST\_ATI}$ définie au § 7.6) ou lors de son utilisation dans un modèle (voir le § 9), l'image supplémentaire requise a pour effet de compliquer la tâche de positionnement des régions S-T (voir le § 7.1.1).

7.6 Caractéristiques fondées sur le produit croisé du contraste et de l'ATI

La quantité de mouvement présente peut avoir une incidence sur la perceptibilité des dégradations spatiales. De même, la quantité de détails spatiaux présents peut avoir une incidence sur la perceptibilité des dégradations temporelles. Une caractéristique déduite du produit croisé de l'information de contraste et de l'information temporelle absolue permet de tenir compte partiellement de ces interactions. Cette caractéristique, désignée par $f_{CONTRAST\_ATI}$, est calculée comme étant le produit des caractéristiques définies aux § 7.4 et 7.5 [12]. Le seuil de perceptibilité recommandé $P = 3$ est appliqué séparément à chaque caractéristique (f_{CONT} et f_{ATI}) avant que leur produit croisé ne soit calculé. Les dégradations seront davantage visibles dans les régions S-T présentant un produit croisé faible que dans les régions S-T présentant un produit croisé élevé. Cela est particulièrement vrai pour les dégradations de type bruit et blocs d'erreurs.

L'image supplémentaire requise pour f_{ATI} a pour effet de compliquer légèrement $f_{CONTRAST\_ATI}$, car les régions S-T utilisées par les deux caractéristiques f_{CONT} et f_{ATI} doivent être positionnées de la même façon. Soit on n'utilise pas la première image de la séquence vidéo pour f_{ATI}, soit les régions S-T situées au début de la séquence vidéo contiennent une image de moins (par exemple, si on considère une dimension temporelle de 6 images, la première région S-T pour f_{ATI} utiliserait 5 images au lieu de 6). Pour les paramètres et modèles spécifiés ici, on suppose que c'est la seconde solution qui est utilisée.

8 Paramètres de qualité

8.1 Introduction

Les paramètres de qualité, qui servent à mesurer les distorsions de qualité vidéo dues aux gains et aux pertes associés aux valeurs de caractéristiques, sont d'abord calculés pour chaque région S-T. Pour cela, on compare les valeurs des caractéristiques du flux d'origine, $f_o(s, t)$, avec les valeurs des caractéristiques du flux traité correspondant, $f_p(s, t)$ (§ 8.2). On utilise plusieurs relations fonctionnelles pour émuler le masquage visuel des dégradations pour chaque région S-T. Des fonctions de regroupement des erreurs dans l'espace et dans le temps émulent ensuite la façon dont l'être humain déduit les évaluations de qualité subjective. Le regroupement d'erreurs dans l'espace est appelé regroupement spatial (§ 8.3) et le regroupement d'erreurs dans le temps est appelé regroupement temporel (§ 8.4). L'application séquentielle des fonctions de regroupement spatial et de regroupement temporel au flux de paramètres de qualité S-T génère des paramètres de qualité pour le clip vidéo tout entier, dont la durée nominale est comprise entre 5 et 10 s. La valeur finale de chaque paramètre après regroupement temporel peut être corrigée puis coupée (§ 8.5) et ce, afin de tenir compte des relations non linéaires entre la valeur du paramètre et la qualité perçue et de réduire encore la sensibilité du paramètre.

---

[12] On utilise un produit croisé standard des caractéristiques f_{CONT} et f_{ATI} (à savoir $f_{CONT} * f_{ATI}$) pour les caractéristiques du flux traité $f_p(s, t)$ et du flux d'origine $f_o(s, t)$ dans les fonctions de comparaison ratio_loss et ratio_gain décrites au § 8.2.1. Toutefois, pour les fonctions de comparaison log_loss et log_gain, les caractéristiques du flux traité et du flux d'origine sont calculées de la manière suivante: $\log_{10}[f_{CONT}] * \log_{10}[f_{ATI}]$, et les fonctions de comparaison utilisent une différence (à savoir $f_p(s, t) - f_o(s, t)$ plutôt que $\log_{10}[f_p(s, t) / f_o(s, t)]$).

En résumé, les étapes du calcul des paramètres sont les suivantes. Dans certains cas, l'étape indiquée comme une [option] ne sera pas nécessaire.

Etape 1: Comparer les valeurs des caractéristiques du flux d'origine avec les valeurs des caractéristiques du flux traité.

Etape 2: Procéder au regroupement spatial.

Etape 3: Procéder au regroupement temporel.

Etape 4: [option] Appliquer une correction non linéaire et/ou procéder à une coupure.

Tous les paramètres sont conçus pour prendre uniquement des valeurs positives ou uniquement des valeurs négatives. Une valeur de paramètre de zéro indique qu'il n'y a pas de dégradation.

8.2 Fonctions de comparaison

La dégradation perçue au niveau de chaque région S-T est calculée au moyen de fonctions qui modélisent le masquage visuel des dégradations spatiales et temporelles. Le présent paragraphe expose les fonctions de masquage qui sont utilisées par les divers paramètres pour produire des paramètres de qualité qui sont fonction de l'espace et du temps.

8.2.1 Fonction de rapport et fonction de logarithme

La perte et le gain sont généralement examinés séparément, car ils produisent des effets fondamentalement différents sur la perception de la qualité (par exemple perte d'activité spatiale due au flou et gain d'activité spatiale dû au bruit ou à la subdivision en blocs). Parmi les nombreuses fonctions de comparaison qui ont été évaluées, deux formes ont produit de manière cohérente une très bonne corrélation avec les évaluations subjectives. Chacune de ces formes peut être utilisée avec les calculs de gain ou de perte pour un total de quatre fonctions de comparaison S-T de base. Les quatre formes primaires sont les suivantes:

$$\text{ratio\_loss}(s,t) = np\left\{\frac{f_p(s,t) - f_o(s,t)}{f_o(s,t)}\right\}$$

$$\text{ratio\_gain}(s,t) = pp\left\{\frac{f_p(s,t) - f_o(s,t)}{f_o(s,t)}\right\}$$

$$\text{log\_loss}(s,t) = np\left\{\log_{10}\left[\frac{f_p(s,t)}{f_o(s,t)}\right]\right\}$$

$$\text{log\_gain}(s,t) = pp\left\{\log_{10}\left[\frac{f_p(s,t)}{f_o(s,t)}\right]\right\}$$

où:

pp: opérateur de partie positive (autrement dit, les valeurs négatives sont remplacées par des zéros)

np: opérateur de partie négative (autrement dit, les valeurs positives sont remplacées par des zéros).

Ces fonctions de masquage visuel impliquent que la perception des dégradations est inversement proportionnelle à la quantité d'activité spatiale ou temporelle localisée qui est présente. En d'autres termes, les dégradations spatiales deviennent moins visibles à mesure que l'activité spatiale augmente (masquage spatial) et les dégradations temporelles deviennent moins visibles à mesure que l'activité temporelle augmente (masquage temporel). Les fonctions de comparaison de type rapport et logarithme ont un comportement très similaire, mais la fonction de logarithme a tendance à être légèrement plus avantageuse pour les gains tandis que la fonction de rapport a tendance à être légèrement plus avantageuse pour les pertes. La fonction de logarithme a une plage dynamique plus grande, ce qui est utile lorsque les valeurs des caractéristiques du flux traité dépassent de beaucoup les valeurs des caractéristiques du flux d'origine.

8.2.2 Distance euclidienne

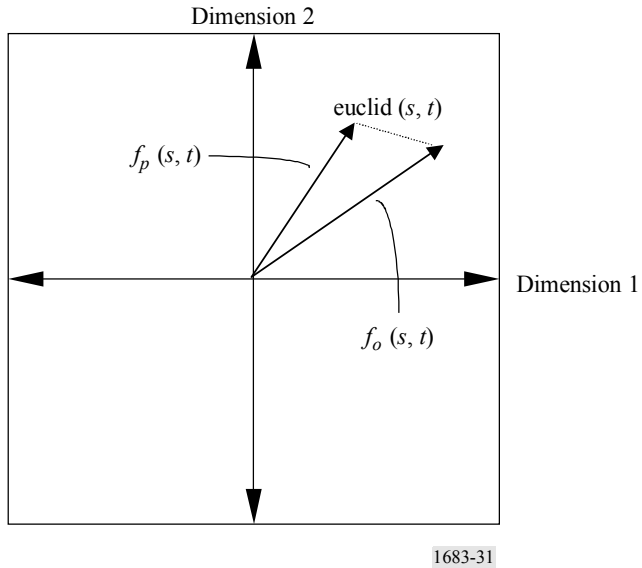
Une autre fonction de comparaison S-T utile est la simple distance euclidienne, représentée par la longueur du vecteur de différence entre le vecteur de caractéristique du flux d'origine $f_o(s, t)$ et le vecteur de caractéristique du flux traité correspondant, $f_p(s, t)$:

$$\text{euclid}(s,t) = \left\| \underline{f}_p(s,t) - \underline{f}_o(s,t) \right\| \tag{110}$$

La Fig. 31 donne une illustration de la distance euclidienne pour un vecteur de caractéristique à deux dimensions extrait d'une région S-T (par exemple le vecteur de caractéristique f_{COHER_COLOR} décrit au § 7.3), où les indices s et t désignent respectivement la position spatiale et la position temporelle de la région S-T dans le flux vidéo d'origine et le flux vidéo traité étalonnés. Le segment en pointillés sur la Fig. 31 illustre la distance euclidienne. La mesure de la distance euclidienne peut être généralisée pour des vecteurs de caractéristique qui ont un nombre arbitraire de dimensions.

FIGURE 31

Illustration de la distance euclidienne, euclid (*s*, *t*), pour un vecteur de caractéristique à deux dimensions



8.3 Fonctions de regroupement spatial

Les paramètres issus des régions S-T (§ 8.2) constituent des matrices à trois dimensions: une dimension temporelle et deux dimensions spatiales (position horizontale et position verticale de la région S-T). On regroupe ensuite les dégradations issues des régions S-T ayant le même indice temporel *t* au moyen d'une fonction de regroupement spatial. Le regroupement spatial génère un historique temporel des valeurs de paramètre. Cet historique, désigné génériquement par *p*(*t*), doit ensuite être regroupé temporellement au moyen d'une fonction de regroupement temporel donnée au § 8.4. Le Tableau 12 présente une récapitulation des fonctions de regroupement spatial les plus couramment utilisées.

Des examens approfondis ont montré que les fonctions de regroupement spatial optimales incluent généralement un certain traitement associé au cas le plus défavorable, comme le calcul de la moyenne des 5% de distorsions les plus mauvaises observées sur l'indice spatial *s* ([Wolf et Pinson, 1998, 1999, 2001 et 2002]). Cela s'explique par le fait que les dégradations localisées ont tendance à attirer l'attention de l'observateur, pour lequel la plus mauvaise partie de l'image constitue le facteur prédominant dans l'évaluation de la qualité subjective. Par exemple, la fonction de regroupement spatial «above95%» est calculée pour chaque indice temporel *t* pour la fonction log_gain(*s*, *t*) décrite au § 8.2.1 comme étant la moyenne des 5% de valeurs positives les plus élevées sur l'indice spatial *s*[13]. Cela revient à trier les distorsions du gain de la valeur la plus faible à la valeur la plus élevée pour chaque indice temporel *t* et à calculer la moyenne des distorsions qui sont au-dessus du seuil de 95% (car plus les valeurs positives sont élevées, plus la distorsion est grande). De même, les distorsions dues à des pertes comme celles qui sont calculées par la fonction ratio_loss(*s*, *t*) décrite au § 8.2.1 sont triées pour chaque indice temporel *t*, mais on calcule la moyenne des distorsions qui sont au-dessous du seuil de 5% (below5%) (car les pertes sont négatives).

8.4 Fonctions de regroupement temporel

L'historique temporel du paramètre *p*(*t*) résultant de la fonction de regroupement spatial (voir le § 8.3) fait ensuite l'objet d'un regroupement au moyen d'une fonction de regroupement temporel et ce, afin de produire un paramètre objectif *p* pour le clip vidéo, dont la durée nominale est comprise entre 4 et 10 s. Les observateurs semblent utiliser plusieurs fonctions de regroupement temporel lorsqu'ils évaluent subjectivement des clips vidéo d'une durée approximative de 10 s. La moyenne, mean dans le temps donne une indication de la qualité moyenne qui est observée pendant la période considérée. Les niveaux à 90% et à 10% dans le temps donnent une indication de la qualité transitoire la plus mauvaise qui est observée respectivement pour les gains et pour les pertes (des erreurs de transmission numérique peuvent par exemple causer une perturbation de 1 à 2 s dans le flux vidéo traité). Après regroupement temporel, un paramètre *p* donné prend uniquement des valeurs négatives ou uniquement des valeurs positives. Le Tableau 13 présente une récapitulation des fonctions de regroupement temporel les plus couramment utilisées.

---

[13] Il est à noter que l'indice temporel *t* ne correspond pas ici à des images individuelles (voir le § 7.1.1). En effet, chaque valeur de *t* correspond aux régions S-T ayant la même dimension temporelle.

TABLEAU 12

Fonctions de regroupement spatial et leur définition

Fonction de regroupement spatial	Définition
below5%	Pour chaque indice temporel *t*, on trie les valeurs de paramètre de la valeur la plus faible à la valeur la plus élevée. On calcule la moyenne de toutes les valeurs de paramètre qui sont inférieures ou égales au seuil de 5%. Pour les paramètres de perte, cette fonction de regroupement spatial produit un paramètre qui donne une indication de la qualité la plus mauvaise dans l'espace
above95%	Pour chaque indice temporel *t*, on trie les valeurs de paramètre de la valeur la plus faible à la valeur la plus élevée. On calcule la moyenne de toutes les valeurs de paramètre qui sont supérieures ou égales au seuil de 95%. Pour les paramètres de gain, cette fonction de regroupement spatial produit un paramètre qui donne une indication de la qualité la plus mauvaise dans l'espace
mean	Pour chaque indice temporel *t*, on calcule la moyenne de toutes les valeurs de paramètre. Cette fonction de regroupement spatial produit un paramètre qui donne une indication de la qualité moyenne dans l'espace
std	Pour chaque indice temporel *t*, on calcule l'écart type de toutes les valeurs de paramètre. Cette fonction de regroupement spatial produit un paramètre qui donne une indication des variations de la qualité dans l'espace
below5%tail	Pour chaque indice temporel *t*, on trie les valeurs de paramètre de la valeur la plus faible à la valeur la plus élevée. On calcule la moyenne de toutes les valeurs de paramètre qui sont inférieures ou égales au seuil de 5% puis on soustrait le niveau à 5% de cette moyenne. Pour les paramètres de perte, cette fonction de regroupement spatial permet de mesurer l'étalement des plus mauvais niveaux de qualité dans l'espace. Elle est utile pour la mesure des effets des distorsions localisées spatialement sur la qualité perçue
above99%tail	Pour chaque indice temporel *t*, on trie les valeurs de paramètre de la valeur la plus faible à la valeur la plus élevée. On calcule la moyenne de toutes les valeurs de paramètre qui sont supérieures ou égales au seuil de 99% puis on soustrait le niveau à 99% de cette moyenne. Pour les paramètres de gain, cette fonction de regroupement spatial permet de mesurer l'étalement des plus mauvais niveaux de qualité dans l'espace. Elle est utile pour la mesure des effets des distorsions localisées spatialement sur la qualité perçue

TABLEAU 13

Fonctions de regroupement temporel et leur définition

Fonction de regroupement temporel	Définition
10%	On trie l'historique temporel des valeurs de paramètre de la valeur la plus faible à la valeur la plus élevée et on prend le seuil de 10%. Pour les paramètres de perte, cette fonction de regroupement temporel produit un paramètre qui donne une indication de la plus mauvaise qualité dans le temps. Pour les paramètres de gain, elle produit un paramètre qui donne une indication de la meilleure qualité dans le temps
25%	On trie l'historique temporel des valeurs de paramètre de la valeur la plus faible à la valeur la plus élevée et on prend le seuil de 25%
50%	On trie l'historique temporel des valeurs de paramètre de la valeur la plus faible à la valeur la plus élevée et on prend le seuil de 50%
90%	On trie l'historique temporel des valeurs de paramètre de la valeur la plus faible à la valeur la plus élevée et on prend le seuil de 90%. Pour les paramètres de perte, cette fonction de regroupement temporel produit un paramètre qui donne une indication de la meilleure qualité dans le temps. Pour les paramètres de gain, elle produit un paramètre qui donne une indication de la plus mauvaise qualité dans le temps
mean	On calcule la moyenne de l'historique temporel des valeurs de paramètre. Cette fonction produit un paramètre qui donne une indication de la qualité moyenne dans le temps
std	On calcule l'écart type de l'historique temporel des valeurs de paramètre. Cette fonction de regroupement temporel produit un paramètre qui donne une indication des variations de la qualité dans le temps
above90%tail	On trie l'historique temporel des valeurs de paramètre de la valeur la plus faible à la valeur la plus élevée et on calcule la moyenne de toutes les valeurs de paramètre qui sont supérieures ou égales au seuil de 90% puis on soustrait le niveau à 90% de cette moyenne. Pour les paramètres de gain, cette fonction de regroupement temporel permet de mesurer l'étalement des plus mauvais niveaux de qualité dans le temps. Elle est utile pour la mesure des effets des distorsions localisées temporellement sur la qualité perçue

8.5 Application d'une correction non linéaire et coupure

On peut appliquer un facteur de correction au paramètre *p* prenant uniquement des valeurs positives ou uniquement des valeurs négatives, issu du regroupement temporel (§ 8.4) afin de tenir compte des relations non linéaires entre la valeur du paramètre et la qualité perçue. Il est préférable de supprimer les éventuelles relations non linéaires avant d'établir les modèles de qualité vidéo (§ 9), car on utilise un algorithme linéaire fondé sur la méthode des moindres carrés pour déterminer les poids optimaux pour les paramètres. Les deux fonctions de correction non linéaire pouvant être appliquées sont la fonction racine carrée, désignée par sqrt, et la fonction carrée, désignée par square. Si la fonction sqrt est appliquée à un paramètre dont toutes les valeurs sont négatives, on commence par faire en sorte que ce paramètre ne prenne que des valeurs positives (on prend la valeur absolue).

Enfin, on peut appliquer une fonction de coupure désignée par clip_T, où T est le seuil de coupure, afin de réduire la sensibilité du paramètre aux faibles dégradations. La fonction de coupure remplace toute valeur du paramètre comprise entre le niveau de coupure et zéro par le niveau de coupure puis le niveau de coupure est soustrait de la valeur résultante du paramètre. La représentation mathématique en est la suivante:

$$\text{clip\_T}(p) = \begin{cases} \max(p,T) - T & \text{si } p \text{ ne prend que des valeurs positives} \\ \min(p,T) - T & \text{si } p \text{ ne prend que des valeurs négatives} \end{cases}$$

8.6 Convention pour la dénomination des paramètres

Le présent paragraphe résume la convention de dénomination technique utilisée pour les paramètres de qualité vidéo. Selon cette convention, on attribue à chaque paramètre un nom très long constitué de mots d'identification (sous-noms) séparés par des soulignés. Le nom de paramètre technique résume le processus exact utilisé pour calculer le paramètre. Chaque sous-nom identifie une fonction ou une étape du processus de calcul du paramètre. Les sous-noms sont énumérés dans l'ordre dans lequel les fonctions ou les étapes se déroulent, de gauche à droite. Le Tableau 14 récapitule les sous-noms utilisés pour créer un nom de paramètre technique, énumérés dans l'ordre susmentionné. Le § 8.6.1 donne quelques exemples de nom de paramètre technique et des sous-noms associés issus du Tableau 14.

TABLEAU 14

Convention utilisée pour la dénomination technique des paramètres de qualité vidéo

Sous-nom	Définition	Exemples
Couleur	Plans d'image de l'espace chromatique utilisés par le paramètre	*Y* pour le plan d'image de luminance. *colour* pour les plans d'image (C_B, C_R)
Propre à la caractéristique	Ce sous-nom décrit les calculs qui rendent unique le paramètre considéré. Tous les autres sous-noms qui suivent correspondent à des processus génériques qui peuvent être utilisés par de nombreux types différents de paramètres. Le sous-nom «propre à la caractéristique» est généralement le nom de la caractéristique qui est extraite du plan «couleur» à ce stade dans le flux, autrement dit à l'emplacement de ce sous-nom. Toutefois, des informations non prises en considération par la convention de dénomination peuvent aussi être incluses ici. Par exemple, le paramètre HV applique le sous-nom «statistique de bloc» séparément aux plans d'image *HV* et $\overline{HV}$. Le rapport entre *HV* et $\overline{HV}$ qui en découle est spécifié par le sous-nom «propre à la caractéristique» (plutôt que d'occuper un sous-nom distinct après le sous-nom «statistique de bloc»)	si13 pour la caractéristique f_{SI13} du § 7.2.2. hv13_angleX.XXX_*rminYY* pour la caractéristique f_{HV13} du § 7.2.2, où *X.XXX* est la valeur de $\Delta\theta$ et *YY* celle de r_{min}. coher_color pour la caractéristique f_{COHER_COLOR} du § 7.3. cont pour la caractéristique f_{CONT} du § 7.4. ati pour la caractéristique f_{ATI} du § 7.5. contrast_ati pour la caractéristique $f_{CONTRAST\_ATI}$ du § 7.6
Décalage de bloc	Présent en cas de superposition de blocs S-T (par exemple des blocs qui se chevauchent dans le temps). Lorsque ce sous-nom est absent, les blocs sont supposés être contigus dans le temps	sliding

TABLEAU 14

Convention utilisée pour la dénomination technique des paramètres de qualité vidéo

Sous-nom	Définition	Exemples
Image complète	Présent lorsque la taille de bloc S-T contient toute la région valable de l'image. Lorsque ce sous-nom est absent, le sous-nom «taille de bloc» doit être présent	image
Taille de bloc	Présent lorsque l'image est subdivisée en blocs S-T (voir le § 7.1.1). Dans un souci de cohérence, la taille de bloc est toujours indiquée en nombre de lignes d'image et en nombre de pixels d'image du plan de luminance (*Y*). Ainsi, pour les séquences vidéo échantillonnées selon le format 4:2:2, les blocs de couleur contiendront en réalité la moitié du nombre de pixels spécifié horizontalement. Lorsque ce sous-nom est absent, le sous-nom «image complète» doit être présent	8×8 pour les blocs comprenant 8 lignes d'image verticalement par 8 pixels d'image horizontalement. 128×128 pour les blocs comprenant 128 lignes d'image verticalement par 128 pixels d'image horizontalement
Images de bloc	Ce sous-nom indique la dimension temporelle des blocs S-T (voir le § 7.1.1), pour une fréquence vidéo de 30 fps. Par exemple, 6*F* représente un cinquième de seconde, quelle que soit la fréquence d'images utilisée (ce qui correspond à 5 images pour un système à 25 fps, 3 images pour un système à 15 fps, 2 images pour un système à 10 fps)	1F pour une dimension temporelle d'une image. 6F pour une dimension temporelle d'un cinquième de seconde
Statistique de bloc	Ce sous-nom indique la fonction statistique qui est utilisée pour extraire la caractéristique de chaque région S-T et qui produit un nombre pour chaque bloc de pixels S-T. Ce sous-nom est présent sauf si la «taille de bloc» est égale à 1×1 (c'est-à-dire un pixel). Avant application de la fonction «statistique de bloc», les résultats intermédiaires contiennent des historiques temporels des images avec un nombre par pixel (images filtrées); après, les résultats intermédiaires contiennent un nombre pour chaque région S-T (images de caractéristiques). Les paramètres associés à deux plans d'image (par exemple *hv*13 et coher_color) appliqueront séparément la fonction «statistique de bloc» aux deux plans d'image, produisant deux images de caractéristiques	mean est la moyenne des valeurs des pixels. std est l'écart type des valeurs des pixels. rms est la valeur quadratique moyenne des valeurs des pixels
Seuil de perceptibilité	Les valeurs produites par la fonction «statistique de bloc» peuvent être coupées à un seuil de perceptibilité *P*. Les valeurs comprises entre zéro et ce seuil sont remplacées par le seuil	*3* pour une valeur minimale de caractéristique de 3,0. *12* pour une valeur minimale de caractéristique de 12,0
Fonction de comparaison	Il s'agit de la fonction utilisée pour comparer les caractéristiques extraites des flux d'origine et traité (voir le § 8.2). Avant application de la fonction de comparaison, les résultats intermédiaires contiennent des historiques temporels des images de caractéristiques des flux d'origine et traité; après, les résultats intermédiaires contiennent un historique temporel des images du paramètre	log_gain (voir le § 8.2.1). ratio_loss (voir le § 8.2.1). euclid (voir le § 8.2.2)

TABLEAU 14

Convention utilisée pour la dénomination technique des paramètres de qualité vidéo

Sous-nom	Définition	Exemples
Fonction de regroupement spatial	Voir le § 8.3. La fonction est appliquée à chaque image du paramètre (par exemple toutes les régions S-T ayant le même indice temporel) et produit un historique temporel des valeurs du paramètre. Avant regroupement spatial, les résultats intermédiaires sont constitués des images du paramètre contenant une valeur pour chaque bloc S-T; après, les résultats intermédiaires correspondent à un historique temporel de nombres (historique temporel du paramètre). Ce sous-nom doit être présent pour tous les paramètres à l'exception des paramètres de type «image complète»	Voir le Tableau 12
Fonction de regroupement temporel	Voir le § 8.4. La fonction est appliquée à l'historique temporel du paramètre et produit une seule valeur du paramètre pour toute la séquence vidéo. Après regroupement temporel, le paramètre prend uniquement des valeurs négatives ou uniquement des valeurs positives. Zéro correspond à aucune dégradation et, plus la valeur du paramètre est éloignée de zéro, plus la dégradation est forte. Ce sous-nom doit être présent pour tous les paramètre	Voir le Tableau 13
Fonction non linéaire	Voir le § 8.5. L'examen des valeurs du paramètre peut indiquer qu'il convient d'appliquer une correction non linéaire au paramètre afin d'assurer une correspondance linéaire avec les données subjectives. C'est la fonction non linéaire qui procède à cette correction finale. Si la fonction sqrt est appliquée à un paramètre prenant uniquement des valeurs négatives, on commence par faire en sorte que le paramètre prenne uniquement des valeurs positives (on prend la valeur absolue)	sqrt pour la racine carrée de la valeur du paramètre issue du regroupement temporel. square pour le carré de la valeur du paramètre issue du regroupement temporel
Fonction de coupure	Voir le § 8.5. L'examen final des valeurs du paramètre peut indiquer qu'il est nécessaire de réduire encore la sensibilité du paramètre aux faibles dégradations (valeurs du paramètre proches de zéro). On remplace toute valeur comprise entre le niveau de coupure *T* et zéro par le niveau de coupure puis on soustrait le niveau de coupure de la valeur résultante du paramètre	clip_0.45 Si le paramètre ne prend que des valeurs positives, on remplace toute valeur inférieure à 0,45 par 0,45 puis on soustrait 0,45 de la valeur résultante du paramètre. Si le paramètre ne prend que des valeurs négatives, on remplace toute valeur supérieure à –0,45 par –0,45 puis on ajoute 0,45 à la valeur résultante du paramètre

8.6.1 Exemples de nom de paramètre

Le présent paragraphe inclut cinq exemples de nom technique, pour lesquels la procédure de sous-dénomination donnée au Tableau 14 est décrite pas à pas.

Y_si13_8x8_6F_std_6_ratio_loss_below5%_mean

Y signifie qu'on utilise le plan d'image de luminance. si13 signifie que l'on filtre les images au moyen des gabarits spatiaux 13×13 du § 7.2.1 en vue de l'extraction de la caractéristique f_{SI13} décrite au § 7.2.2. 8×8_6F signifie que l'on subdivise le flux vidéo en régions S-T contenant huit lignes d'image verticalement par huit pixels horizontalement par un cinquième de seconde temporellement (c'est-à-dire 6 images NTSC, 5 images PAL). std signifie que l'on prend l'écart type de chaque bloc. *6* signifie que l'on applique un seuil de perceptibilité et que l'on remplace toute valeur de l'écart type inférieure à 6,0 par 6,0. ratio_loss signifie que l'on compare les caractéristiques des flux d'origine et traité provenant de chaque bloc au moyen de la fonction ratio_loss. below5% sgnifie que l'on regroupe spatialement les valeurs du paramètre pour chaque indice temporel au moyen de la fonction below5%. mean signifie que l'on regroupe temporellement l'historique temporel du paramètre au moyen de la fonction mean.

color_coher_color_8x8_1F_mean_euclid_std_10%_clip_0.8

color signifie que l'on utilise les plans d'image C_B et C_R. coher_color signifie que l'on préserve la relation de phase entre les images C_B et C_R (en les traitant séparément) en vue de l'extraction de la caractéristique f_{COHER_COLOR} décrite au § 7.3. 8×8_1F signifie que l'on subdivise chaque image en blocs qui font 8 lignes d'image verticalement par 4 pixels C_B et C_R horizontalement (en raison du sous-échantillonnage 4:2:2 des plans d'image C_B et C_R) par 1 image temporellement. mean signifie que l'on prend la valeur moyenne pour chaque bloc. euclid signifie que l'on calcule la distance euclidienne entre le vecteur (C_B, C_R) issu du flux d'origine et le vecteur (C_B, C_R) issu du flux traité pour chaque bloc S-T. std signifie que l'on utilise la fonction de regroupement spatial std. *10%* signifie que l'on utilise la fonction de regroupement temporel 10%. clip_0.8 signifie que l'on applique une coupure de la valeur finale du paramètre à 0,8 (autrement dit, on remplace toute valeur inférieure à 0,8 par 0,8 puis on soustrait 0,8).

Y_hv13_angle0.225_rmin20_8x8_6F_mean_3_ratio_loss_below5%_mean_square_clip_0.05

Y signifie que l'on utilise le plan d'image de luminance. hv13 signifie que l'on filtre les images *Y* au moyen des gabarits spatiaux 13×13 du § 7.2.1 en vue de l'extraction de la caractéristique f_{HV13} décrite au § 7.2.2 (autrement dit, les images *HV* et $\overline{HV}$ sont créées et traitées séparément jusqu'à l'application du seuil de perceptibilité). angle0.225 et rmin20 signifient que l'on utilise un $\Delta\theta$ de 0,225 radians et un r_{min} de 20 pour le calcul de la caractéristique f_{HV13}. 8×8_6F signifie que l'on subdivise le flux vidéo en régions S-T contenant huit lignes d'image verticalement par huit pixels horizontalement par un cinquième de seconde temporellement (c'est-à-dire 6 images NTSC, 5 images PAL). mean signifie que l'on prend la valeur moyenne de *HV* et $\overline{HV}$ pour chaque bloc S-T. *3* signifie que l'on applique un seuil de perceptibilité à ces moyennes et que l'on remplace toute valeur inférieure à 3,0 par 3,0. On calcule ensuite la caractéristique f_{HV13} décrite au § 7.2.2 comme étant le rapport entre la moyenne coupée de *HV* et la moyenne coupée de $\overline{HV}$, comme spécifié dans hv13_angle0.225_rmin20, le sous-nom propre à la caractéristique. ratio_loss signifie que l'on applique la fonction de comparaison ratio_loss à la caractéristique f_{HV13} du flux d'origine et à la caractéristique f_{HV13} correspondante du flux traité pour chaque bloc S-T. below5% spécifie la fonction de regroupement spatial. mean spécifie la fonction de regroupement temporel. square spécifie la fonction non linéaire appliquée à la valeur du paramètre issue du regroupement temporel. clip_0.05 représente la fonction de coupure pour laquelle toute valeur inférieure à 0,05 est

remplacée par 0,05 puis 0,05 est soustrait de la valeur résultante (on rappelle qu'un paramètre ne prenant que des valeurs négatives devient un paramètre ne prenant que des valeurs positives après application de la fonction non linéaire *square*).

Y_contrast_ati_4x4_6F_std_3_ratio_gain_mean_10%

Y signifie que l'on utilise le plan de luminance. *contrast_ati* signifie que l'on calcule deux versions filtrées distinctes de l'image en vue de l'extraction de la caractéristique $f_{CONTRAST\_ATI}$ décrite au § 7.6. Le premier filtre, contrast, prend directement en considération les plans de luminance (§ 7.4). Le second filtre, ati, prend en considération les images correspondant aux différences entre les plans de luminance successifs (§ 7.5). Les images contrast et ati sont traitées séparément jusqu'à l'application du seuil de perceptibilité. 4×4_6F signifie que les deux flux vidéo sont subdivisés en régions S-T contenant quatre lignes d'image verticalement par quatre pixels horizontalement par un cinquième de seconde temporellement (par exemple 6 images NTSC, 5 images PAL). Le premier bloc S-T d'images *ati* ne contiendra en réalité que 5 images et non pas 6 car une image ati ne peut pas être générée pour la première image de la séquence (en effet, il n'existe pas d'image antérieure dans le temps disponible). Cette exception est spécifiée dans le cadre du sous-nom propre à la caractéristique. std signifie que l'on calcule l'écart type pour chaque bloc. Ensuite, comme spécifié au § 7.6, on applique un seuil de perceptibilité de 3 aux deux caractéristiques contrast et ati (on remplace toute valeur inférieure à 3 par 3,0). Ensuite, on multiplie la valeur de contrast avec la valeur de *ati* pour chaque bloc S-T (on trouvera dans la note de bas de page du § 7.6 les instructions particulières sur la manière de procéder à cette multiplication) et on poursuit les calculs avec cette image de caractéristique combinée. ratio_gain est la fonction de comparaison utilisée pour comparer la caractéristique du flux d'origine et la caractéristique du flux traité pour chaque bloc S-T. mean est la fonction de regroupement spatial. 10% est la fonction de regroupement temporel.

9 Modèle général

Le présent paragraphe contient une description complète du calcul de la qualité VQM selon le modèle général (désignée par VQM_G). Ce calcul est optimisé de manière à obtenir la corrélation maximale entre les mesures objectives et les mesures subjectives pour une large plage de niveaux de qualité vidéo et de débits binaires. Le modèle général comporte des paramètres objectifs pour la mesure des effets perçus d'une grande variété de dégradations telles que le flou, la distorsion due à la subdivision en blocs, les mouvements saccadés/non naturels, le bruit (à la fois dans les canaux de luminance et de chrominance) et les blocs erronés (par exemple ce que l'on peut généralement voir lorsque des erreurs de transmission numérique sont présentes). Le calcul décrit ici consiste en une combinaison linéaire de paramètres de qualité vidéo dont les conventions de dénomination sont décrites au § 8.6. Les paramètres de qualité vidéo ont été choisis sur la base des critères d'optimisation donnés ci-dessus. Ce calcul donne des valeurs comprises entre zéro (pas de dégradation perçue) et environ un (dégradation perçue maximale). Pour pouvoir comparer les résultats avec ceux obtenus par la méthode à double stimulus utilisant une échelle de qualité continue (DSCQS), on multiplie les résultats VQM_G par 100.

La conception du modèle général repose sur des séquences vidéo conformes à la Recommandation UIT-R BT.601 qui ont été évaluées subjectivement à une distance de visualisation de six hauteurs d'image. Lorsqu'on analyse les séquences vidéo pour différentes distances de visualisation, il faut appliquer un facteur de correction aux résultats. Plus la distance de visualisation est grande, moins les dégradations sont visibles; plus la distance de visualisation est petite, plus les dégradations sont visibles. Il convient de faire attention lorsqu'on compare les résultats pour des séquences vidéo qui sont observées à des distances de visualisation différentes.

La qualité VQM_G est donnée par une combinaison linéaire de sept paramètres. Quatre paramètres sont fondés sur des caractéristiques extraites des gradients spatiaux de la composante de luminance Y (voir le § 7.2.2), deux paramètres sont fondés sur des caractéristiques extraites du vecteur formé par les deux composantes de chrominance (C_B, C_R) (voir le § 7.3) et un paramètre est fondé sur les caractéristiques de contraste et d'information temporelle absolue, toutes deux extraites de la composante de luminance Y (voir les § 7.4 et 7.5, respectivement). La qualité VQM_G est donnée par:

$$VQM_G = \{-0.2097 * Y\_si13\_8\times8\_6F\_std\_12\_ratio\_loss\_below5\%\_10\%$$

$$+0.5969 * Y\_hv13\_angle0.225\_rmin20\_8\times8\_6F\_mean\_3\_ratio\_loss\_below5\%\_mean\_square\_clip\_0.06$$

$$+0.2483 * Y\_hv13\_angle0.225\_rmin20\_8\times8\_6F\_mean\_3\_log\_gain\_above95\%\_mean$$

$$+0.0192 * color\_coher\_color\_8\times8\_1F\_mean\_euclid\_std\_10\%\_clip\_0.6$$

$$-2.3416 * [Y\_si13\_8\times8\_6F\_std\_8\_log\_gain\_mean\_mean\_clip\_0.004 \mid ^{0.14}]$$

$$+0.0431 * Y\_contrast\_ati\_4\times4\_6F\_std\_3\_ratio\_gain\_mean\_10\%$$

$$+0.0076 * color\_coher\_color\_8\times8\_1F\_mean\_euclid\_above99\%tail\_std\} \mid_{0.0}$$

Il est rappelé que les caractéristiques ci-dessus pour le modèle général avec une dimension temporelle de «6F» correspondent en réalité à cinq images vidéo PAL (625 lignes).

L'élévation au carré du paramètre hv_loss est nécessaire pour linéariser la réponse du paramètre par rapport aux données subjectives. Il est à noter que, comme le paramètre hv_loss devient positif après l'élévation au carré, on utilise un poids multiplicatif positif. Il est par ailleurs à noter que le paramètre hv_loss est coupé à 0,06, le paramètre colour est coupé à 0,6 et le paramètre si_gain est coupé à 0,004. Le paramètre si_gain est le seul paramètre du modèle correspondant à une amélioration de la qualité (comme le paramètre si_gain est positif, un poids négatif conduit à des contributions négatives à la qualité VQM, autrement dit à des améliorations de la qualité). Le paramètre si_gain mesure les améliorations de la qualité qui résultent de l'accentuation des contours. Une coupure du paramètre à un seuil supérieur de 0,14 immédiatement avant la multiplication par le poids associé au paramètre empêche toute amélioration excessive de la qualité VQM de plus de 1/3 d'une unité de qualité, qui est l'amélioration maximale observée dans l'ensemble général des données subjectives (autrement dit, l'accentuation des contours opérée par un HRC ne permet d'améliorer la qualité que dans une faible mesure).

La qualité VQM totale (une fois que les contributions de tous les paramètres ont été ajoutées) est coupée à un seuil inférieur de 0,0 pour éviter les valeurs VQM négatives. Enfin, une fonction d'écrasement autorisant un maximum de 50% de dépassement est appliquée aux valeurs VQM supérieures à 1,0 afin de limiter les valeurs VQM qui sont associées à des séquences vidéo comportant de fortes distorsions et qui se situent en dehors de la plage des données subjectives disponibles.

$$\text{Si } VQM_G > 1{,}0, \text{ alors } VQM_G = (1 + c) * VQM_G / (c + VQM_G), \text{ où } c = 0{,}5.$$

Les valeurs de la qualité VQM_G calculées comme indiqué ci-dessus seront supérieures ou égales à zéro et présenteront une valeur nominale maximale de un. La valeur de la qualité VQM_G peut parfois être supérieure à un pour des scènes vidéo comportant de très fortes distorsions.

10 Références bibliographiques

JAIN, A. K. [1989] *Fundamentals of Digital Image Processing*. Englewood Cliffs, NJ: Prentice-Hall Inc., p. 348-357.

PINSON, M. et WOLF, S. [février 2002] Video Quality Measurement User's Manual. NTIA Handbook 02-1. National Telecommunications and Information Administration.

SMPTE [1995a] Norme SMPTE 125M. Television – Component Video Signal 4:2:2 – Bit-Parallel Digital Interface. Society of Motion Picture and Television Engineers, 595 West Hartsdale Avenue, White Plains, NY 10607.

SMPTE [1995b] Norme SMPTE Recommended Practice 187. Center, Aspect Ratio, and Blanking of Video Images. Society of Motion Picture and Television Engineers, 595 West Hartsdale Avenue, White Plains, NY 10607.

SMPTE [1999] Norme SMPTE 170M. Television – Composite Analog Video Signal – NTSC for Studio Applications. Society of Motion Picture and Television Engineers, 595 West Hartsdale Avenue, White Plains, NY 10607.

WOLF, S. et PINSON, M. [12-13 novembre 1998] In-service performance metrics for MPEG-2 video systems. Proc. Made to Measure 98 – Measurement Techniques of the Digital Age Technical Seminar, conférence technique cofinancée par l'International Academy of Broadcasting (IAB), l'UIT et la Technical University of Braunschweig (TUB), Montreux, Suisse.

WOLF, S. et PINSON, M. [septembre 1999] Spatial-temporal distortion metrics for in-service quality monitoring of any digital video system. Proc. SPIE International Symposium on Voice, Video, and Data Communications, Boston, MA.

WOLF, S. et PINSON, M. [juillet 2001] The relationship between performance and spatial-temporal region size for reduced-reference, in-service video quality monitoring systems. Proc. SCI/ISAS 2001 (Systematics, Cybernetics, and Informatics/Information Systems Analysis and Synthesis), p. 323-328. National Telecommunications and Information Administration.

WOLF, S. et PINSON, M. [juin 2002] Video Quality Measurement Techniques. NTIA Report 02-392. National Telecommunications and Information Administration.

Annexe 5a

Données objectives brutes sur les mesures VQM de la NTIA

La présente Annexe expose l'ensemble des données objectives brutes de la NTIA sur les mesures VQM.

Résumé concernant les données brutes

Le modèle général élaboré par la NTIA a été conçu au départ pour des valeurs en sortie sur une échelle nominale de 0 à 1 où 0 correspond à la perception d'aucune dégradation et 1 à la perception d'une dégradation maximale. Toutefois, l'exécutable binaire soumis au test VQEG FR-TV Phase II a transformé les valeurs (0, 1) du modèle général à (0, 100) dans un souci d'adaptation avec la DSCQS. Etant donné que toutes les valeurs du modèle devraient désormais être mises à l'échelle à (0, 1), nous avons supprimé le facteur de multiplication par 100 (c'est-à-dire multiplication par 100) pour retrouver l'échelle originale (0, 1) du modèle général.

Les valeurs du modèle général calculées ici ont utilisé les 8 s centrales de chaque clip vidéo rejetant les 10 trames supplémentaires au début et à la fin de chaque fichier vidéo comme décrit dans le programme de tests VQEG Phase II FR-TV. Pour les routines d'étalonnage, on a utilisé une incertitude de 30 trames et une fréquence de 15 images (voir le § 6 de l'Annexe 5). Par ailleurs, la région SROI utilisée pour calculer la valeur VQM pour chaque clip a été choisie comme suit:

Etape 1: Pour les systèmes vidéo à 525 lignes, on utilise une région SROI par défaut de 672 pixels × 448 lignes centrée dans l'image vidéo. Pour les systèmes vidéo à 625 lignes, on utilise une région SROI par défaut de 672 pixels × 544 lignes centrée dans l'image vidéo. Ces régions SROI par défaut peuvent être modifiées comme indiqué dans les Etapes 2 et 3.

Etape 2: Le modèle a besoin de 6 pixels/lignes valables supplémentaires sur tous les côtés de la région SROI susmentionnée pour que les filtres spatiaux puissent fonctionner correctement. Si la PVR (calculée automatiquement selon la méthode donnée au § 6.2 de l'Annexe 5) n'est pas suffisamment large pour englober la région SROI par défaut + 6 pixels/lignes Etape 1), la région SROI est alors réduite par des multiples de 8 pixels/lignes, uniquement dans la direction nécessaire (horizontale ou verticale).

Etape 3: La région SROI est toujours centrée horizontalement de façon à ce que l'échantillon gauche commence en un point d'échantillonnage identique pour la luminance/chrominance de la Recommandation UIT-R BT.601. La région SROI est centrée verticalement de façon à ce que lorsqu'elle est subdivisée en deux trames, le même nombre de lignes est rejeté depuis le haut de chaque trame. Si la taille de la région SROI a été réduite dans l'Etape 2, un centrage parfait de la région SROI dans l'image vidéo ne sera peut-être pas possible.

Il est possible de télécharger le logiciel d'évaluation permettant de mettre en oeuvre le modèle général et ses routines d'étalonnage à l'adresse suivante:

http://www.its.bldrdoc.gov/n3/video/vqmsoftware.htm

TABLEAU 15

Données objectives brutes pour un système à 525 lignes

N° Source	N° HRC	NTIA: Modèle H
1	1	0,660[1]
1	2	0,347
1	3	0,286
1	4	0,178
2	1	0,449
2	2	0,246
2	3	0,119
2	4	0,061
3	1	0,321
3	2	0,167
3	3	0,076
3	4	0,049
4	5	0,396
4	6	0,280
4	7	0,222
4	8	0,183
5	5	0,329
5	6	0,217
5	7	0,159
5	8	0,115
6	5	0,542
6	6	0,266
6	7	0,189
6	8	0,139
7	5	0,258
7	6	0,161
7	7	0,108
7	8	0,076
8	9	0,911
8	10	0,717
8	11	0,721
8	12	0,526
8	13	0,424
8	14	0,311
9	9	0,827
9	10	0,453
9	11	0,512
9	12	0,264
9	13	0,188
9	14	0,124
10	9	0,666
10	10	0,250
10	11	0,375
10	12	0,129
10	13	0,078
10	14	0,153
11	9	0,513
11	10	0,534
11	11	0,407
11	12	0,161
11	13	0,148
11	14	0,159
12	9	0,600
12	10	0,410
12	11	0,471
12	12	0,244
12	13	0,171
12	14	0,114
13	9	0,537
13	10	0,425
13	11	0,346
13	12	0,215
13	13	0,188
13	14	0,169

[1] Pour la source 1, HRC 1, le logiciel d'étalonnage soumis au VQEG a abouti à une erreur d'alignement spatial/temporel qui a estimé de façon incorrecte la séquence vidéo traitée qui sera resynchronisée (c'est-à-dire décalée d'une trame, voir le § 6.1.2 de l'Annexe 5). Pour les autres scènes de HRC 1, l'alignement spatial/temporel a été correctement estimé. Il est recommandé au § 6.1.5.7 de l'Annexe 5 de soumettre les résultats de l'étalonnage à un filtrage médian pour toutes les scènes d'un HRC donné afin d'obtenir des estimations d'étalonnage plus fiables pour ce HRC. Toutefois, le programme des essais de Phase II du VQEG précisait que tous les logiciels VQM donnent une estimation de qualité unique pour chaque clip vidéo. Par conséquent, un filtrage médian des résultats d'étalonnage pour toutes les scènes d'un HRC donné n'a pas été autorisé par le programme des essais. Si le filtrage médian des résultats d'étalonnage avait été autorisé, le logiciel VQM aurait correctement aligné ce clip vidéo et la note objective brute aurait été de 0,529.

TABLEAU 16

Données objectives brutes pour un système à 625 lignes

N° Source	N° HRC	NTIA: Modèle H
1	2	0,421
1	3	0,431
1	4	0,264
1	6	0,205
1	8	0,155
1	10	0,123
2	2	0,449
2	3	0,473
2	4	0,312
2	6	0,260
2	8	0,226
2	10	0,145
3	2	0,472
3	3	0,506
3	4	0,308
3	6	0,239
3	8	0,183
3	10	0,146
4	2	0,409
4	3	0,458
4	4	0,384
4	6	0,354
4	8	0,280
4	10	0,232
5	2	0,470
5	3	0,521
5	4	0,260
5	6	0,234
5	8	0,132
5	10	0,083
6	2	0,391
6	3	0,364
6	4	0,290
6	6	0,252
6	8	0,181
6	10	0,169
7	4	0,422
7	6	0,385
7	9	0,336
7	10	0,270
8	4	0,345
8	6	0,311
8	9	0,280
8	10	0,242
9	4	0,344
9	6	0,285
9	9	0,246
9	10	0,192
10	4	0,410
10	6	0,355
10	9	0,313
10	10	0,241
11	1	0,739
11	5	0,468
11	7	0,199
11	10	0,201
12	1	0,548
12	5	0,441
12	7	0,367
12	10	0,307
13	1	0,598
13	5	0,409
13	7	0,321
13	10	0,277

Appendice 1

Résultats des tests FR-TV Phase II du Groupe d'experts sur la qualité vidéo

1 Introduction

On a évalué les résultats des modèles de qualité perceptuelle décrits dans la présente Recommandation dans le cadre de deux évaluations parallèles. Dans la première évaluation, on a utilisé une méthode subjective normalisée, la méthode de DSCQS pour obtenir des indices subjectifs de la qualité des séquences vidéo auprès de groupes d'observateurs humains. Dans la seconde évaluation, les indices objectifs de la qualité ont été obtenus à l'aide de modèles de calcul objectifs. Pour chaque modèle, on a mesuré à plusieurs reprises la précision et la cohérence avec lesquelles les notes objectives prédisent les notes subjectives.

Le présent Appendice décrit la partie évaluation subjective de l'essai ainsi que les résultats des modèles de calcul objectifs soumis par les modèles suivants:

– Modèle 1 (British Telecom; modèle D dans les tests FR-TV Phase II du VQEG);

– Modèle 2 (Yonsei University/Radio Research Laboratory/SK Telecom; modèle E dans les essais FR-TV Phase II du VQEG);

– Modèle 3 (CPqD; modèle F dans les essais FR-TV Phase II du VQEG);

– Modèle 4 (NTIA; modèle H dans les essais FR-TV Phase II du VQEG).

Des laboratoires indépendants ont mené les essais subjectifs. Deux laboratoires, Communications Research Center (CRC, Canada) et Verizon (Etats-Unis d'Amérique), ont réalisé les essais avec des séquences 525/60 Hz et un troisième laboratoire Fondazione Ugo Bordoni (FUB, Italie), a effectué des essais avec des séquences 625/50 Hz.

Une description détaillée des essais FR-TV Phase II du VQEG est donnée dans le Document[1] mentionné.

2 Séquences vidéo

Le format des séquences vidéo d'essai 525/60 ou 625/50 lignes était le format vidéo en composante 4:2.2 de la Recommandation UIT-R BT.601 avec un format d'image de 4:3.

2.1 SRC et HRC

Pour chacun des essais à 525 ou 625 lignes, on a utilisé 13 séquences source (SRC) avec des caractéristiques différentes (format, information temporelle ou spatiale, couleur, etc.) (voir Tableaux 17 et 18).

Dans les deux essais, les circuits HRC ont été choisis de façon à représenter des conditions types de distribution secondaire de signaux vidéonumériques de qualité télévision. Dans l'essai avec un système à 625 lignes, on a utilisé 10 circuits HRC; leurs caractéristiques sont présentées dans le Tableau 19. Dans l'essai avec un système à 525 lignes, on a utilisé 14 circuits HRC dont les caractéristiques sont présentées dans le Tableau 20.

Dans un essai comme dans l'autre, les séquences SRC et les circuits HRC ont été combinés en une matrice creuse (voir les Tableaux 23 à 26).

TABLEAU 17

Séquences de format 625/50

Numéro de la séquence SRC	Caractéristiques
1	Vue de l'horizon prise depuis un bateau en mouvement; au départ film 16:9 converti en télécinéma à 576i/50
2	Danseurs sur plancher en bois; mouvements rapides, détail modéré; initialement, format D5
3	Match de volley-ball masculin en intérieur; format D5
4	Match de football féminin, plan rapide; format D5
5	Animation classique 12 fps; source convertie en un film à 24 fps puis en télécinéma à 576i/50
6	Globe en fil de fer tournant lentement; DigiBetCam
7	Mouvements rapides (scène et caméra), effets d'éclairage
8	Gros plan d'un guitariste jouant de son instrument, effets d'éclairage
9	Couleur, mouvements, détail
10	Beaucoup de détails, fond avec texture visible, mouvements
11	Couleur, mouvements, détail
12	Match de rugby en extérieur; mouvement, couleur
13	Mouvements, détail, étendue d'eau en mouvement
14 (démo)	Mouvements rapides (scène et caméra), avec effets d'éclairage
15 (démo)	Mouvements rapides (scène et caméra), avec effets d'éclairage
16 (démo)	Gros plan de visage suivi d'un plan large de chantier

TABLEAU 18

Séquences format 525/60 (SRC)

Numéro de la séquence SRC	Caractéristiques
1	Match de football en extérieur, avec couleur, mouvements et fond avec texture visible
2	Paysage d'automne avec couleur détaillée, travelling optique lent
3	Animation contenant des mouvements, des couleurs et des plans de coupe
4	Scène dans un parc avec de l'eau, beaucoup de détails; original TVHD
5	Couleur et mouvements rapides; original TVHD
6	Couleur, large étendue d'eau; original TVHD
7	Match de football de rue, mouvements modérés; original TVHD
8	Parc d'attractions aquatiques (DigiBetaCam)
9	Excursion dans un parc d'attractions, mouvements modérés, beaucoup de détails, travelling optique lent (DigiBetaCam)
10	Couleur, mouvements, illumination faible modérée (DigiBetaCam)
11	Animation classique 12 fps, convertie en film 24 fps puis télécinéma à 480i/60
12	Fontaine en extérieur, détail, avec zoom; (DigiBetaCam)
13	Plans de coupe, gros plan sur la clé de contact puis plan large distant et retour; film converti au télécinéma à 480i/60
14 (démo)	Gros plan d'une rose, brise légère, mouvements, couleur et détail; (DigiBetaCam)
15 (démo)	Beaucoup de détails, mouvements lents; original TVHD
16 (démo)	Rotation lente de statues, branches d'arbres qui oscillent; (DigiBetaCam)

TABLEAU 19

HRC 625/50

Numéro du HRC	Débit	Résolution	Méthode	Observations
1	768 kbit/s	CIF	H.263	Ecran plein (HRC15 de VQEG 1)
2	1 Mbits/s	320H	MPEG2	Modèle codé
3	1,5 Mbit/s	720H	MPEG2	Codé par FUB
4	2,5→4 Mbit/s	720H	MPEG2	Mis en cascade par FUB
5	2 Mbit/s	3/4	MPEG2 sp@ml	HRC13 de VQEG 1
6	2,5 Mbit/s	720H	MPEG2	Codé par FUB
7	3 Mbit/s	totale	MPEG2	HRC9 de VQEG 1
8	3 Mbit/s	704H	MPEG2	Modèle codé
9	3 Mbit/s	720H	MPEG2	Codé par FUB
10	4 Mbit/s	720H	MPEG2	Codé par FUB

TABLEAU 20

HRC 525/60

Numéro du HRC	Débit	Résolution	Méthode	Observations
1	768 kbit/s	CIF	H.263	Ecran plein (HRC15 de VQEG 1)
2	2 Mbit/s	3/4	MPEG2, sp@ml	HRC13 de VQEG 1
3	3 Mbit/s	totale	MPEG2	HRC9 de VQEG 1
4	5 Mbit/s	720H	MPEG2	Codé par CRC
5	2 Mbit/s	704H	MPEG2	Codé par CRC
6	3 Mbit/s	704H	MPEG2	Codé par CRC
7	4 Mbit/s	704H	MPEG2	Codé par CRC
8	5 Mbit/s	704H	MPEG2	Codé par CRC
9	1 Mbit/s	704H	MPEG2	Faible débit binaire combiné à une résolution élevée
10	1 Mbit/s	480H	MPEG2	Codé par CRC; faible débit binaire, faible résolution
11	1,5 Mbit/s	528H	MPEG2	Modèle codé; modulation MAQ-64; Sortie NTSC composite convertie en une sortie en composante
12	4->2 Mbit/s	720H	MPEG2	Modèle codé; codeurs en cascade
13	2,5 Mbit/s	720H	MPEG2	Codé par CRC
14	4 Mbit/s	720H	MPEG2	Modèle codé; utilisation d'un codec logiciel

3 Méthode d'évaluation des résultats des modèles objectifs

On a utilisé la méthode DSCQS de la Recommandation UIT-R BT.500 pour les essais subjectifs. Pour les essais avec les systèmes à 525 lignes, des notes d'opinion moyenne de dégradation (DMOS, *difference mean opinion scores*) ont été recueillies pour 63 combinaisons SRC × HRC. Pour les essais avec les systèmes à 625 lignes, les notes d'opinion moyenne de dégradation ont été recueillies pour 64 combinaisons SRC × HRC. Pour les mêmes combinaisons SRC × HRC, des données objectives ont également été obtenues pour chaque modèle de calcul objectif.

Pour les besoins de l'évaluation des modèles, les données subjectives ont été mises à l'échelle et les données objectives ont subi une transformation non linéaire vers une échelle variant de 0 (pas distinguable de la source) à 1. La transformation non linéaire est donnée par l'équation:

$$DMOS_p = b1/(1 + \exp(-b2*(VQR - b3)))$$

où:

VQR: (indice de qualité vidéo), valeur en sortie effective du modèle de calcul objectif

$DMOS_p$: valeur ayant subi la transformation non linéaire.

Les performances des modèles objectifs ont été évaluées sous trois angles, du point de vue de leur aptitude à faire une évaluation subjective de la qualité vidéo:

– précision du modèle prédictif – capacité du modèle à prévoir les indices de qualité subjectifs avec un faible pourcentage d'erreur;
– monotonicité du modèle prédictif – degré de concordance entre les prédictions du modèle et les fourchettes relatives des indices de qualité subjectifs; et
– cohérence du modèle prédictif – mesure dans laquelle l'exactitude de prédiction du modèle est maintenue sur toute la série de séquences vidéo tests; en d'autres termes, le résultat obtenu n'est pas affecté par diverses dégradations vidéo.

Ces attributs ont été évalués dans le cadre de sept mesures de performance qui sont décrites ci-après.

Mesure 1: Coefficient de corrélation linéaire de Pearson entre $DMOS_p$ et *DMOS*.

Mesure 2: Coefficient de corrélation de rang de Spearman entre $DMOS_p$ et *DMOS*.

La corrélation de Spearman et la corrélation de Pearson ainsi que toutes les autres statistiques ont été calculées simultanément pour toutes les combinaisons SRC × HRC.

Mesure 3: Proportion de «points éloignés» par rapport au nombre total de points *N*.

Proportion des points éloignés = (nombre total de points éloignés)/*N*

où un point éloigné est un point pour lequel ABS[Qerror[i]] > 2∗DMOSStandardError[i].

L'erreur type DMOS a servi, à deux reprises, de valeur seuil pour définir un point éloigné.

Mesures 4, 5, 6: Ces mesures ont été évaluées sur la base de la méthode décrite dans le Rapport T1.TR.72-2001 [ATIS, 2001]:

4. Erreur quadratique moyenne;

5. Puissance de résolution; et

6. Erreurs de classification.

A noter que pour l'évaluation des modèles à l'aide de cette méthode, la procédure d'étalonnage croisé décrite a été sautée, étant donné qu'elle ne concerne pas les mesures de performance des différents modèles.

Mesure 7: Cette mesure est basée sur le test *F*. Deux mesures de test *F* ont été effectuées. La première utilise l'erreur quadratique moyenne calculée à partir d'indices de sujets individuels. L'erreur quadratique moyenne a été calculée pour un modèle «nul ou optimal», ce qui correspondait à la note DMOS observée et les résidus associés, et pour chacun des modèles objectifs. Des tests *F* ont été effectués pour comparer l'erreur quadratique moyenne associée au modèle nul à celle associée à chaque modèle et l'erreur quadratique moyenne associée au modèle le plus performant à celle associée aux autres modèles. La deuxième mesure *F* a été basée sur l'erreur quadratique moyenne calculée à partir d'indices moyens, c'est-à-dire la note DMOS. Plus précisément, les erreurs quadratiques moyennes ont été calculées pour chaque modèle en utilisant la variation résiduelle entre la note DMOS prévisionnelle et la note DMOS observée. Des tests *F* ont été effectués pour comparer l'erreur quadratique moyenne associée au modèle le plus performant à celle associée aux autres modèles.

4 Evaluation des résultats

Les résultats des calculs de mesure sont présentés dans les Tableaux 21 et 22, un pour les données concernant les systèmes à 525 lignes et un pour les données concernant les systèmes à 625 lignes.

Les sept mesures dans les tableaux concordent toutes presque parfaitement. Un modèle objectif qui donne de bons résultats pour une méthode de mesure le fait également pour les autres. Par ailleurs, le classement des modèles objectifs en fonction des différentes méthodes de mesure est fondamentalement le même pour l'un et l'autre format vidéo. Toutefois, les résultats des deux essais (525 et 625 lignes) sont similaires mais non identiques. Les changements apparents dans le classement d'un test à l'autre étaient peu nombreux.

Les données subjectives mises à l'échelle qui ont été utilisées pour calculer ces mesures sont présentées dans les Tableaux 23 à 26. Les données objectives correspondantes obtenues avec les quatre modèles de calcul objectifs sont présentées dans les Annexes 2 à 5.

5 Données relatives au rapport PSNR

La valeur du rapport PSNR est une mesure simple de la qualité vidéo. Les résultats des méthodes de mesure de la qualité vidéo peuvent être comparés à ceux du rapport PSNR. On a calculé le rapport PSNR pour les séquences d'essai, pour plusieurs modèles. Les résultats des méthodes de mesure pour le rapport PSNR le plus élevé sont consignés dans les Tableaux 21 et 22.

TABLEAU 21

Résumé des analyses pour les systèmes à 525 lignes

Numéro de la ligne	Mesure	D525	E525	F525	H525	PSNR525
1	1. Corrélation de Pearson	0,937	0,857	0,835	0,938	0,804
2	2. Correlation de Spearman	0,934	0,875	0,814	0,936	0,811
3	3. Proportion des points éloignés	33/63 = 0,52	44/63 = 0,70	44/63 = 0,70	29/63 = 0,46	46/63 = 0,73
4	4. Erreur quadratique moyenne, 63 points de données	0,075	0,11	0,117	0,074	0,127
5	5. Puissance de résolution, delta VQM (puissance de résolution faible est préférable)	0,2177	0,2718	0,3074	0,2087	0,3125
6	6. Pourcentage d'erreurs de classification (minimum par rapport à VQM delta)	0,1889	0,2893	0,3113	0,1848	0,3180
7	7. Modèle MSE/modèle MSE optimal	1,262	1,59	1,68	1,256	1,795
8	F = Modèle MSE/MSE modèle H	1,005	1,266	1,338	1	1,429
9	Modèle MSE, 4 219 points de données	0,02421	0,03049	0,03223	0,02409	0,03442
10	Modèle MSE optimal, 4 219 points de données	0,01918	0,01918	0,01918	0,01918	0,01918
11	Modèle MSE, 63 points de données	0,00559	0,01212	0,01365	0,00548	0,01619
12	F= Modèle MSE63/MSE63 modèle H	1,02	2,212	2,491	1	2,954

NOTE 1 – Les Mesures 5 et 6 ont été effectuées à l'aide du code Matlab® publié dans T1.TR.72-2001.

NOTE 2 – La Mesure 5 est une mesure visuelle à partir de diagrammes de dispersion indiqués dans les documents.

NOTE 3 – Les valeurs de Mesure 7 qui sont inférieures à 1,07 indiquent que le modèle n'est pas différent de façon fiable du modèle optimal.

NOTE 4 – Les valeurs à la ligne 8 qui sont supérieures à 1,07 indiquent que le modèle a des résidus beaucoup plus importants que le meilleur modèle, à savoir le modèle *H* en l'occurrence.

NOTE 5 – Les valeurs à la ligne 12 qui sont supérieures à 1,81 indiquent que le modèle a des résidus beaucoup plus importants que le meilleur modèle considéré, à savoir le modèle *H* en l'occurrence.

TABLEAU 22

Résumé des analyses pour les systèmes à 625 lignes

Numéro de la ligne	Mesure	D625	E625	F625	H625	PSNR625
1	1. Corrélation de Pearson	0,779	0,87	0,898	0,886	0,733
2	2. Correlation de Spearman	0,758	0,866	0,883	0,879	0,74
3	3. Proportion des points éloignés	28/64 = 0,44	24/64 = 0,38	21/64 = 0,33	20/64 = 0,31	30/64 = 0,47
4	4. Erreur quadratique moyenne, 64 points de données	0,113	0,089	0,079	0,083	0,122
5	5. Puissance de résolution, VQM delta (une puissance de résolution faible est préférable)	0,321	0,281	0,270	0,267	0,313
6	6. Pourcentage d'erreurs de classification (minimum par rapport à VQM delta)	0,305	0,232	0,204	0,199	0,342
7	7. Modèle MSE/modèle MSE nul	1,652	1,39	1,303	1,339	1,773
8	F = Modèle MSE/MSE modèle F	1,268	1,067	1	1,028	1,361
9	Modèle MSE, 1 728 points de données	0,02953	0,02484	0,02328	0,02393	0,03168
10	Modèle MSE nul, 1 728 points de données	0,01787	0,01787	0,01787	0,01787	0,01787
11	Modèle MSE, 64 points de données	0,0127	0,00786	0,00625	0,00693	0,01493
12	F= Modèle MSE64/MSE64 modèle F	2,032	1,258	1	1,109	2,389

NOTE 1 – Les Mesures 5 et 6 ont été effectuées à l'aide du code Matlab® publié dans T1.TR.72-2001.

NOTE 2 – La Mesure 5 est une mesure visuelle à partir de diagrammes de dispersion indiqués dans les documents.

NOTE 3 – Les valeurs de Mesure 7 qui sont inférieures à 1,12 indiquent que le modèle n'est pas différent de façon fiable du modèle optimal.

NOTE 4 – Les valeurs à la ligne 8 qui sont supérieures à 1,12 indiquent que le modèle a des résidus beaucoup plus importants que le meilleur modèle, à savoir le modèle *F en l'occurrence*.

NOTE 5 – Dans le cas de données 625 lignes avec 1 728 observations, la valeur critique des statistiques de *F* est de 1,12.

NOTE 6 – Les valeurs à la ligne 12 qui sont supérieures à 1,81 indiquent que le modèle a des résidus beaucoup plus importants que le meilleur modèle considéré, à savoir le modèle *F* en l'occurrence.

TABLEAU 23

Données subjectives pour toutes les combinaisons HRC-SRC 525/60 – (valeurs DMOS)

	HRC													
SRC (Image)	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**
1	0,5402368	0,5483205	0,4024097	0,3063528										
2	0,5025558	0,3113346	0,1881739	0,1907347										
3	0,4682724	0,3088831	0,1300389	0,1293293										
4					0,6742005	0,4250873	0,3762656	0,2972294						
5					0,4682559	0,3203024	0,2071702	0,1652752						
6					0,5690291*	0,4370961	0,3591788	0,2482169						
7					0,3796362	0,2276934	0,1644409	0,1819566						
8									0,9513387	0,789748	0,8405916	0,5221555	0,4572049	0,4614104
9									0,8262912	0,660339	0,7100111	0,4921708	0,3656559	0,2960957
10									0,9084171	0,5908784	0,7302376	0,3345703	0,2565459	0,2953144
11									0,6675853	0,7054929	0,5761193	0,32761	0,310495	0,331051
12									0,7883371	0,6295301	0,6809288	0,3651402	0,2714356	0,2782449
13									0,7211194	0,5545722	0,5525494	0,2708744	0,27549	0,2733771

NOTE 1 – La valeur SRC = 6, HRC = 5 (*) a été prise de l'analyse car elle dépassait les caractéristiques d'alignement temporel du programme des essais.

TABLEAU 24

Données subjectives pour toutes les combinaisons HRC-SRC 625/50 – (valeurs DMOS)

	HRC									
SRC (Image)	**1**	**2**	**3**	**4**	**5**	**= 6**	**7**	**8**	**9**	**10**
1		0,59461	0,64436	0,40804		0,34109		0,2677		0,26878
2		0,54173	0,70995	0,27443		0,22715		0,21133		0,16647
3		0,73314	0,76167	0,49848		0,38613		0,34574		0,26701
4		0,58528	0,90446	0,62361		0,61143		0,43329		0,26548
5		0,61973	0,68987	0,41648		0,4218		0,27543		0,2022
6		0,38852	0,44457	0,27983		0,28106		0,23726		0,17793
7				0,59953		0,55093			0,45163	0,35617
8				0,32528		0,32727			0,30303	0,26366
9				0,47656		0,49924			0,39101	0,37122
10				0,70492		0,58218			0,49711	0,37854
11	0,79919				0,59256		0,34337			0,30567
12	0,61418				0,6661		0,53242			0,44737
13	0,74225				0,66799		0,42065			0,33381

TABLEAU 25

Données subjectives pour toutes les combinaisons HRC-SRC 525/60 – (valeurs des erreurs types)

	HRC													
SRC (Image)	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**
1	0.02109499	0,0223858	0,0202654	0,0200377										
2	0,02072424	0,0186353	0,0164296	0,0179823										
3	0,02075164	0,021336	0,0131301	0,0141977										
4					0,0224479	0,0200094	0,0221945	0,0216022						
5					0,0254351	0,0217278	0,0179396	0,0145813						
6						0,0215159	0,0176766	0,0180308						
7					0,0197204	0,0171224	0,0147712	0,0188843						
8									0,010892	0,0180687	0,0185947	0,0249537	0,0272349	0,0258362
9									0,0167711	0,018702	0,0281708	0,0226776	0,0193788	0,0203533
10									0,0144376	0,0263593	0,0171287	0,0202314	0,01996	0,018688
11									0,0186046	0,0189571	0,0213137	0,0188185	0,020292	0,0183653
12									0,0175106	0,0223805	0,0216039	0,0192717	0,0183	0,0202472
13									0,0213225	0,023069	0,0238845	0,0196748	0,0187747	0,0201108

NOTE 1 – Pour convertir aux écarts types, multiplier par la racine carrée du nombre d'observations, 66.

NOTE 2 – La valeur SRC = 6, HRC = 5 a été prise de l'analyse car elle dépassait les critères d'alignement temporel du programme des essais.

TABLEAU 26

Données subjectives pour toutes les combinaisons HRC-SRC 625/60 – (valeurs des erreurs types)

	HRC									
SRC (Image)	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1		0,040255	0,039572	0,038567		0,040432		0,040014		0,036183
2		0,038683	0,033027	0,040957		0,038301		0,042618		0,033956
3		0,039502	0,039111	0,039109		0,042553		0,044151		0,036685
4		0,031762	0,024408	0,036375		0,031371		0,02973		0,042911
5		0,034299	0,044757	0,0407		0,03597		0,033742		0,041272
6		0,040602	0,040035	0,03707		0,043341		0,035289		0,040621
7				0,037894		0,032156		0,038034		0,036946
8				0,036819		0,041563		0,036988		0,037467
9				0,040289		0,040265		0,04015		0,039649
10				0,030283		0,038334		0,037966		0,041339
11	0,034761				0,034838		0,041778			0,041516
12	0,037332				0,036964		0,031253			0,035114
13	0,035205				0,038385		0,038371			0,043687

NOTE – Pour convertir aux écarts types, multiplier par la racine carrée du nombre d'observations, 27.

6 Références bibliographiques

ATIS [octobre 2001] Technical Report T1.TR.72-2001 – Methodological Framework for Specifying Accuracy and Cross-Calibration of Video Quality Metrics, Alliance for Telecommunications Industry Solutions, 1200 G Street, NWn Suite 500, Washington DC.