

UIT-R

Secteur des Radiocommunications de l'UIT

Recommandation UIT-R BT.1867
(03/2010)

**Techniques de mesure objective de la
qualité vidéo perceptuelle pour les
applications de télédiffusion utilisant la
télévision basse définition en
présence d'un signal de référence
à largeur de bande réduite**

Série BT
Service de radiodiffusion télévisuelle



Avant-propos

Le rôle du Secteur des radiocommunications est d'assurer l'utilisation rationnelle, équitable, efficace et économique du spectre radioélectrique par tous les services de radiocommunication, y compris les services par satellite, et de procéder à des études pour toutes les gammes de fréquences, à partir desquelles les Recommandations seront élaborées et adoptées.

Les fonctions réglementaires et politiques du Secteur des radiocommunications sont remplies par les Conférences mondiales et régionales des radiocommunications et par les Assemblées des radiocommunications assistées par les Commissions d'études.

Politique en matière de droits de propriété intellectuelle (IPR)

La politique de l'UIT-R en matière de droits de propriété intellectuelle est décrite dans la «Politique commune de l'UIT-T, l'UIT-R, l'ISO et la CEI en matière de brevets», dont il est question dans l'Annexe 1 de la Résolution UIT-R 1. Les formulaires que les titulaires de brevets doivent utiliser pour soumettre les déclarations de brevet et d'octroi de licence sont accessibles à l'adresse <http://www.itu.int/ITU-R/go/patents/fr>, où l'on trouvera également les Lignes directrices pour la mise en oeuvre de la politique commune en matière de brevets de l'UIT-T, l'UIT-R, l'ISO et la CEI et la base de données en matière de brevets de l'UIT-R.

Séries des Recommandations UIT-R

(Egalement disponible en ligne: <http://www.itu.int/publ/R-REC/fr>)

Séries	Titre
BO	Diffusion par satellite
BR	Enregistrement pour la production, l'archivage et la diffusion; films pour la télévision
BS	Service de radiodiffusion sonore
BT	Service de radiodiffusion télévisuelle
F	Service fixe
M	Services mobile, de radiorepérage et d'amateur y compris les services par satellite associés
P	Propagation des ondes radioélectriques
RA	Radio astronomie
RS	Systèmes de télédétection
S	Service fixe par satellite
SA	Applications spatiales et météorologie
SF	Partage des fréquences et coordination entre les systèmes du service fixe par satellite et du service fixe
SM	Gestion du spectre
SNG	Reportage d'actualités par satellite
TF	Emissions de fréquences étalon et de signaux horaires
V	Vocabulaire et sujets associés

Note: Cette Recommandation UIT-R a été approuvée en anglais aux termes de la procédure détaillée dans la Résolution UIT-R 1.

Publication électronique
Genève, 2011

© UIT 2011

Tous droits réservés. Aucune partie de cette publication ne peut être reproduite, par quelque procédé que ce soit, sans l'accord écrit préalable de l'UIT.

RECOMMANDATION UIT-R BT.1867

Techniques de mesure objective de la qualité vidéo perçue pour les applications de télédiffusion utilisant la télévision basse définition* en présence d'un signal de référence à largeur de bande réduite**

(2010)

Domaine d'application

La présente Recommandation spécifie les méthodes d'évaluation de la qualité vidéo perçue pour les applications de radiodiffusion utilisant la télévision basse définition (LDTV: Low Definition TV) lorsqu'on peut disposer d'un signal de référence réduit (RR), par exemple via une voie de données auxiliaires, un tatouage numérique, des métadonnées, etc.

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 le secteur d'activité concerné;
- b) que la Recommandation UIT-R BT.1683 décrit des méthodes objectives de mesure de la qualité vidéo perçue de la télévision numérique télédiffusée à définition normale en présence d'une image de référence complète;
- c) que la Recommandation UIT-R BT.1833 décrit des systèmes multimédia pour la diffusion d'applications multimédias et d'applications de données destinées à la réception mobile au moyen de récepteurs portatifs;
- d) que la télévision basse définition (LDTV) est de plus en plus utilisée pour la diffusion d'applications multimédias et d'applications de données destinées à une réception mobile;
- e) que la Recommandation UIT-R J.246¹ spécifie les techniques de mesure objective de la qualité perçue de la vidéo pour les applications LDTV en présence d'une référence réduite;
- f) que la mesure objective de la qualité vidéo perçue peut compléter utilement les méthodes d'évaluation subjective,

reconnaissant

- a) que la LDTV est principalement destinée à la visualisation sur de petits écrans tels ceux des récepteurs portatifs et mobiles,

* Par télévision basse définition (LDTV), on entend des résolutions vidéo dont le nombre de pixels est inférieur à ceux définis dans la Recommandation UIT-R BT.601. Une Recommandation UIT-R sur la LDTV est actuellement à l'étude.

** La méthode de mesure avec une référence réduite, pour la mesure objective de la qualité vidéo perçue, permet d'évaluer les performances de systèmes en comparant d'une part les caractéristiques extraites d'un signal de référence ou d'un signal exempt de distorsion appliqué à l'entrée du système avec d'autre part le signal dégradé à la sortie du système.

¹ La Recommandation UIT—T J.246 est disponible à l'adresse <<http://www.itu.int/rec/T-REC-J.246-200808-P/en>>.

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 identifiés au §2 du dispositif de la présente Recommandation;
- 2 d'utiliser le modèle d'évaluation objective de la qualité vidéo perceptuelle spécifiée dans l'Annexe 2 pour les applications de télédiffusion utilisant la télévision à basse définition lorsqu'un signal de référence réduit est disponible, tel que décrit dans l'Annexe 2.

Annexe 1

1 Introduction

La présente Recommandation spécifie des méthodes d'évaluation de la qualité perceptuelle vidéo des applications de télédiffusion utilisant la télévision à basse définition lorsqu'un signal de référence réduit est disponible.

La méthode mesure avec signal de référence réduit peut être utilisée lorsque les caractéristiques extraites du signal vidéo sont directement disponibles au point de mesure comme cela peut être le cas de mesures sur des équipements individuels ou une chaîne dans un laboratoire ou dans un environnement clos. Ces méthodes d'évaluation sont fondées sur le traitement vidéo pour les résolutions VGA, CIF et QCIF.

Les séquences utilisées pour les tests de validation contenaient à la fois plusieurs dégradations de codage et diverses conditions d'erreur de transmission (erreurs binaires, pertes de paquets par exemple). Dans le cas où l'on étudie les distorsions dues au codage des signaux vidéo, le codeur peut utiliser diverses méthodes de compression (MPEG-2, H.264, etc.). Les modèles définis dans la présente Recommandation peuvent être utilisés pour monitorer la qualité des réseaux en place afin de s'assurer de leur aptitude opérationnelle. Les effets visuels des dégradations peuvent se traduire aussi bien par des dégradations spatiales que temporelles (répétitions de trame, sauts de trame, diminution de la fréquence de trame par exemple). Les modèles définis dans la présente Recommandation peuvent également être utilisés pour les tests en laboratoire de systèmes vidéo. Lorsqu'ils sont utilisés pour comparer différents systèmes vidéo, il est recommandé d'utiliser une méthode quantitative (telle celle décrite dans la Recommandation UIT-T J.149) pour déterminer la l'exactitude d'un modèle pour ce contexte particulier. La présente Recommandation est considérée comme étant applicable aux services fournis à 4 Mbit/s ou moins à destination de récepteurs mobiles. Les conditions suivantes sont été prises pour hypothèse pour les tests de validation pour chaque résolution:

- QCIF (quarter common intermediate format (176 × 144 pixels)): 16 kbit/s à 320 kbit/s.
- CIF (common intermediate format (352 × 288 pixels)): 64 kbit/s – 2 Mbit/s.
- VGA (video graphics array (640 × 480 pixels)): 128 kbit/s – 6 Mbit/s.

TABLEAU 1
Facteurs utilisés pour l'évaluation des modèles

Facteurs de test
Erreurs de transmission avec perte de paquets
Résolution vidéo QCIF, CIF et VGA
Débits binaires vidéo – QCIF: 16 kbit/s à 320 kbit/s – CIF: 64 kbit/s – 2 Mbit/s – VGA: 128 kbit/s – 4 Mbit/s
Erreurs temporelles (pause avec saut) de 2 s au maximum
Fréquences de trame vidéo comprises entre 5 et 30 trames/s
Schémas de codage
H.264/AVC (MPEG-4 Partie 10), MPEG-4 Partie 2, et trois autres schémas de codage propriétaires. (Voir la Note 1.)
Applications
Monitoring de la qualité en temps réel, en service au niveau de la source
Monitoring de la qualité à la destination distante lorsque des canaux latéraux sont disponibles pour des caractéristiques extraites de séquences de vidéo source
Mesure de la qualité pour le monitoring de systèmes de stockage ou de transmission vidéo qui utilisent des techniques de compression et de décompression vidéo ou une concaténation de ces deux techniques
Test en laboratoire de systèmes vidéo

NOTE 1 – Les tests de validation des modèles comportaient des séquences codées et portaient sur 15 codecs vidéo différents. Les 5 codecs indiqués dans le Tableau 1 étaient généralement utilisés pour coder des séquences de test et tout modèle recommandé pouvait être considéré comme approprié pour l'évaluation de ces codecs. Outre ces cinq codecs, une plus faible proportion de séquences de test a été créée en utilisant les codecs suivants: H.261, H.263, H.263+², JPEG-2000, MPEG-1, MPEG-2, H.264 SVC et autres systèmes propriétaires. On notera que certains de ces codecs ont uniquement été utilisés pour les résolutions CIF et QCIF car ils devraient être utilisés dans la pratique principalement pour ces résolutions.

Avant d'appliquer un modèle aux séquences codées au moyen de ces codecs, l'utilisateur devrait étudier soigneusement leurs performances prédictives afin de déterminer si les performances prédictives du modèle sont acceptables.

2 Applications

Les applications des modèles d'évaluation décrits dans la présente Recommandation sont entre autres les suivantes:

- 1) évaluation, spécification et tests de réception des codecs, compatibles avec l'exactitude limitée tel que décrit ci-dessous;
- 2) monitoring de la qualité en temps réel, en service;
- 3) monitoring de la qualité à la destination distante lorsqu'on dispose de canaux latéraux pour l'extraction des caractéristiques à partir de séquences vidéo source;

² La H.263+ est une configuration particulière de la H.263 (1998).

- 4) mesure de la qualité en vue du monitoring de systèmes de stockage ou de transmission qui utilisent des techniques de compression et de décompression vidéo, par application unique ou concaténation de ces techniques;
- 5) test en laboratoire de systèmes vidéo.

3 Limitations

Les modèles d'évaluation décrits dans la présente Recommandation ne remplacent en aucune façon les tests subjectifs. Les valeurs de corrélation entre deux tests subjectifs conçus et exécutés avec soin (par exemple dans deux laboratoires différents) se situent normalement dans la fourchette 0,95-0,98. Si la présente Recommandation est utilisée pour comparer des systèmes vidéo (comparaison de deux codecs par exemple), il est recommandé d'utiliser une méthode quantitative (telle celle décrite dans la Recommandation UIT-T J.149) pour déterminer l'exactitude du modèle dans ce contexte particulier.

Les modèles décrits dans la présente Recommandation ont été validés pour des mesures sur des séquences vidéo présentant un gel de trame pouvant durer jusqu'à 2 s.

Les modèles décrits dans la présente Recommandation n'ont pas été validés pour des mesures sur des séquences vidéo présentant un retard augmentant régulièrement (par exemple une vidéo dans laquelle les trames manquantes ne sont pas éliminées après un gel de trame).

Il convient de noter que, dans le cas de nouvelles technologies de codage et de transmission introduisant des défauts (artéfacts) qui n'ont pas été incluses dans la présente évaluation, les modèles de mesure objective pourraient donner des résultats erronés. Dans ce cas, une évaluation subjective est indispensable.

4 Description des modèles

Les modèles suivants sont décrits dans l'Annexe 2:

Modèle A (Annexe 2) – Proposition du Groupe d'experts sur la qualité vidéo (VQEG: Vidéo Quality Expert Group), Université de Yonsei (République de Corée).

Appendice 1 à l'Annexe 1

Résultats des travaux du VQEG

Les études relatives aux mesures de la qualité vidéo perceptuelle sont effectuées par un groupe officieux appelé VQEG (Video Quality Expert Group: Groupe d'experts sur la qualité vidéo) qui rend compte de ses travaux aux Commissions d'étude 9 et 12 de l'UIT-T et à la Commission d'étude 9 de l'UIT-R. La Phase 1 des tests multimédias récemment menée à bien, ont permis d'évaluer les performances des algorithmes de mesure de la qualité vidéo perceptuelle en présence d'un signal de référence réduit pour les formats QCIF, CIF et VGA.

Sur la base des présents résultats, l'UIT-R peut actuellement recommander la méthode suivante:

Modèle A (Annexe 2) – Initiateur: Groupe VQEG, Université de Yonsei, Corée (République de).

Les Tableaux 2, 3 et 4 contiennent des informations détaillées sur les performances du modèle pour la Phase 1 des tests multimédias du VQEG.

TABLEAU 2

Résolution VGA: description pour information des performances du modèle dans la Phase 1 des tests multimédias du VQEG: moyenne sur 13 tests subjectifs

Statistique	Yonsei RR10k	Yonsei RR64k	Yonsei RR128k	PSNR ⁽¹⁾
Corrélation	0,803	0,803	0,803	0,713
RMSE ⁽²⁾	0,599	0,599	0,598	0,714
Taux d'aberration	0,556	0,553	0,552	0,615

⁽¹⁾ PSNR: peak signal-to-noise ratio (valeur de crête du rapport signal à bruit).

⁽²⁾ RMSE: root mean square error (erreur quadratique moyenne).

TABLEAU 3

Résolution CIF: description pour information des performances du modèle dans la Phase 1 des tests multimédias du VQEG: moyenne sur 14 tests subjectifs

Statistique	Yonsei RR10k	Yonsei RR64k	PSNR
Corrélation	0,780	0,782	0,656
RMSE	0,593	0,590	0,720
Taux d'aberration	0,519	0,511	0,632

TABLEAU 4

Résolution QCIF: description pour information des performances du modèle dans la Phase 1 des tests multimédias du VQEG: moyenne sur 14 tests subjectifs

Statistique	Yonsei RR1k	Yonsei RR10k	PSNR
Corrélation	0,771	0,791	0,662
RMSE	0,604	0,578	0,721
Taux d'aberration	0,505	0,486	0,596

Les valeurs de corrélation moyennes des principales analyses des modèles VGA à référence réduite (RR VGA) étaient toutes de 0,80 et le rapport PSNR était de 0,71. Les valeurs de corrélation de modèles individuels pour certaines expériences atteignaient 0,93. Pour les modèles VGA à référence réduite, les erreurs RSME étaient toutes égales à 0,60 et le PSNR était de 0,71. Pour les modèles VGA à référence réduite, le taux d'aberration moyen était compris entre 0,55 et 0,56 et le PSNR était de 0,62. Pour 7 des 13 expérimentations, tous les modèles proposés donnaient des résultats statistiquement meilleurs comparativement au PSNR. Sur la base de chaque mesure, chaque modèle VGA à référence réduite se classait parmi les modèles donnant les meilleures performances pour le nombre de fois suivant:

Statistique	Yonsei RR10k	Yonsei RR64k	Yonsei RR128k	PSNR
Corrélation	13	13	13	7
RMSE	13	13	13	6
Taux d'aberration	13	13	13	10

Les valeurs de corrélation moyennes des principales analyses des modèles CIF à référence réduite étaient de 0,78 et le rapport PNSR était de 0,66. Les valeurs de corrélation des modèles individuels pour certaines expérimentations atteignaient 0,90. Pour les modèles CIF à référence réduite, les erreurs RSME moyenne étaient toutes égales à 0,93 et le PNSR était de 0,63. Pour les modèles CIF à référence réduite, le taux d'aberration moyen était compris étaient de 0,51 et 0,52 et le PNSR était de 0,63. Pour 10 des 14 expérimentations, tous les modèles proposés donnaient des résultats statistiquement meilleurs comparativement au PNSR. Sur la base de chaque mesure, chaque modèle CIF à référence réduite se classait parmi les modèles donnant les meilleures performances pour le nombre de fois suivant:

Statistique	Yonsei RR 10k	Yonsei RR64k	PSNR
Corrélation	14	14	5
RMSE	14	14	4
Taux d'aberration	14	14	5

Les valeurs de corrélation moyennes des principales analyses des modèles QCIF à référence réduite étaient de 0,77 et 0,79 et le rapport PNSR était de 0,66. Les valeurs de corrélation des modèles individuels pour certaines expérimentations atteignaient 0,89. Pour les modèles QCIF à référence réduite, les erreurs RSME moyennes étaient égales à 0,49 et 0,51 et le PNSR était égal à 0,60. Pour les modèles QCIF à référence réduite, les taux d'aberration moyen étaient de 0,49 et 0,5 et le PNSR était de 0,60. Pour 9 des 14 expérimentations, tous les modèles proposés donnaient des résultats statistiquement meilleurs comparativement au PNSR. Sur la base de chaque mesure, chaque modèle QCIF à référence réduite se classait parmi les modèles donnant les meilleures performances pour le nombre de fois suivant:

Statistique	Yonsei RR1k	Yonsei RR10k	PSNR
Corrélation	14	14	5
RMSE	14	14	4
Taux d'aberration	12	13	4

Annexe 2

Méthodes à référence réduite pour le Modèle A^{***}

TABLE DES MATIÈRES

	<i>Page</i>
1 Introduction	7
2 Rapport EPSNR pour les modèles à référence réduite	8
2.1 Rapport PNSR pour les contours	8
2.2 Choix des caractéristiques à partir des séquences vidéo source	13
2.3 Alignement spatial/temporel et réglage du gain/décalage	16
2.4 Calcul de l'EPSNR et post-traitement	19
2.5 Largeur de bande optimale du canal latéral	20
Appendice	20

1 Introduction

Bien que le rapport PSNR soit largement utilisé pour la mesure de la qualité vidéo objective, on signale qu'il ne représente pas bien la qualité vidéo perceptuelle. En analysant comment l'être humain perçoit la qualité vidéo, on constate que le système visuel humain est sensible aux dégradations sur les contours. En d'autres termes, lorsque les pixels sur les contours d'une séquence vidéo sont flous, les évaluateurs tendent à donner des notes faibles à la séquence vidéo même si le rapport PSNR est élevé. A partir de cette observation, on a mis au point des modèles à référence réduite qui mesurent principalement les dégradations sur les contours.

La Fig. 1 illustre la façon dont le modèle à référence réduite fonctionne. Les caractéristiques qui seront utilisées pour mesurer la qualité vidéo au point de mesure sont extraites de la séquence vidéo source et transmises. Le Tableau 5 montre les largeurs de bande d'un canal latéral pour les caractéristiques, qui ont été testées lors du test VQEGMM.

^{***} Ce modèle est identique à celui spécifié dans l'Annexe A de la Recommandation UIT-T J.246.

FIGURE 1

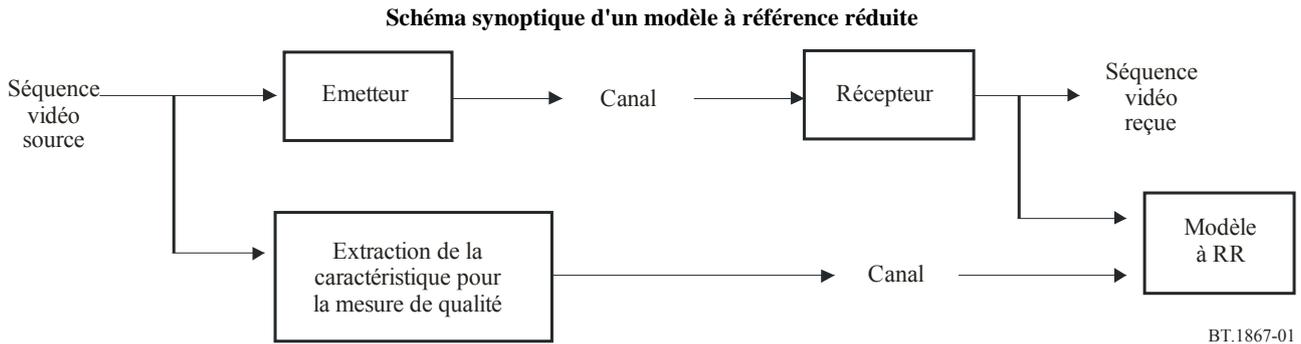


TABLEAU 5

Largeurs de bande du canal latéral

Format vidéo	Largeurs de bande testées
QCIF	1 kbit/s, 10 kbit/s
CIF	10 kbit/s, 64 kbit/s
VGA	10 kbit/s, 64 kbit/s, 128 kbit/s

2 Rapport EPSNR pour les modèles à référence réduite

2.1 Rapport PNSR pour les contours

Les modèles à référence réduite (RR) permettent de mesurer principalement les dégradations sur les contours. Dans ces modèles, un algorithme de détection des contours est d'abord appliqué à la séquence vidéo source pour localiser les pixels des contours. Puis, la dégradation de ces pixels des contours est mesurée en calculant l'erreur quadratique moyenne. A partir de cette erreur quadratique moyenne, on calcule le rapport EPNSR (EPNSR: Edge PNSR) (PNSR des contours)).

On peut utiliser tout algorithme de détection des contours, bien que ces algorithmes puissent donner des différences mineures dans les résultats. Par exemple, on peut utiliser un opérateur gradient pour localiser les contours. Un certain nombre d'opérateurs gradient ont été proposés. Dans de nombreux algorithmes de détection des contours, l'image en gradients horizontaux $g_{horizontal}(m,n)$ et l'image en gradients verticaux $g_{vertical}(m,n)$ ont d'abord été calculées au moyen d'opérateurs gradients. On peut alors calculer la valeur de l'image en gradients de valeurs $g(m,n)$ comme suit:

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

Enfin, on applique une opération de seuillage à l'image en gradient de valeurs $g(m,n)$ pour trouver les pixels des contours. En d'autres termes, les pixels dont les gradients de valeur sont supérieurs à une valeur seuil sont considérés comme étant des pixels de contour.

Les Fig. 2 à 6 illustrent la procédure. La Fig. 2 montre une image source. La Fig. 3 montre une image en gradients horizontaux $g_{horizontal}(m,n)$, qui est obtenue en appliquant un opérateur gradient horizontal à l'image source de la Fig. 2. La Fig. 4 montre une image en gradients verticaux $g_{vertical}(m,n)$, qui est obtenue en appliquant un opérateur gradient vertical à l'image source de la Fig. 2. La Fig. 5 montre l'image en gradients de valeurs (image des contours) et la Fig. 6 montre une

image binaire des contours obtenue en appliquant un seuillage à l'image en gradients de valeurs de la Fig. 5.

FIGURE 2

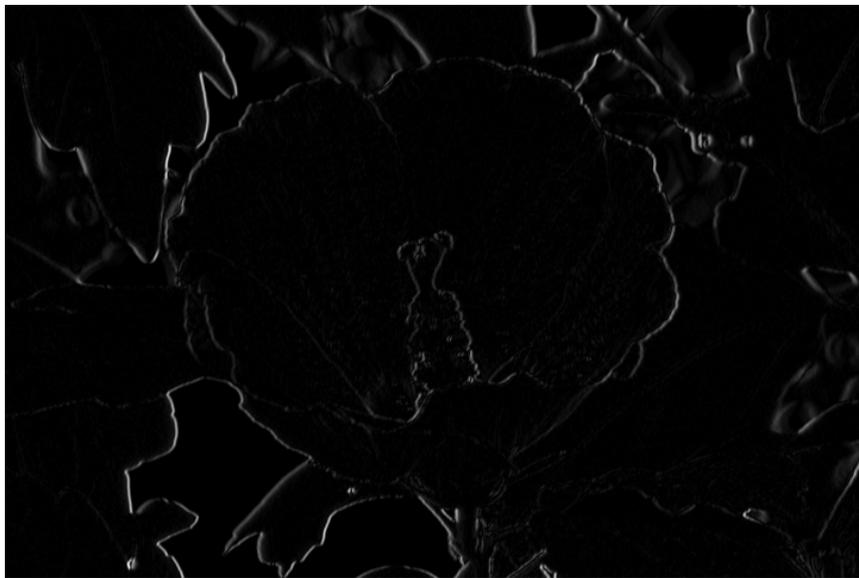
Image source (image originale)



BT.1867-02

FIGURE 3

Image en gradients horizontaux obtenue en appliquant un opérateur gradient horizontal à l'image source de la Fig. 2



BT.1867-03

FIGURE 4

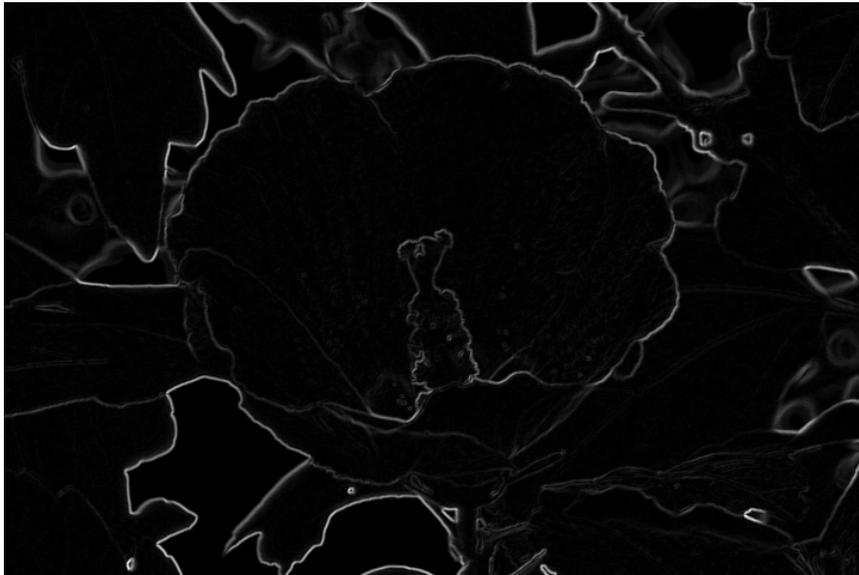
Image en gradients verticaux obtenue en appliquant un opérateur gradient vertical à l'image source de la Fig. 2



BT.1867-04

FIGURE 5

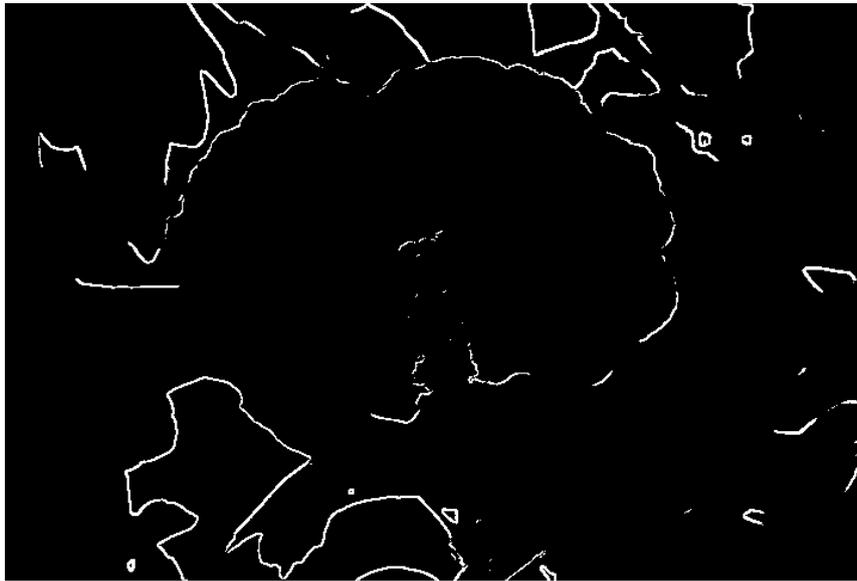
Image en gradients de valeur



BT.1867-05

FIGURE 6

Image à contours binaire (image de masque) obtenue en appliquant un seuillage à l'image en gradients de valeurs de la Fig. 5

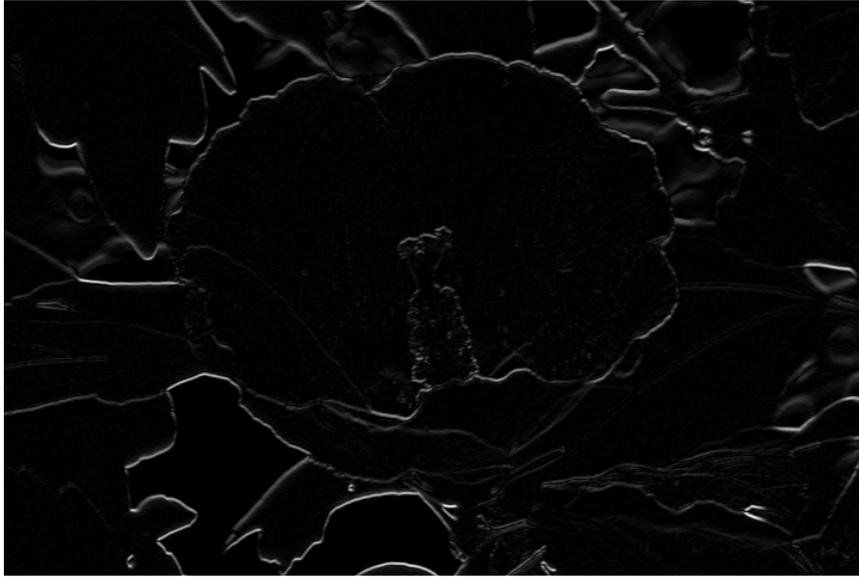


BT.1867-06

On peut également utiliser une procédure modifiée pour localiser les pixels de contour. Par exemple, on peut tout d'abord appliquer un opérateur gradient vertical à l'image source, ce qui donne une image en gradients verticaux. On applique ensuite un opérateur gradient horizontal à l'image en gradients verticaux qui donne une image en gradients successifs (image en gradients horizontaux et en gradients verticaux). On peut alors appliquer un seuillage à l'image en gradients successifs pour trouver les pixels des contours. En d'autres termes, les pixels de l'image en gradients successifs modifiée qui dépassent une valeur seuil sont considérés comme pixels de contour. Les Fig. 7 à 9 illustrent la procédure modifiée. La Fig. 7 montre une image en gradients verticaux $g_{vertical}(m,n)$, laquelle est obtenue par application d'un opérateur gradient vertical à l'image source de la Fig. 2. La Fig. 8 montre une image en gradients successifs modifiée (image en gradients horizontaux et en gradients verticaux), laquelle est obtenue par application d'un opérateur gradient horizontal à l'image en gradients verticaux de la Fig. 7. La Fig. 9 montre l'image binaire des contours (image masque) obtenue en appliquant un seuillage à l'image en gradients successifs modifiée de la Fig. 8.

FIGURE 7

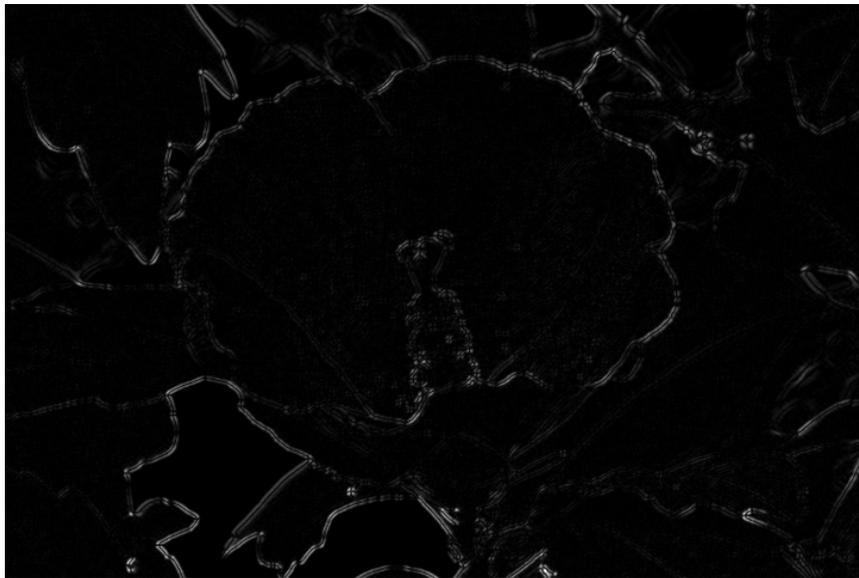
Image en gradients verticaux obtenue en appliquant un opérateur gradient vertical à l'image source de la Fig. 2



BT.1867-07

FIGURE 8

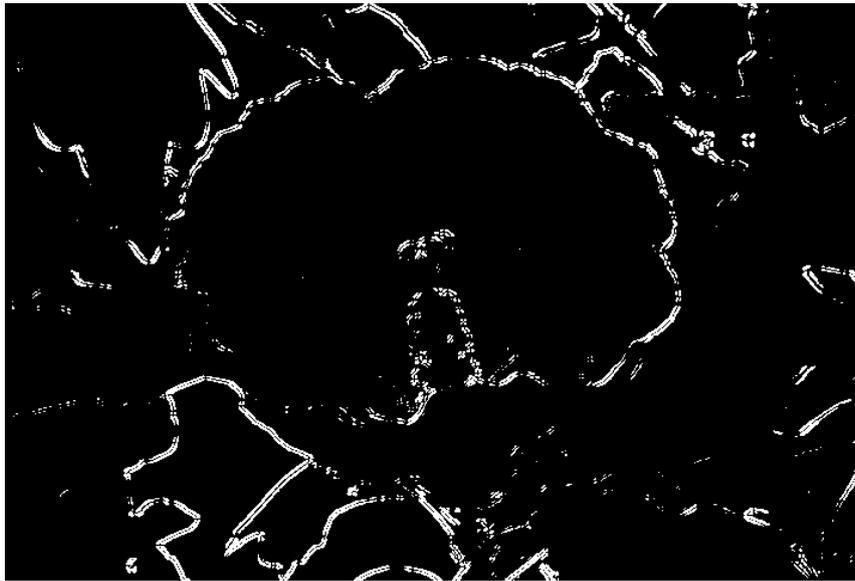
Image en gradients successifs (image du gradient horizontal et du gradient vertical) obtenue en appliquant un opérateur gradient horizontal à l'image du gradient horizontal de la Fig. 7



BT.1867-08

FIGURE 9

Image à contours binaire (image de masque) obtenue en appliquant un seuillage à l'image en gradients successifs modifiée de la Fig. 8



BT.1867-09

Il est à noter que les deux méthodes peuvent être considérées comme étant des algorithmes de détection des contours. On peut choisir tout algorithme de détection des contours selon la nature des séquences vidéo et les algorithmes de compression. Toutefois, certaines méthodes peuvent donner de meilleurs résultats que d'autres.

Ainsi, dans le modèle, on applique tout d'abord un opérateur de détection des contours qui donne des images des contours (Fig. 5 et 8). Puis, on produit une image de masque (image binaire des contours) en appliquant un seuillage à l'image des contours (Fig. 6 et 9). 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 au seuil sont mis à une valeur non nulle. Les Fig. 6 et 9 montrent certaines images de masque. Etant donné qu'une séquence vidéo peut être considérée comme une séquence de trames ou d'images, la procédure précitée peut être appliquée à chaque trame ou image des séquences vidéo. Puisque le modèle peut être utilisé pour les séquences vidéo à trames ou à images, le terme «image» sera utilisé pour désigner indifféremment une image ou une trame.

2.2 Choix des caractéristiques à partir des séquences vidéo source

Etant donné que le modèle est un modèle à référence réduite (RR), un ensemble de caractéristiques peut être extrait de chaque image d'une séquence vidéo source. Dans le modèle EPSNR RR, un certain nombre de pixels de contour est choisi sur chaque image. Les positions et les valeurs des pixels de contour sont ensuite codées et transmises. Toutefois, pour certaines séquences vidéo, le nombre de pixels de contour peut être très faible lorsqu'on utilise une valeur seuil fixe. Dans le scénario le plus défavorable, ce nombre peut être nul (images supprimées ou images à cadence très faible). Afin de traiter du problème, si le nombre de pixels de contour d'une image est inférieur à une valeur donnée, l'utilisateur peut abaisser le seuil jusqu'à ce que le nombre de pixels de contour soit supérieur à une certaine valeur. On peut aussi choisir les pixels de contour qui correspondent aux plus grandes valeurs de l'image en gradients horizontaux et en gradients verticaux. Lorsqu'il n'y a pas de pixels d'image (image supprimée par exemple) dans une trame, on peut choisir aléatoirement le nombre de pixels requis ou sauter la trame. Par exemple, si 10 pixels de contour doivent être extraits de chaque trame, on peut trier les pixels de l'image en gradients horizontaux et en gradients verticaux selon leurs valeurs et retenir les 10 plus grandes valeurs. Toutefois, cette

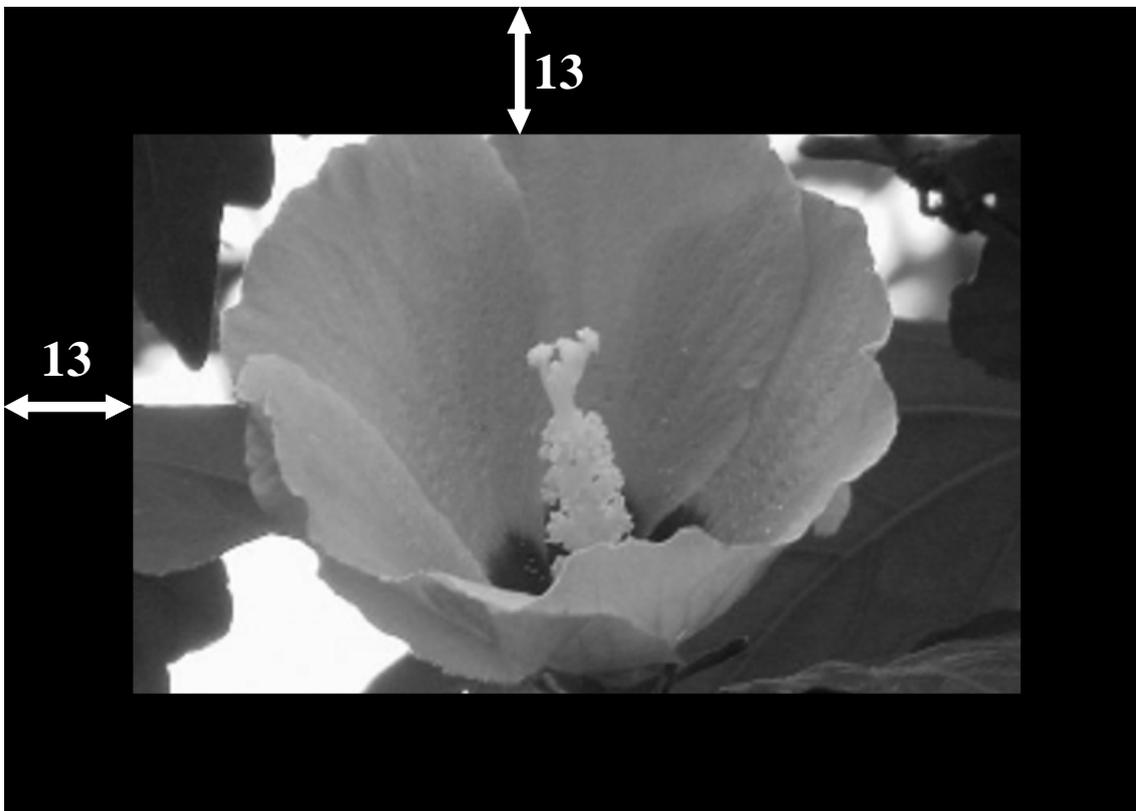
procédure peut produire plusieurs pixels de contour sur des positions identiques. Pour traiter ce problème, on peut d'abord choisir plusieurs fois le nombre souhaité de pixels de l'image en gradients horizontaux et verticaux et choisir ensuite aléatoirement le nombre souhaité de pixels parmi les pixels de l'image en gradients horizontaux et en gradients verticaux. Dans les modèles testés lors du test multimédia VQEG, le nombre de pixels souhaité est choisi aléatoirement parmi un grand pool de pixels de contour. Le pool des pixels de contour est obtenu en appliquant une opération de seuillage de l'image en gradients.

Dans les modèles EPSNR RR, les positions et les valeurs des pixels de contour sont codés. Il est à noter que pendant le processus de codage, on peut appliquer un recadrage. Afin d'éviter le choix des pixels de contour dans les zones recadrées, le modèle choisit les pixels de contour dans la zone centrale (Fig. 10). Le Tableau 6 montre les dimensions après recadrage. Ce tableau montre aussi le nombre de bits nécessaire pour coder la position et la valeur de pixel d'un pixel de contour.

TABLEAU 6
Bits nécessaires par pixel de contour

Format vidéo	Dimensions	Dimensions après recadrage	Bits de position	Bits de valeur de pixel	Nbre total de bits par pixel
QCIF	176 × 144	168 × 136	15	8	23
CIF	352 × 288	338 × 274	17	8	25
VGA	640 × 480	614 × 454	19	8	27

FIGURE 10
Exemple de recadrage (VGA) et zone centrale



Le modèle choisit les pixels de contour dans chaque trame conformément à la largeur de bande autorisée (Tableau 5). Les Tableaux 7 et 8 montrent le nombre de pixels de contour par trame qui peut être transmis pour les largeurs de bande testées.

TABLEAU 7

Nombre de pixels de contour par trame (30 trames/s)

Format vidéo	1 kbit/s	10 kbit/s	64 kbit/s	128 kbit/s
QCIF	1	14		
CIF		13	85	
VGA		12	79	158

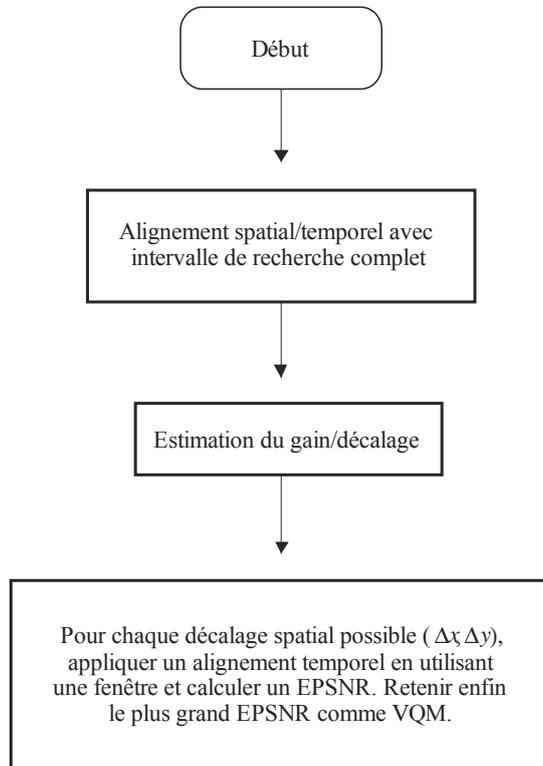
TABLEAU 8

Nombre de pixels de contour par trame (25 trames/s)

Format vidéo	1 kbit/s	10 kbit/s	64 kbit/s	128 kbit/s
QCIF	1	17		
CIF		16	102	
VGA		14	94	189

FIGURE 11

Organigramme du modèle



2.3 Alignement spatial/temporel et réglage du gain/décalage

Avant de calculer la différence entre les pixels de contour de la séquence vidéo source et ceux de la séquence vidéo traitée qui est la séquence vidéo reçue sur le récepteur, le modèle procède d'abord à un alignement spatial/temporel et à un réglage du gain/décalage. On applique d'abord un algorithme de recherche complet afin de trouver les décalages spatiaux et temporels globaux ainsi que les valeurs de gain et de décalage (Fig. 11). Pour chaque décalage spatial ($\Delta x, \Delta y$), un alignement temporel est donc effectué et l'EPSNR calculé. Enfin, la plus grande valeur d'EPSNR est retenue comme mesure de la qualité vidéo (VQM: Video Quality Metric).

Au point de monitoring, la séquence vidéo traitée doit être ajustée avec les pixels de contour extraits de la séquence vidéo source. Toutefois, si la largeur de bande des canaux latéraux est petite, seuls quelques pixels de contour de la séquence vidéo source sont disponibles (Fig. 12). En conséquence, l'alignement temporel peut être faussé si l'alignement temporel est effectué en utilisant une seule trame (Fig. 13). Pour résoudre ce problème, le modèle fait appel à une fenêtre pour l'alignement temporel. Au lieu d'utiliser une seule trame de la séquence vidéo traitée, le modèle construit une fenêtre qui est composée de trames adjacentes pour trouver le décalage temporel optimal. La Fig. 14 illustre la procédure. L'erreur quadratique moyenne dans la fenêtre est calculée comme suit:

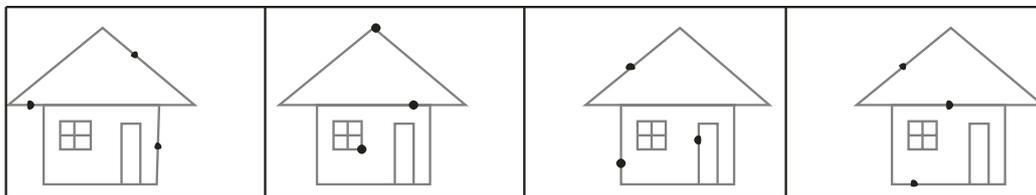
$$MSE_{window} = \frac{1}{N_{win}} \sum (E_{SRC}(i) - E_{PVS}(i))^2$$

où MSE_{window} est l'erreur quadratique moyenne dans la fenêtre, $E_{SRC}(i)$ est un pixel de contour à l'intérieur de la fenêtre qui a un pixel correspondant dans la séquence vidéo traitée, $E_{PVS}(i)$ est un pixel de la séquence vidéo traitée correspondant au pixel de contour, et N_{win} est le nombre total de pixels de contour utilisé pour calculer MSE_{window} . L'erreur quadratique moyenne est utilisée comme différence entre une trame de la séquence vidéo traitée et la trame de la séquence vidéo source correspondante.

La longueur de la fenêtre vidéo peut être déterminée en prenant en considération de la nature de la séquence vidéo traitée. Pour une application type, une fenêtre correspondant à deux secondes est recommandée. On peut aussi appliquer différentes longueurs de fenêtre et utiliser la meilleure de celles-ci, c'est à dire celle qui donne la plus faible erreur quadratique moyenne.

FIGURE 12

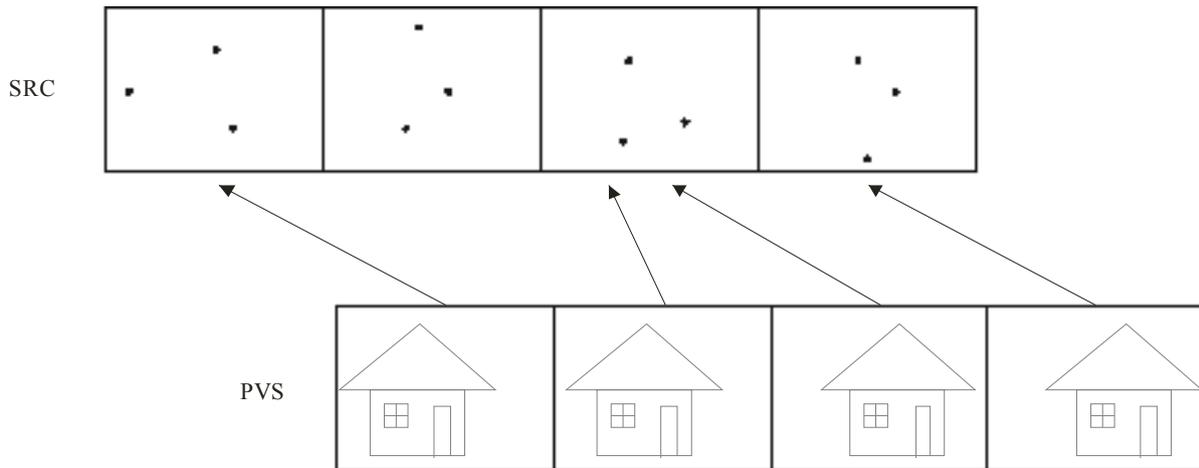
Choix des pixels de contour dans une séquence vidéo source



BT.1867-12

FIGURE 13

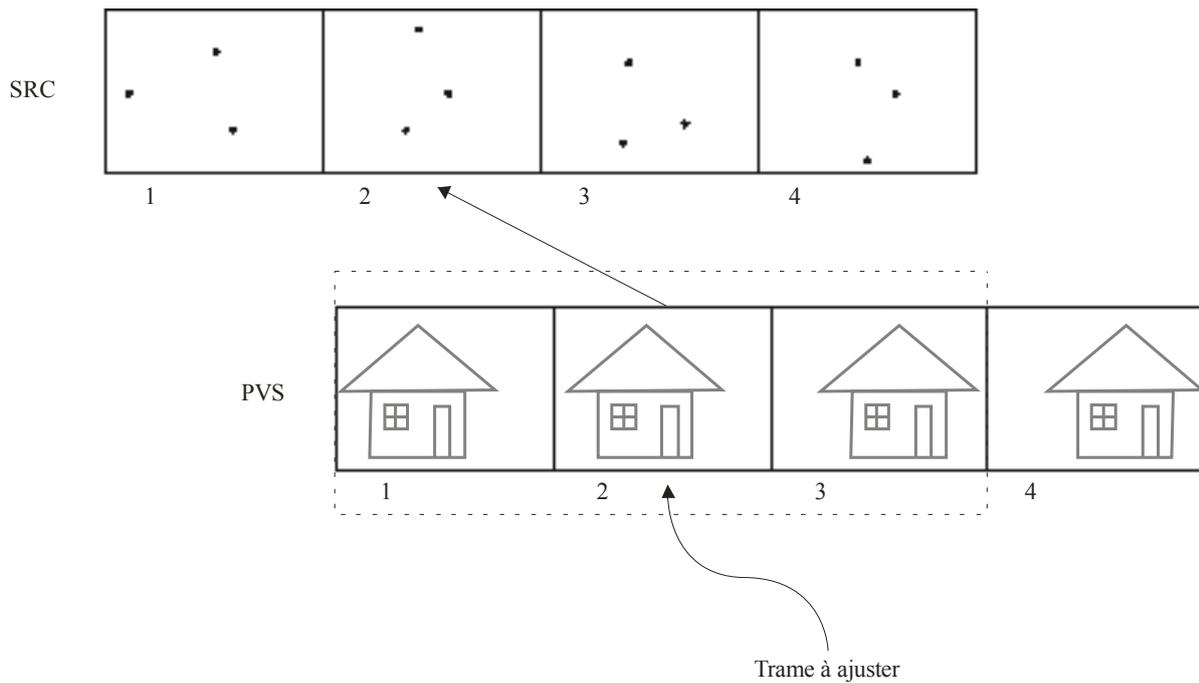
Ajustage de la séquence vidéo traitée sur les pixels de contour de la séquence vidéo source



BT.1867-13

FIGURE 14

Ajustage de la séquence vidéo traitée sur les pixels de contour en utilisant une fenêtre



BT.1867-14

Lorsque la séquence vidéo source est codée avec des taux de compression élevés, le codeur doit réduire le nombre de trames par seconde et la séquence vidéo traitée comporte des trames répétées (Fig. 15). Dans la Fig. 15, la séquence vidéo traitée ne comporte pas de trames correspondant à certaines trames de la séquence vidéo source (2ème, 4ème, 6ème et 8ème trame). Dans ce cas, le modèle n'utilise pas les trames répétées dans le calcul de l'erreur quadratique moyenne. En d'autres termes, le modèle effectue un alignement temporel en utilisant la première trame (trame valide) de chaque bloc répété. Ainsi, seules trois trames (3ème, 5ème et 6ème trame) dans la fenêtre sont utilisées pour l'alignement temporel.

FIGURE 15

Exemple de trames répétées

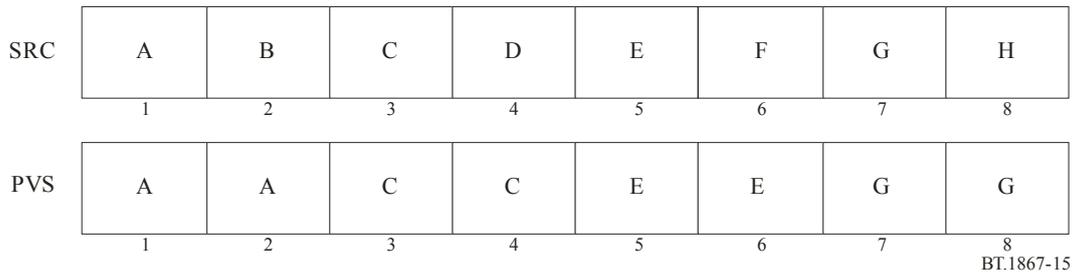
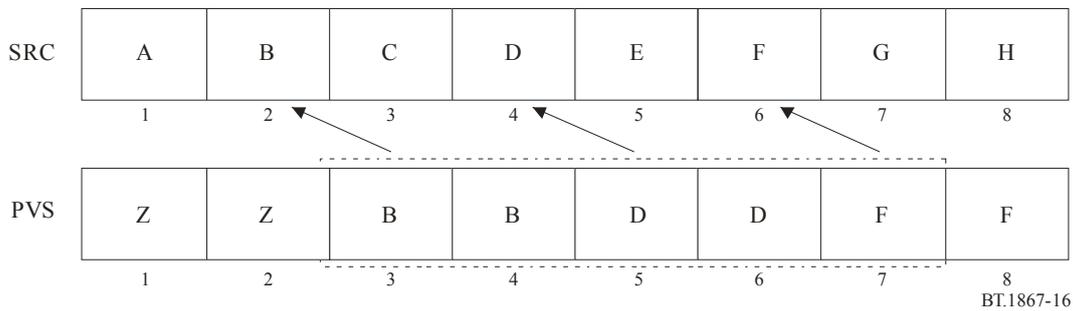


FIGURE 16

Traitement de trames répétées



Il est possible d'avoir une séquence vidéo traitée avec une répétition irrégulière de trames, ce qui peut aboutir à ce que la méthode d'alignement temporel utilisant une fenêtre, donne des résultats erronés. Pour résoudre ce problème, il est possible d'ajuster localement chaque trame de la fenêtre dans les limites d'une certaine valeur (± 1 par exemple) comme le montre la Fig. 18 après l'alignement temporel au moyen d'une fenêtre. Puis, l'ajustement local qui donne la plus petite erreur quadratique moyenne est utilisé pour calculer l'EPSNR.

FIGURE 17

Fenêtre de longueurs diverses

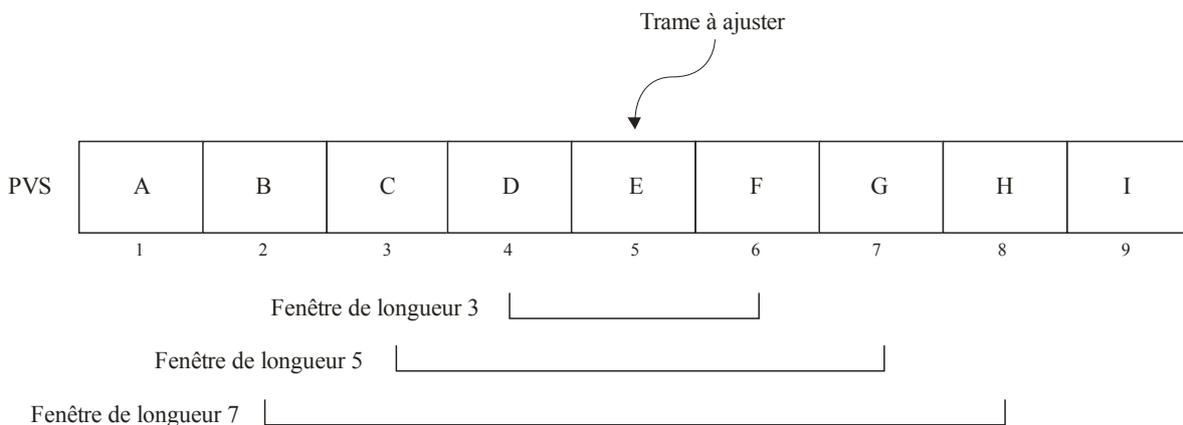
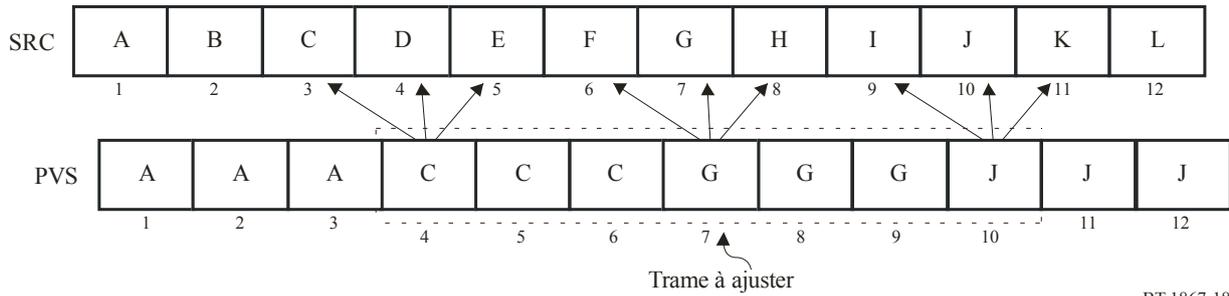


FIGURE 18

Ajustement local pour l'alignement temporel en utilisant une fenêtre



BT.1867-18

2.4 Calcul de l'EPSNR et post-traitement

Après l'alignement temporel, on calcule la moyenne des différences entre les pixels de contour de la séquence vidéo source et les pixels correspondants de la séquence vidéo traitée, différence qui peut être interprétée comme étant l'erreur quadratique moyenne des contours de la séquence vidéo traitée (MSE_{edge}). On calcule enfin l'EPSNR (PSNR des contours) comme suit:

$$EPSNR = 10 \log_{10} \left(\frac{P^2}{MSE_{edge}} \right)$$

où

p est la valeur crête de l'image.

Dans le codage vidéo multimédia, il peut y avoir une répétition de trames causée par des cadences de trame réduites et un gel de trame dus à des erreurs de transmission, ce qui entraîne une dégradation de la qualité vidéo perçue. Afin de traiter de cet effet, le modèle applique l'ajustage suivant avant de calculer l'EPSNR:

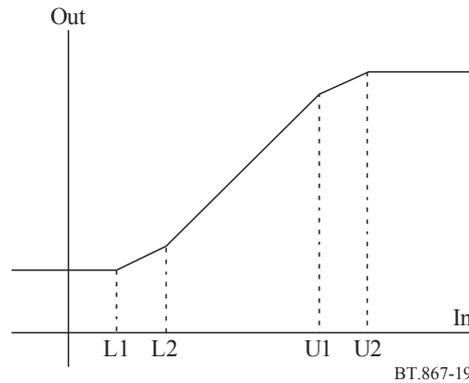
$$MSE_{frozen_frame_considered} = MSE_{edge} \times \frac{K \times N_{total_frame}}{N_{total_frame} - N_{total_frozen_frame}}$$

où $MSE_{frozen_frame_considered}$ est l'erreur quadratique moyenne qui tient compte des trames répétées et gelées, N_{total_frame} est le nombre total de trames, $N_{total_frozen_frame}$, est le nombre total de trames gelées, K est une constante. Dans le modèle testé lors du test multimédia VQEG, K a été fixé à 1.

Lorsque l'EPSNR est supérieur à une certaine valeur, la qualité vidéo perçue devient saturée. Dans ce cas, il est possible de fixer la limite supérieure de l'EPSNR. De plus, lorsqu'une relation linéaire entre l'EPSNR et la note d'opinion moyenne différentielle (DMOS: Difference Mean Opinion Score) est souhaitable, on peut appliquer une fonction linéaire par paliers comme le montre la Fig. 19. Dans le modèle testé lors du test VQEG multimédia, seule la limite supérieure a été fixée à 50 étant donné qu'on a utilisé un ajustement polynomial.

FIGURE 19

Fonction linéaire par paliers pour la relation linéaire entre l'EPSNR et la DMOS



2.5 Largeur de bande optimale du canal latéral

L'Appendice contient une comparaison des performances lorsque la largeur de bande du canal latéral augmente. Pour le format QCIF, on observe que les coefficients de corrélation sont pour la plupart saturés à 10 kbit/s environ. Puis, l'augmentation de largeur de bande apporte une amélioration de 1% environ. Pour le format CIF, on observe que les coefficients de corrélation sont pour la plupart saturés à 15 kbit/s environ. Puis, l'augmentation de la largeur de bande apporte une amélioration de 0,5% environ. Pour le format VGA, on observe que les coefficients de corrélation sont pour la plupart saturés à 30 kbit/s environ. Puis, l'augmentation de la largeur de bande apporte une amélioration de 0,5% environ.

Les modèles d'EPSNR à référence réduite pour la mesure objective de la qualité vidéo sont fondés sur la dégradation des contours. Les modèles peuvent être mis en œuvre en temps réel avec utilisation modérée de la puissance de calcul. Les modèles sont bien adaptés aux applications qui nécessitent un monitoring en temps réel de la qualité vidéo lorsque des canaux latéraux sont disponibles.

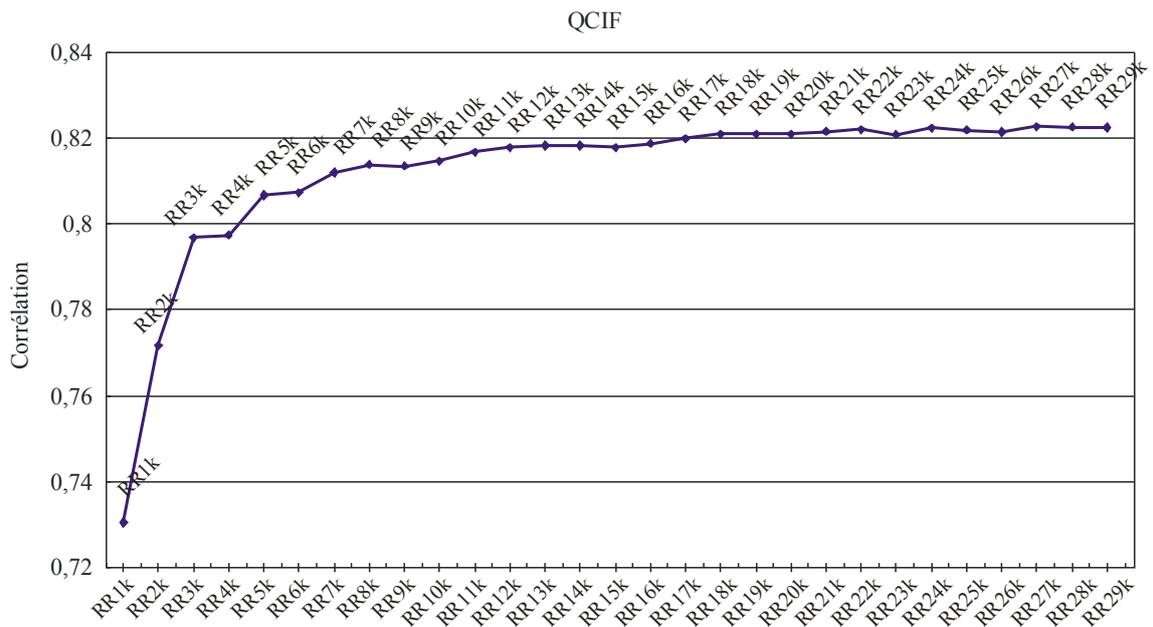
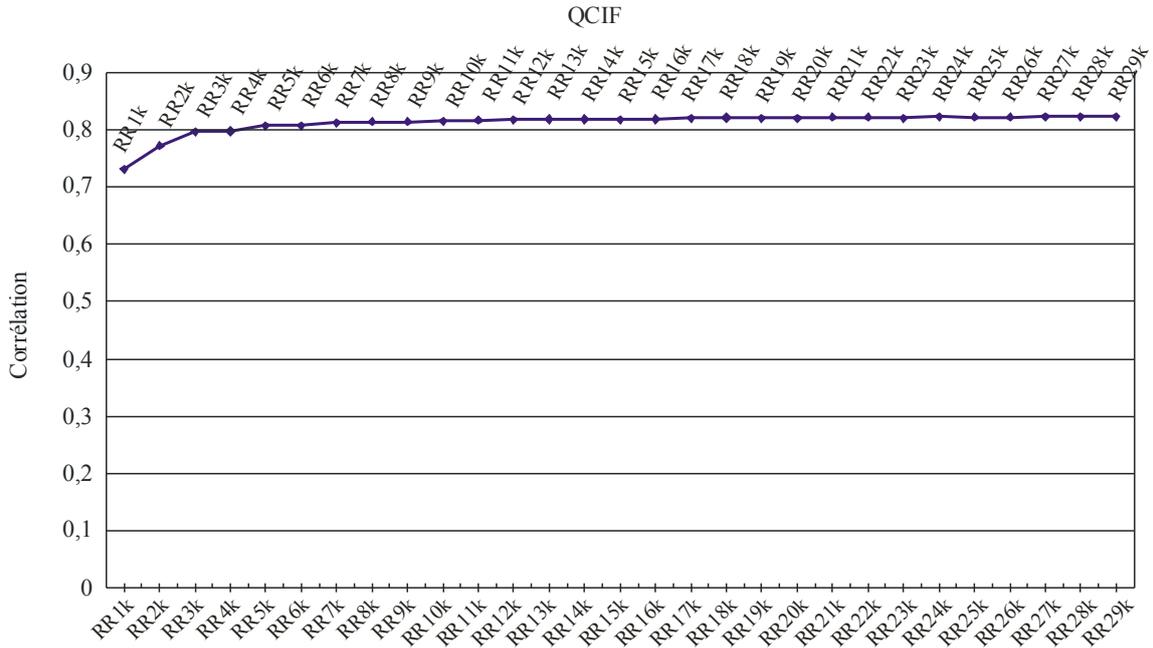
Appendice 1 de l'Annexe 2

1 Largeurs de bande optimales des canaux latéraux

La Fig. 20 montre les coefficients de corrélation pour différentes largeurs de bande des canaux latéraux pour le format QCIF. On peut constater que les coefficients de corrélation sont pour la plupart saturés à 10 kbit/s environ. Puis, l'augmentation de la bande passante produit une amélioration de 1% environ.

FIGURE 20

Amélioration des performances lorsque la largeur de bande des canaux latéraux augmente (QCIF)

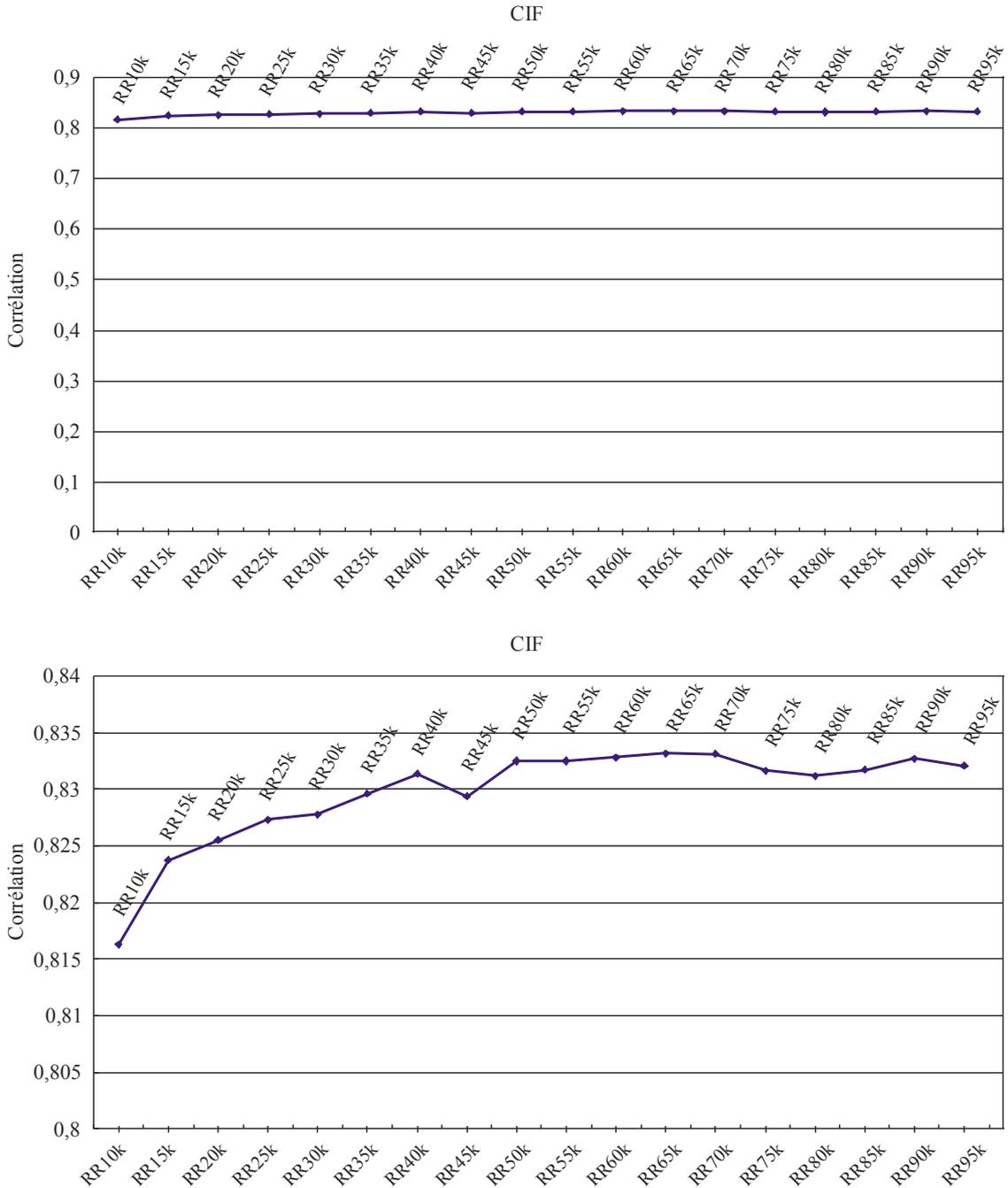


BT.1867-20

La Fig. 21 montre les coefficients de corrélation pour différentes largeurs de bande des canaux latéraux pour le format CIF. On peut constater que les coefficients de corrélation sont pour la plupart saturés à 15 kbit/s environ. Puis, l'augmentation de la bande passante produit une amélioration de 0,5% environ.

FIGURE 21

Amélioration des performances lorsque la largeur de bande des canaux latéraux augmente (CIF)



BT.1867-21

La Fig. 22 montre les coefficients de corrélation pour différentes largeurs de bande des canaux latéraux pour le format VGA. On peut constater que les coefficients de corrélation sont pour la plupart saturés à 30 kbit/s environ. Puis, l'augmentation de la bande passante produit une amélioration de 0,5% environ.

FIGURE 22

Amélioration des performances lorsque la largeur de bande des canaux latéraux augmente (VGA)

