



CODAGE D'IMAGES ANIMEES AVEC INTERPOLATION PAR COMPENSATION DE MOUVEMENT - APPLICATION AU CODAGE DE HAUTE QUALITE D'IMAGES TV COULEUR A 5 MBIT/S

José HANEN - Fabrice LALLAURET

LATI/IRESTE - La Chantrerie CP 3003 - 44087 NANTES CEDEX 03
Tel : (33) 40 68 30 48 - Fax : (33) 40 68 30 66 - Email : lati@lati.ireste.fr

Pour conserver un très haut niveau de qualité dans le codage des images animées TV couleur avec des taux de compression importants, la décomposition en sous-bandes spatiales et le codage différentiel par compensation de mouvement avec quantification psychovisuelle ont déjà été utilisés avec succès pour atteindre un débit de 10 Mbit/s.

Ces techniques efficaces ne peuvent cependant permettre à elles seules d'atteindre un débit de 5 Mbit/s sans réduire de manière significative soit la résolution spatiale soit la qualité de la quantification.

Nous présentons ici une méthode de codage en sous-bandes permettant de maintenir simultanément la pleine résolution spatiale et la qualité de quantification. Cette méthode utilise un sous-échantillonnage temporel de la séquence à coder (une image sur deux est codée, l'autre étant interpolée) avec transmission d'un unique champ de vecteurs.

1. INTRODUCTION

Le codage numérique des images animées de télévision couleur à faible débit et à bonne qualité visuelle nécessite la mise en oeuvre de stratégies efficaces de représentation combinée des signaux en espace et en temps.

Notre objectif est ici de réduire le débit naturel de transmission, égal à 216 Mbit/s, d'une séquence de télévision couleur à la norme CCIR 601 pour atteindre un débit de 5 Mbit/s.

Dans un schéma précédent associé à des débits de 15 et 10 Mbit/s, les techniques de décomposition en sous-bandes associées à un codage différentiel par compensation de mouvement et à une régulation du débit de sortie ont permis d'atteindre une très bonne qualité visuelle à la distance normalisée de 6 fois la hauteur de l'écran de visualisation (D.BARBA, F. LALLAURET [1], [2]).

Pour passer de 10 à 5 Mbit/s, une amélioration des techniques précédentes aurait pu être envisagée avec en particulier une modification importante des pas de quantification, reliés aux seuils différentiels de visibilité calculé par sous-bande. L'expérience du passage de 15 à 10 Mbit/s nous a permis de

To obtain a high level of quality when coding a TV color video signal with a high compression rate, the spatial subband decomposition and the differential coding with motion compensation and psychovisual quantization have been previously successfully used for coding a color television sequence at a rate of 10 Mbit/s.

However these techniques are not sufficiently efficient to achieve the objective of a rate of 5 Mbit/s without any significant reduction of the spatial resolution and of the quantification quality.

We present hereunder a subband coding method which allows simultaneously to maintain the full spatial resolution and the quantification quality. This method is based on a frame skipping algorithm (one frame is coded, the other is interpolated) and uses only one motion vector field.

prévoir qu'une telle modification aurait conduit à une dégradation considérable de la qualité visuelle de la séquence reconstruite.

Nous avons donc recherché une méthode de codage permettant de conserver la qualité de la reconstruction au débit de 5 Mbit/s.

2. PRINCIPE

Le principe à la base du schéma de codage proposé est le sous-échantillonnage temporel de la séquence : seule une image sur deux est codée, l'autre étant reconstruite par interpolation avec compensation de mouvement.

L'originalité de la méthode tient dans le fait qu'aucune information n'est fournie pour interpoler les images non codées. L'interpolation est effectuée par compensation de mouvement en utilisant les seules informations de mouvement fournies pour décoder les images codées : un seul champ de vecteurs est donc transmis. Le traitement est effectué sur chaque trame de chacune des images codées.

Le schéma général du codeur est celui de la figure numéro 1.



En début de séquence et à chaque changement de plan, chaque trame de l'image courante est codée en mode spatial : les trois composantes Y, dR et dB du signal sont tout d'abord décomposées dans le plan fréquentiel en sous-bandes, puis un codage (prédicatif ou non) est appliqué suivi d'une quantification scalaire adaptée aux capacités de perception différentielles du système visuel humain.

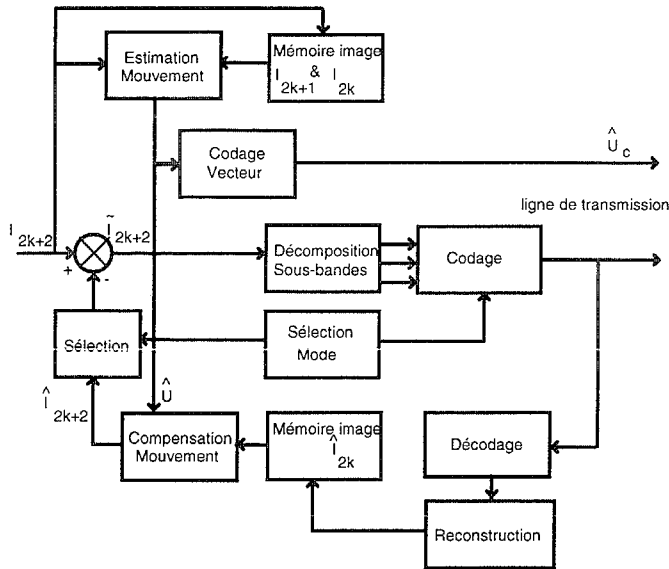


Figure n°1 : Schéma de principe du codage d'une séquence d'images TV à 5 Mbit/s avec décomposition en sous-bandes et compensation de mouvement.

Les images suivantes d'ordre pair sont traitées en mode temporel : une prédiction de l'image courante I_{2k+2} est d'abord effectuée par compensation de mouvement à partir de l'image I_{2k} et du champs de vecteurs de mouvement correspondant; puis le signal d'erreur est décomposé en sous-bandes et quantifié en mode scalaire en utilisant les mêmes seuils différentiels de visibilité que dans le mode spatial.

Les images d'ordre impair ne sont pas codées. Elles seront reconstruites au niveau du récepteur par interpolation à partir des deux images codées encadrantes et du champs de vecteurs de mouvement.

3. MODE SPATIAL

Nous avons conservé le mode de codage spatial utilisé pour le 10 Mbit/s [2] qui fait apparaître trois étapes :

a) 1ère Etape : Décomposition du signal d'entrée

Dans ce mode chaque composante de chacune des trames codées est tout d'abord décomposée en utilisant des filtres 2D séparables de type PQMF ([3], [4]) qui assurent une décomposition-reconstruction pratiquement exacte pour une longueur des bancs de filtres suffisamment importante ($\approx 8 \times N$ sous-bandes), tout en conservant la localisation spatiale des informations.

La composante de luminance est décomposée en 8×4 sous-bandes et celles de chrominances le sont en 4×4 sous-bandes (figure n°2) compte tenu du tramage des images.

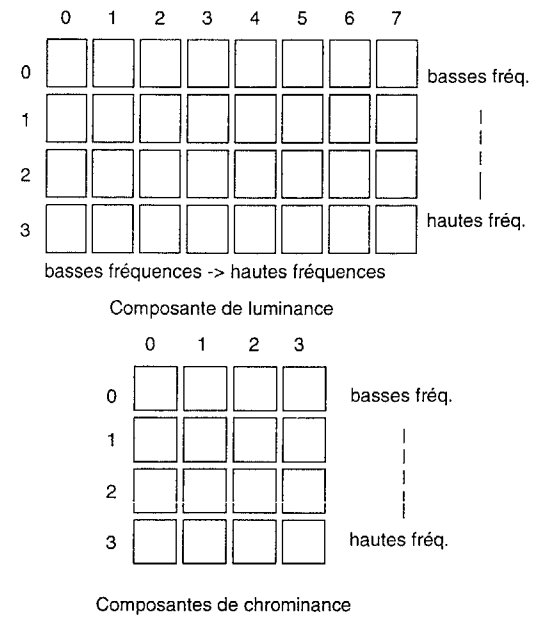


Figure n°2 : Décomposition en sous-bandes.

b) 2ème étape : codage prédictif dans les sous-bandes

Des analyses statistiques dans les sous-bandes ont permis de mettre en évidence la présence de corrélations spatiales dans les seules sous-bandes basses fréquences pour lesquelles un codage de type MIC différentiel est mis en oeuvre.

Pour les autres sous-bandes, un codage de type MIC nous a paru plus adapté compte tenu des très faibles corrélations observées.

c) 3ème étape : quantification

Une quantification de type scalaire a été retenue.

- Les données pour les sous-bandes très basses fréquences sont quantifiées en utilisant un nombre fixe et élevé de niveaux de quantification.

- En basses fréquences un critère d'activité (égal au nombre de pixels à coder dans chaque sous-bande) permet de sélectionner un nombre de niveaux de quantification parmi deux prédéfinis (figure n°3 pour la composante Y).

- Pour les autres sous-bandes, une quantification psychovisuelle utilisant le Seuil Différentiel de Visibilité (SDV) du système visuel humain pour chaque sous-bande considérée est mise en oeuvre, c'est directement le SDV qui est utilisé comme pas de quantification. Certaines sous-bandes hautes fréquences ont de plus été systématiquement éliminées d'après des études menées sur la qualité de reconstruction finale[2].

Après transmission, la reconstruction est effectuée par le même banc de filtres que celui utilisé pour la décomposition, les effets de bord étant traités par effet miroir.

	0	1	2	3	4	5	6	7
0	64	31	17/11	11/7	SDV	SDV	SDV	XXX
1	31	23/15	13/9	SDV	SDV	SDV	XXX	XXX
2	17/11	13/9	SDV	SDV	SDV	SDV	XXX	XXX
3	11/7	SDV	SDV	SDV	SDV	XXX	XXX	XXX

figure n°3 : niveaux de quantification prédéfinis dans les sous-bandes (composante de luminance)

4. MODE TEMPOREL

Après codage en mode spatial de l'image initiale de la séquence, les images suivantes sont codées en mode temporel par compensation de mouvement.

a). Compensation de mouvement

La compensation de mouvement est basée sur une approche de mise en correspondance de blocs utilisant un léger recouvrement entre les blocs à compenser.

Un seul champ de vecteurs étant utilisé à la fois pour la prédiction et l'interpolation, un critère mixte d'appariement de blocs a été développé et testé : ce critère est de la forme

$$\text{crit}(\Delta) = \alpha \text{crit}_{\text{int}}(\Delta) + (1 - \alpha) \text{crit}_{\text{pred}}(\Delta)$$

- ou crit_{int} : critère prenant en compte la seule qualité de l'interpolation des blocs de l'image I_{2k+1} .
- $\text{crit}_{\text{pred}}$: critère prenant en compte la seule qualité de prédiction de l'image I_{2k+2}

pour l'interpolation : chaque bloc de l'image I_{2k+1} est comparé avec la moyenne des blocs déplacés de Δ (en sens inverse) dans les images I_{2k} et I_{2k+2} (figure numéro 4).

pour la compensation : chaque bloc de l'image I_{2k+2} est comparé avec le bloc translaté de -2Δ de l'image I_{2k} (figure numéro 5).

Chacun de ces critères est de type Minimum de l'Erreur Absolue (MEA). Le vecteur Δ retenu correspond au minimum de $\text{crit}(\Delta)$ calculé sur l'ensemble des déplacements admissibles et au demi pixel près (un déplacement est admissible s'il provient d'une région interne à l'image et dans une fenêtre de ± 8 pixels en horizontal et de ± 4 pixels en vertical). La valeur du coefficient α a été fixée à 0.5.

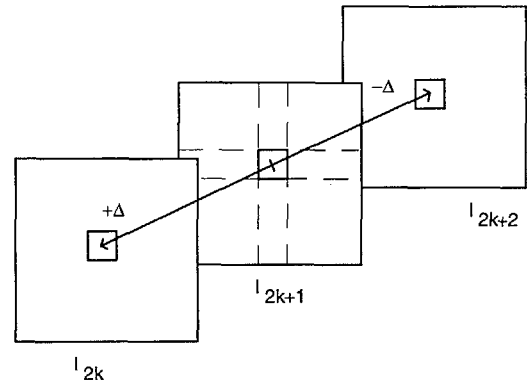


Figure n°4 : Principe de l'interpolation

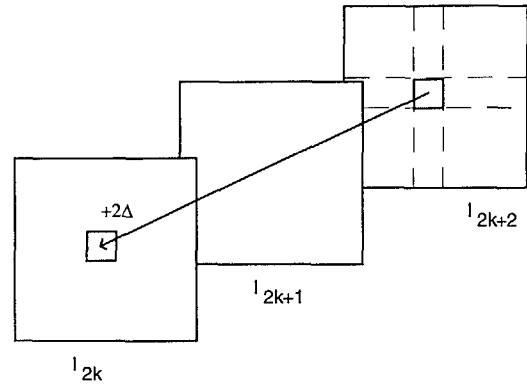


Figure n°5 : Principe de la compensation

Le nombre de bits à transmettre peut être important pour les séquences très bruitées ("voiture" par exemple). Pour réduire l'entropie moyenne du champs de vecteurs mouvement, un post-filtrage est appliqué ayant pour effet de dupliquer le maximum de vecteurs tout en préservant une bonne qualité de compensation et d'interpolation. Un découpage du champs de vecteurs en blocs 2×2 et un codage entropique sont ensuite utilisés pour coder ces informations de mouvement.

Résultats de la seule compensation de mouvement :

L'emploi d'un critère mixte (donc non optimal pour la seule compensation) dans la détermination des vecteurs de mouvement ainsi que la plus grande distance temporelle entre deux images codées conduisent à une dégradation de la qualité de l'image compensée. Les résultats suivants ont été relevés :

séquence (moyennes sur 30 images)	entropie par pixel du champ de vecteur	Y rapport signal/bruit (en dB)	dR rapport signal/bruit (en dB)	dB rapport signal/bruit (en dB)
mobile	0.04	26.3 / 31.4	38.8 / 43.2	37.4 / 40.7
voiture	0.16	24.1 / 26.2	37.8 / 39.9	33.9 / 36.2
girl	0.06	26.5 / 30.4	34.3 / 38.6	34.9 / 37.8
flower	0.10	24.3 / 30.8	36.8 / 41.8	34.6 / 40.4

Figure n°6 : statistiques sur l'erreur de compensation/interpolation



b) Signal d'erreur de compensation

Le signal d'erreur de compensation est décomposé en sous-bandes comme dans le mode spatial mais un certain nombre de sous-bandes en hautes fréquences sont systématiquement mises à zéro. Ainsi 12 sous-bandes sur 32 sont ignorées en luminance, 10 sur 16 pour le signal de chrominance dR et 13 sur 16 pour le signal de chrominance dB.

Le signal dans les sous-bandes est ensuite quantifié scalairement avec pour pas de quantification le Seuil Différentiel de Visibilité SDV : la quantification du pixel $e(m,n)$ de la sous-bande (i,j) est donnée par :

$$e_q(m,n) = \delta_{ij} Q[e(m,n) / \delta_{ij}]$$

$Q[x] =$ entier le plus proche

δ_{ij} = seuil différentiel de visibilité dans la sous-bande i,j

c) Régulation de débit

De façon à conserver un débit de transfert constant, un tampon de quatre images ainsi qu'une régulation de remplissage de la mémoire tampon ont été mis en oeuvre. Cette régulation agit par modulation des seuils de visibilité à l'aide de deux coefficients R_f et R_g :

$$\delta_{ij}' = \delta_{ij} * (1 + R_f + R_g)$$

R_g : coefficient de régulation grossier (dépendant de manière non linéaire de l'état de remplissage du tampon)

R_f : coefficient de régulation fin (dépendant de la valeur lissée de la dérivée de l'état de remplissage)

δ_{ij}' : nouvelle valeur de δ_{ij} recalculée 10 fois par image

d) Résultats de la reconstruction à 5Mbit/s, mode régulé

séquence images comp./recons.	Y : rapport signal/bruit (en dB)	dR : rapport signal/bruit (en dB)	dB : rapport signal/bruit (en dB)
mobile	24.2 / 27.1	35.9 / 37.9	31.7 / 32.2
voiture	23.2 / 27.9	36.8 / 38.5	32.2 / 33.4
girl	25.4 / 30.5	32.4 / 34.9	33.0 / 33.8
flower garden	22.9 / 28.3	34.2 / 35.5	30.9 / 31.9

Figure n°7 : statistiques en mode régulé sur l'erreur de compensation/reconstruction

séquence images interpolées	Y : rapport signal/bruit (en dB)	dR : rapport signal/bruit (en dB)	dB : rapport signal/bruit (en dB)
mobile	29.4	40.6	36.0
voiture	25.9	39.2	35.0
girl	29.6	38.9	36.3
flower garden	29.4	38.5	35.7

Figure n°8 : statistiques en mode régulé sur l'erreur d'interpolation

Les valeurs moyennes des rapports signal à bruit des figures précédentes ne permettent pas d'apprécier la bonne qualité visuelle obtenue. En effet, l'utilisation de critères psychovisuels peut conduire à accepter des défauts qui restent inférieurs aux seuils de visibilité mais contribuent à la diminution des rapports signal à bruit.

L'interpolation des images impaires se comporte comme un filtrage passe bas ; nous avons donc appliqué, pour la visualisation des séquences reconstruites, un léger post-filtrage spatial passe bas des images codées de façon à conserver une qualité visuelle constante pendant toute la séquence.

5. CONCLUSIONS et PERSPECTIVES

Nous avons présenté un nouveau schéma de codage de séquences d'images de télévision couleur à 5Mbit/s validé sur quatre séquences d'images de télévision représentatives (MOBILE, VOITURE, GIRL et FLOWER GARDEN).

Les résultats obtenus montrent qu'avec une configuration optimisée des paramètres, une bonne qualité (à six fois la hauteur de l'écran) peut être obtenue même dans le cas de séquences très critiques, à un débit régulé de 5 Mbit/s à pleine résolution spatiale.

Afin d'atteindre des débits plus faibles sans diminution de la qualité visuelle, des études sont en cours concernant l'emploi d'une quantification vectorielle, le codage simultané de l'erreur de compensation de mouvement des deux trames d'une même image, l'emploi d'une modélisation globale des champs de vecteurs de mouvement et l'amélioration des performances de la régulation de débit.

Bibliographie :

[1] F. LALLAURET, D. BARBA "Motion Compensation by block matching and vector post processing in subband coding of TV signals at 15 Mbit/s" Visual Communications and Image Processing' 91, Boston.

[2] F. LALLAURET, "Codage en sous-bandes de séquences d'images animées : Application au codage d'images de télévision couleur à 15 et 10 Mbit/s", Thèse de Doctorat, Université de NANTES, Novembre 1992.

[3] J.B. RAULT, "Algorithme de réduction de débit pour le codage des voies son haute qualité", Thèse, Université de Rennes 1, 1987.

[4] M. VETTERLI, "Multi-dimensional sub-band coding : some theory and algorithms", Signal Processing Vol 6, 1984, pp. 97-112.