

Stratégie de codage conjoint pour la transmission de séquences d'images via un canal bruité

WANG S., PERRINE C., OLIVIER C. et CHATELLIER C.

Laboratoire SIC, Université de Poitiers
Téléport 2, Bvd M. et P. Curie, BP 30179, F-86962 Futuroscope-Chasseneuil cedex
Wang.Olivier.Chatellier,Perrine@sic.sp2mi.univ-poitiers.fr

Résumé

Dans cet article, qui a trait au codage conjoint source canal, notre objectif est de contribuer à l'amélioration de la qualité visuelle d'images vidéo reçues dans des conditions de TEB (Taux d'Erreurs Binaires) assez important (supérieur à 10^{-4}), tout en minimisant la complexité globale de la chaîne de transmission.

Afin de réaliser ce codage, nous proposons une association entre une transformation en ondelettes classique, une quantification vectorielle par sous bande et une modulation de type MAQ-M. On exploite conjointement l'ensemble des éléments constituant la chaîne de transmission numérique pour rendre le schéma proposé robuste à la fois dans des conditions de transmission difficiles et pour des canaux bas débit.

Mots clés

Codage conjoint source canal, quantification vectorielle, transmission vidéo, GOP.

1. Introduction

Aujourd'hui, la nécessité de transmettre des images ou de la vidéo croit rapidement avec le développement des communications modernes. De plus, la quantité et la qualité de l'information à transmettre ne cessant d'augmenter, l'étape de compression joue un rôle essentiel. Ceci est d'autant plus vrai pour le signal vidéo. Les normes récentes comme H.264 ou MPEG-4 ([1], [2]) ont démontré de bonnes performances, avec des algorithmes de plus en plus intéressants en termes de compression (notamment par les derniers travaux sur le codage optimisé des vecteurs mouvements) et de scalabilité de l'information à transmettre. Cependant, celles-ci restent fragiles aux attaques susceptibles d'apparaître sur le canal de transmission.

L'objectif de cet article est de présenter une stratégie de codage vidéo conjoint source canal, basée sur des décompositions en ondelettes, et qui soit robuste aux erreurs de transmission ($TEB > 10^{-4}$) tout en offrant un bon compromis débit/distorsion. La contrainte d'une application temps réel est également prise en compte dans la mesure où la complexité de la méthode présentée est

réduite. L'application visée est de type visioconférence, ce qui autorise un nombre réduit de changements de plans pour les canaux bas débits.

Le codeur vidéo proposé dans cet article est basé sur une quantification vectorielle (QV) issue de [3], appliquée à des coefficients en ondelettes et couplée à une modulation de type MAQ (-16 ou -256). Ce codeur porte le nom de WTSOM (pour Wavelet Transform Self Organized Map) vidéo. Il présente le grand avantage, par son codage à longueur fixe, d'être bien plus robuste aux erreurs que les codes à longueur variable usuels (VLC), au prix bien sûr d'un moindre taux de compression.

Dans le paragraphe 2, nous décrivons la méthode de codage conjoint WTSOM vidéo initiale, directement issue de la stratégie WTSOM d'origine développée sur des images fixes [4]. Nous montrons son adéquation à une modulation MAQ-256 par ses bonnes performances vis-à-vis d'erreurs de transmission trop nombreuses. Nous proposons, comme pour MPEG-4, d'intégrer un codage prédictif basé GOP, mais à partir de décompositions en ondelettes, en considérant 12 images successives. Il s'agit donc de la mise en œuvre d'une stratégie de codage conjoint source canal dont l'application visée est la transmission de la vidéo sur canaux complexes perturbés.

Dans le troisième paragraphe, nous montrons des résultats en transmission réelle de notre méthode WTSOM vidéo, via un canal gaussien caractérisé par son TEB à taux d'erreur binaire supérieur à 10^{-4} . La bonne tenue de cette méthode est claire dans le cas présenté, sur le plan strictement robuste, mais non en termes de taux de compression. C'est pourquoi nous terminons cette contribution par des perspectives à court terme pouvant améliorer les performances de ce travail, mais en occultant volontairement l'introduction de codes correcteurs, dont l'utilisation dans notre cas ne varierait pas des codeurs usuels dès lors qu'ils offrent des propriétés de scalabilité comparables.

2. Principe du WTSOM Vidéo

Après avoir montré la difficulté à transmettre des séquences vidéo dès que le taux d'erreur binaire du canal est assez élevé, nous rappelons le principe de WTSOM [4] sur une image (un plan), son intérêt en superposant le

dictionnaire à une constellation donnée, et sa stratégie de codage prédictif à partir des GOP.

2.1 Problématique

Un exemple de transmission d'une vidéo codée par MPEG-4 est illustré sur la figure 1 avec un TEB de l'ordre de 10^{-4} , sur canal gaussien et en utilisant une modulation MAQ-16 classique. Ce qui est observé se justifie par ce que nous savons: la grande sensibilité des codeurs usuels en cas d'erreurs sur le canal, due en partie aux codages entropiques employés (MDL).



Fig. 1: Séquence transmise sur canal idéal (sans erreurs) et sur canal gaussien:
(a) séquence originale (b) MPEG-4, TEB = $1,9 \cdot 10^{-4}$

Pour palier ce handicap, nous introduisons un codeur source basé sur des quantifications vectorielles, et donc à longueur fixe.

2.2. Construction d'un dictionnaire sur un plan

Considérons une image. Nous utilisons tout d'abord les ondelettes bi-orthogonales de Daubechies (9/7) sur trois niveaux, pour lesquels nous ne conservons que les cinq sous-bandes les plus significatives : LL_3, HL_3, LH_3, HL_2 et LH_2 . Ce choix est un bon compromis entre taux de compression et qualité visuelle de l'image. Il est conforté par une étude de l'entropie de l'information sur chacune des 10 sous-bandes des trois niveaux sur notre base d'apprentissage de séquences d'images. Dans un second temps, nous appliquons une quantification vectorielle (QV), différente pour chaque sous bande en utilisant les cartes topologiques de Kohonen et l'algorithme qui en découle [5] : le codeur WTSOM. Cette méthode nécessite donc le calcul de cinq dictionnaires (pour chacune des 5 sous-bandes retenues), connus de l'émetteur et du récepteur. Le nombre et la forme des vecteurs ou blocs de chaque dictionnaire dépendent des sous bandes, du taux de compression et de la qualité voulue de l'image

reconstruite. Pour notre exemple de la figure 2, chaque dictionnaire est composé de 256 vecteurs V_i dont la taille est d'autant plus petite que la sous-bande traitée est plus informative.

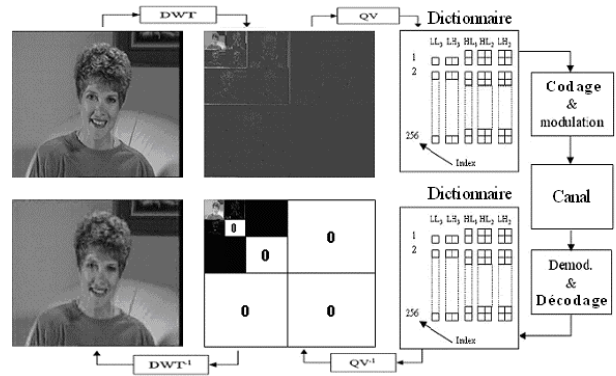


Fig. 2: Méthode WTSOM vidéo: Principe de la QV sur un plan.

2.3 Robustesse aux erreurs de transmission

Seuls les indices des dictionnaires sont ainsi transmis par l'intermédiaire d'une modulation adaptée sur le canal considéré. Le nombre d'éléments du dictionnaire est justifié par la modulation utilisée (MAQ-256 ou MAQ-16) : en effet, on montre sur la figure 3 la représentation de la constellation correspondant à une MAQ-256 superposée à un dictionnaire de 256 vecteurs et sa carte de Kohonen. Une erreur n'aura que peu d'impact visuel si le symbole erroné est décodé par l'indice d'un vecteur proche du vecteur réel. Ceci est dû à la carte de Kohonen montré en figure 3 sous les points de la constellation (en rouge), où les indices voisins géographiquement représentent des images (vecteurs) proches visuellement.

Cette interdépendance entre le codeur WTSOM et la modulation est la cause de la bonne robustesse du schéma conjoint proposé.

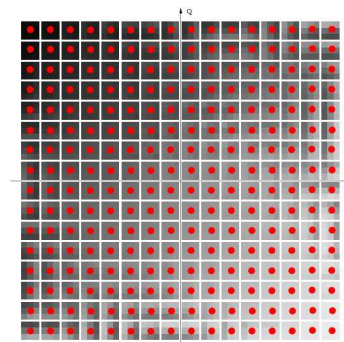


Fig. 3: Superposition des 256 éléments d'un dictionnaire d'une carte de Kohonen sur les symboles d'une modulation MAQ-256

2.4 Codage prédictif

Nous reprenons la stratégie proposée dans MPEG-4 mais sur 12 images successives [2] et qui diffère de celle sur 3 images successives proposées initialement dans [6]. Notre méthode requière la fabrication de 2 fois cinq dictionnaires appris sur la base d'images des différences de référence, soit 4 séquences vidéo : *Mom-daughter*, *Claire*, *Missa*, *Grande-mom*). Pour les dictionnaires, chaque vecteur est de taille variable, choisie comme indiqué au paragraphe 2.2. Nous appliquons notre méthode sur des séquences de douze images suivant la technique GOP utilisée dans MPEG-4 ou H264, mais ici sur des sous-bandes DWT.

Pour le traitement des redondances temporelles, nous retenons les solutions de MPEG-4 en instituant trois types d'images qui seront traitées selon des procédés spécifiques :

Les images intra ou images de référence, symbolisées par la lettre **I**;

Les images prédites ou images **P**;

Les images bidirectionnelles ou images **B**.

La figure 4 montre leur répartition sur une séquence de treize images.

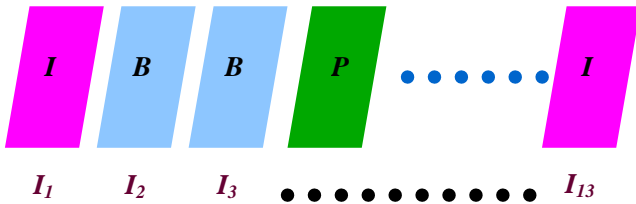


Fig. 4: GOP sur une séquence de 13 images

Les images de référence (I) ou intrapicture sont transformées selon l'algorithme DWT. Ces images sont codées en mode intra, par l'intermédiaire de la technique de codage d'image fixe WTSOM déjà détaillée ci-dessus. Ces images serviront de référence pour le calcul des images P et B ainsi que pour le décodage. Pour les images P, elles sont codées en mode prédictif. Une compensation de mouvement d'une image précédente est effectuée afin de réaliser la prédiction de cette image. Les images B sont codées en mode bidirectionnel. La prédiction est obtenue par l'intermédiaire de l'évaluation du mouvement basée sur une image précédente et sur une image future.

On calcule alors une différence sur les coefficients d'ondelettes, par sous bande conservée, entre ces images. Dif_1 , Dif_2 , Dif_3 , ..., seront les vecteurs des différences. Ainsi nous pouvons construire 5 dictionnaires sur les images de références, et à partir des vecteurs des différences nous pouvons construire des nouveaux dictionnaires des différences.

Les équations sur les différences sont les suivantes:

$$Dif_1 = WT(I_4) - WT(I_1)$$

$$Dif_2 = WT(I_2) - (WT(I_1) + 1/3Dif_1)$$

$$Dif_3 = WT(I_3) - (WT(I_4) - 1/3Dif_1)$$

....

où $WT(I_j)$ indique un coefficient d'ondelette du plan I_j . Dif_4 , Dif_5 , Dif_6 , etc ..., sont obtenues de la même façon.

Sur la figure 5, nous montrons le principe de WTSOM GOP.

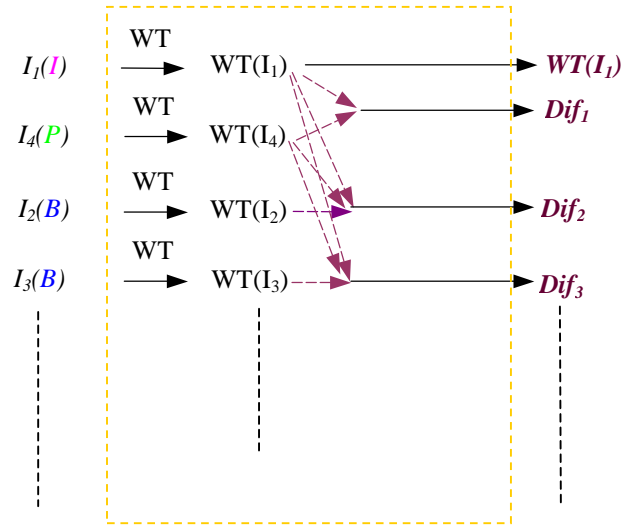


Fig. 5: WTSOM GOP

3. Résultats de transmission sur canal gaussien

Dans cette partie, tous les résultats sont donnés par le schéma codeur WTSOM vidéo décrit ci-dessus et une modulation MAQ-16. Cette modulation (symboles codés sur 4 bits) est facile à mettre en œuvre [3]. Elle présente l'avantage par rapport à une MAQ-256 d'une plus grande simplicité, une plus grande robustesse au bruit, et une moindre complexité, sans déstructurer le schéma initial avec les 256 vecteurs des dictionnaires (qui sont codés sur $8 = 2 \times 4$ bits). Le taux de compression est d'environ 44:1. Nous transmettons la séquence vidéo *Mom* sur canal gaussien (figure 6a: extrait de la vidéo *Mom*: originale: *Mom_024*, *Mom_026*, ... *Mom_035*). Nous construisons les dictionnaires dans les mêmes conditions que précédemment relativement au taux de compression choisi et sur la base des 4 séquences vidéo *Mom-daughter*, *Claire*, *Missa*, *Grande-mom*.

Le résultat de la reconstruction de cette séquence par le codeur WTSOM vidéo est donné par la figure 6b. Les séquences sont visuellement satisfaisantes avec un PSNR moyen de 27.28 dB, mais évidemment moindre qu'avec MPEG-4 ou H264.

La séquence originale de la figure 6a. est ensuite transmise via un canal gaussien avec les caractéristiques suivantes: TEB de l'ordre de 2.10^{-4} , SNR = 8 dB, $T_c = 63,5$

pour WTSOM vidéo et $T_c = 61$ pour MPEG-4, et une MAQ-16. Les résultats obtenus avec notre schéma conjoint sont présentés sur la figure 6c et ceux obtenus avec MPEG-4 sont illustrés sur la figure 6d. Sur la figure 6c, la valeur de PSNR moyen ne diminue pas (0.02dB) par rapport à la figure 6b. On montre aussi en figure 6e le résultat avec un fort TEB ($TEB = 1,34 \cdot 10^{-2}$) et le même taux de compression, ce qui prouve la robustesse de la méthode présentée (diminution 1,7 dB encore acceptable) et son intérêt pour le codage de vidéo comparé au standard MPEG-4 où le résultat est visuellement inacceptable.

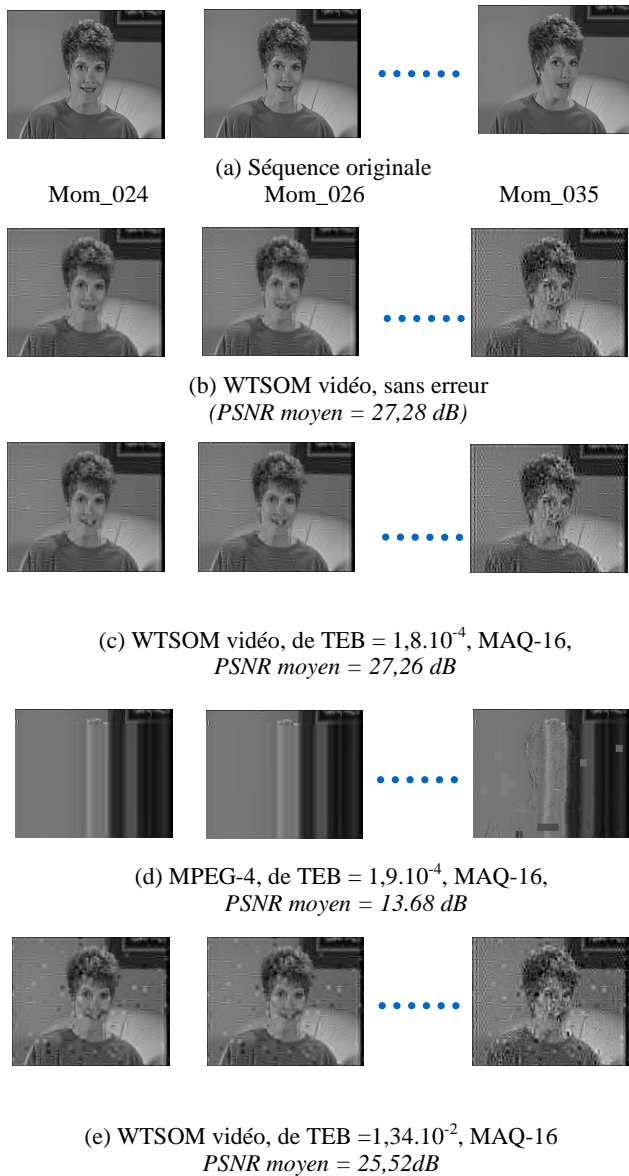


Fig. 6: Séquence transmise sur canal idéal et sur canal gaussien, à taux de compression comparable: (a) séquence originale (b) WTSOM vidéo, sans erreur, (c) WTSOM vidéo, $TEB = 1,8 \cdot 10^{-4}$, (d) MPEG-4, $TEB = 1,9 \cdot 10^{-4}$ (e) WTSOM vidéo, $TEB = 1,32 \cdot 10^{-2}$

4. Conclusion et perspectives

Dans cet article, nous avons proposés un schéma de codage conjoint intitulé WTSOM vidéo. Il est basé sur des quantifications vectorielles à partir de décomposition en ondelette et d'un codage prédictif classique sur des séquences de 12 images. Nous montrons que cette stratégie se prête bien à la transmission d'images sur canal moyennement perturbé, comparé au standard MPEG-4, et sans introduction de codes correcteurs qui amélioreraient les méthodes au détriment du débit.

En termes de qualité visuelle, nous pouvons améliorer cette méthode en introduisant, comme dans les standards usuels de vidéo, un vecteur de mouvement, ce qui permettra une diminution significative de la distorsion résiduelle sur les images reconstruites [7]. Son codage se fera lui aussi par une quantification vectorielle adaptée à la modulation pour conserver sa robustesse à l'algorithme global, tout en veillant à minimiser son impact sur le débit.

Enfin, certaines erreurs (voir figures 6c et 6e) pourront être éliminées par un algorithme de restauration de vidéo basé EDP que nous avons développé [8] sur images fixes.

Bibliographie

- [1] H264/MPEG-4 Part 10 Tutorials (Richardson), <http://www.vcodex.com/h264.html>
- [2] MPEG-4 Video Group, Coding of audio-visual objects: Video, ISO/IEC JTC1/SC29/WG11 N2202, March 1998
- [3] Aitsab O. "Turbo codes et codage conjoint source canal : application à la transmission d'images". PhD thesis, Ecole Nationale Supérieure des Télécommunications de Bretagne, Février 1998
- [4] Souhard, B. "Codage conjoint source canal : Application à la transmission d'images fixes sur canal ionosphérique", Thèse de l'Université de Poitiers, Mars 2004.
- [5] Kohonen T. "Self Organization and associative memory". Springer-Verlag, 1994.
- [6] Wang S., Chatellier C., et Olivier C. "Codage conjoint source canal pour des séquences d'images visioconférences", CORESA, Caen, pp 89-93, Novembre 2006
- [7] Agostini M.A., Antonini M., Barlaud M. "Model-based bit allocation between wavelet subbands and motion information in MCWT video coders", EUSIPCO 2006, Florence (Italy), September 2006
- [8] Bourdon P., Augereau B., Chatellier C., Olivier C.: "A multi-resolution, geometry-driven error concealment method for corrupted JPEG color images" - *Signal Processing: Image and Communication*, 20, pp 681-694, August 2005