

# Codage conjoint source canal appliqué aux séquences d'images visioconférences

WANG S., CHATELLIER C. et OLIVIER C.

Laboratoire SIC, Université de Poitiers  
Téléport 2, Bvd M. et P. Curie, BP 30179, F-86962 Futuroscope-Chasseneuil cedex  
[Wang@sic.sp2mi.univ-poitiers.fr](mailto:Wang@sic.sp2mi.univ-poitiers.fr)

## Résumé

Dans cet article, nous nous intéressons au codage conjoint source canal appliqué à des séquences d'images de type visioconférences transmises sur canal gaussien. La méthode proposée repose sur une DWT et une quantification vectorielle adaptée aux différentes sous bandes d'ondelettes considérée. Après avoir rappelé l'algorithme le principe de ce codage appelé WTSOM (Wavelett Transform Self Organize Map) sur des images fixes, nous proposons une extension aux images en mouvement à partir de l'utilisation de dictionnaires 3D, puis une amélioration en terme de distorsion basée sur l'utilisation de différences d'images. L'exécution de cette nouvelle méthode nous permet d'obtenir une bonne qualité visuelle des images en mouvement même avec des TEB très élevés.

## Mots clés

Codage conjoint source canal, visio conférence, Quantification vectorielle, transmission d'images.

## 1. Introduction

Afin de résister aux erreurs de transmission liées au canal gaussien, nous proposons dans cette contribution un codage d'image basé sur une quantification vectorielle (QV) appliquée à des coefficients en ondelettes et couplée à une modulation [1][2][3]. Le nombre d'états de la modulation étant idéalement identique au nombre d'indices du dictionnaire utilisé pour la QV [4]. Si, en terme de performance de débit, cette méthode est inférieure aux standards, elle présente le grand avantage d'être, par son codage à longueur fixe, bien plus robuste aux bruits que les classiques codes VLC usuelles. Déjà présentée dans le cas de transmission d'image fixes [3], la méthode employée est ici étendue et adaptée au codage d'images vidéo.

Dans le paragraphe 2, nous décrivons ainsi la méthode de codage conjoint WTSOM et présentons ses bonnes performances pour des images fixes transmises sur un canal gaussien. Nous proposons un algorithme appelé WTSOM 3D dans le 3<sup>ème</sup> paragraphe, qui adapte le précédent aux séquences d'images en considérant 3 images successives et WTSOM appliqué directement sur la série de 3 images (3D). Après avoir donné les limites de

cette première extension, nous choisissons dans le 4<sup>ème</sup> paragraphe de travailler avec les différences entre images, en accord avec les stratégies de codages vidéo usuels. Nous proposons un nouvel algorithme basé sur ce principe, puis nous comparons ces deux méthodes en terme de qualité de reconstruction. Dans le 5<sup>ème</sup> paragraphe, nous montrons des résultats en transmission de cette dernière méthode WTSOM 3D avec différence sur différentes séquences via un canal gaussien caractérisé par son TEB (Taux d'Erreur Binaire). La robustesse de cette méthode est claire dans le cas présenté. Nous terminons cette contribution par une conclusion et les perspectives nombreuses pouvant faire suite à ce travail.

## 2. WTSOM

L'objectif du codage conjoint source canal appliqué à la transmission d'images est d'améliorer la qualité visuelle de l'image reçue dans des conditions de fort TEB, tout en minimisant la complexité globale de la chaîne de transmission. Afin de réaliser ce codage, nous avons proposé une association entre une transformation en ondelettes classique, une quantification vectorielle par sous bande et une modulation de type M-QAM.

Dans cette application, nous utilisons tout d'abord les ondelettes bi-orthogonales de Daubechies (9/7) [5] 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  $LH_2$  représentées sur la figure 1.



Fig. 1 : Décomposition DWT et sous-bandes retenues.

Dans un second temps, nous appliquons une quantification vectorielle différente pour chaque sous bande en utilisant les cartes topologiques de Kohonen et l'algorithme qui en découle : Self Organize Map (SOM) [6]. Seul les indices des dictionnaires sont transmis par l'intermédiaire d'une modulation choisie en fonction du canal : Gaussien [3] ou Rayleigh [7]. Cette méthode nécessite 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 application, chaque dictionnaire est composé de 256 vecteurs dont la taille est d'autant plus petite que la sous-bande traitée est plus informative.

La figure 2 représente la chaîne de transmission utilisant le codage WTSOM. On peut y observer la taille et la forme choisie pour les vecteurs ou blocs du dictionnaire suivant les 5 sous-bandes LL<sub>3</sub>, HL<sub>3</sub>, LH<sub>3</sub>, HL<sub>2</sub> et LH<sub>2</sub>. Pour une sous-bande donnée, l'indice du vecteur du dictionnaire le plus ressemblant au vecteur extrait de la sous-bande, au sens de la distance euclidienne, est transmis.

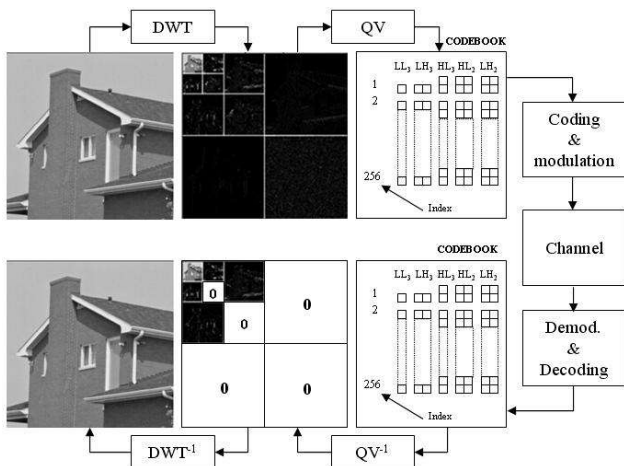


Fig. 2 : chaîne de transmission pour le codage WTSOM.

Un exemple de reconstruction d'images qui ne met pas en jeu les aléas dus à une transmission réelle est représenté sur la figure 3. Le taux de compression ( $T_c$ ) a été fixé ici à 25. Ce résultat montre que la qualité d'image reste bonne même si elle ne peut évidemment pas rivaliser avec les autres codeurs (JPEG, JPEG2000).

La figure 4 permet de mettre en évidence le comportement du codeur WTSOM en présence d'erreurs lors d'une transmission. Pour cette simulation, les erreurs sont générées aléatoirement sur le flux binaire suivant une loi gaussienne. Afin de vérifier la robustesse des différents algorithmes en présence de bruit, nous avons testé JPEG et JPEG2000 pour des TEB faibles de l'ordre de  $10^{-4}$  puis

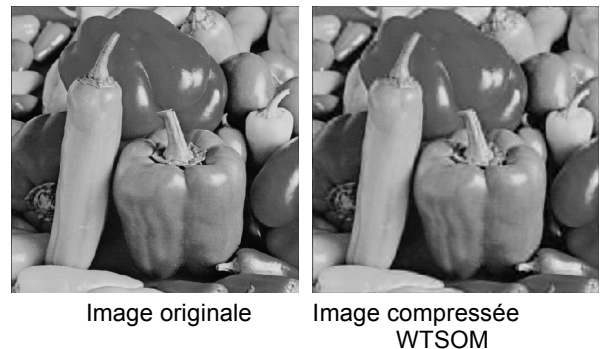


Fig. 3 : Image compressée WTSOM : PSNR=29,37 dB

WTSOM pour des TEB beaucoup plus élevés de l'ordre de  $10^{-2}$ . Ce choix se justifie d'une part, par le fait que pour des TEB trop élevés il n'est pas possible de reconstruire une image JPEG ou JPEG2000, et que d'autre part, pour un TEB trop faible avec WTSOM il n'y a pas d'artefacts visibles sur l'image reconstruite. Ce qui est observé se justifie par ce que nous savons : la grande sensibilité des codeurs usuels en cas d'erreurs. Les résultats montrent clairement que malgré quelques artefacts dus aux erreurs dans la sous-bande LL<sub>3</sub>, la qualité de l'image WTSOM est visuellement bien meilleure.

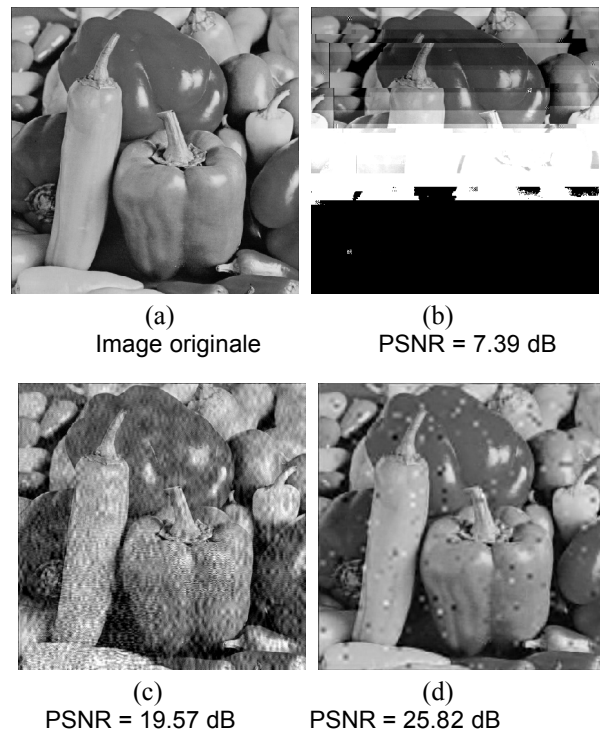


Fig. 4 : transmission d'images sur canal gaussien  
a - image originale; b - JPEG, TEB =  $2 \times 10^{-4}$   
c - JPEG2000, TEB =  $2 \times 10^{-4}$   
d - WTSOM, TEB =  $1,6 \times 10^{-2}$

La méthode WTSOM se prête bien aux techniques de protections hiérarchiques d'informations. En effet, les seuls artefacts visibles sur l'image reçue (taches claires ou

foncées) ne sont dus qu'aux erreurs localisées sur la sous bande  $LL_3$ . Un code correcteur de type Reed Solomon dédié à la protection de cette sous bande a déjà été testé et a permis d'éliminer la quasi-totalité des artefacts visibles pour des TEB de l'ordre de  $10^{-2}$  sur les deux types de canaux déjà cités [2,3].

### 3. WTSOM 3D

L'objet de ce paragraphe est de montrer comment il est possible d'appliquer directement le codage WTSOM à des séquences d'images, ce que nous appelons WTSOM 3D. Ce nouvel algorithme permet de traiter des séquences de trois images consécutives sur le même principe que pour les images fixes. Comme dans le cas précédent, cette méthode requière la fabrication de cinq dictionnaires dans lesquels chaque vecteur ou bloc est de taille variable mais de dimension 3 ou volumique. Les dictionnaires sont construits à partir d'une base d'apprentissage de séquences issues des vidéos suivantes : missa, mom, mom\_daughter, grandmom, claire, susie. Une fois les dictionnaires connus, le processus de transmission s'effectue de la façon suivante :

- 1- Chaque image subie une transformation en ondelettes à 3 niveaux de décomposition.
- 2- Comme l'illustre la figure 5, un vecteur ou bloc 3D est extrait d'une sous bande (ici  $LL_3$ ) puis comparé à l'ensemble des vecteurs du dictionnaire correspondant.
- 3- L'indice du vecteur le plus ressemblant, au sens de la distance euclidienne, est transmis.
- 4- Le récepteur décode l'indice reçu et reconstruit la sous bande à partir des vecteurs d'approximation.

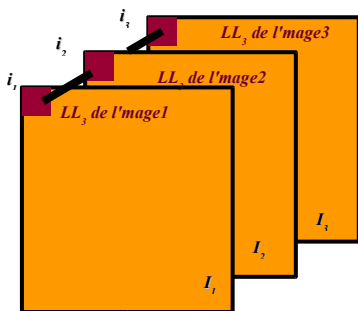


Fig. 5 : représentation d'un vecteur 3D  $(i_1, i_2, i_3)$  composé de trois coefficients d'ondelette des sous bandes  $LL_3$  de la séquence  $I_1, I_2, I_3$

La qualité visuelle de la séquence reconstruite dépend de la taille des vecteurs, du nombre de séquences d'apprentissage et du nombre de vecteur par dictionnaire. Pour notre cas, chaque dictionnaire est constitué de 256 vecteurs et nous avons utilisé une petite base d'apprentissage de 21 séquences de trois images issues de 6 vidéos pour construire les 5 dictionnaires. Nous

présentons un résultat sur une séquence ne faisant pas parti de la base d'apprentissage. Ces premiers résultats obtenus sur des extraits de la séquence vidéo *Foreman* sont représentés sur la figure 6.



Fig. 6 : Reconstruction d'une séquence de 3 images avec WTSOM 3D. PSNR moyen = 19,81 dB

Ils permettent de conclure sur un point : la qualité de la séquence reconstruite n'est pas très bonne comparée à la séquence originale et ceci est le fait, notamment, du nombre restreint de vecteurs retenus (toujours 256) parmi beaucoup plus de vecteurs que dans le cas 2D, ainsi que du trop petit nombre de séquences d'apprentissages utilisées pour construire chaque dictionnaire. Augmenter le nombre de vecteur par dictionnaire reviendrait à augmenter la complexité de la chaîne de transmission et diminuer le débit utile. C'est pourquoi la seconde stratégie que nous mettons en oeuvre utilise les différences d'images comme cela est déjà utilisé dans d'autres codeurs vidéo et ceci afin de minimiser la variabilité des vecteurs d'approximation.

### 4. WTSOM 3D avec différence

Afin de minimiser la diversité des vecteurs présents dans les dictionnaires, nous exploitons la faible différence entre deux images consécutives après leur transformation par DWT. Il s'agit donc d'appliquer une transformée en ondelette, comme décrit au chapitre 2, sur trois images successives. On calcule alors une différence sur les coefficients d'ondelette, par sous bande conservée, entre les deux premières images et entre les deux suivantes :

$$\text{Diff}_1 = i_1 - i_2, \text{Diff}_2 = i_2 - i_3,$$

où  $i_1, i_2, i_3$  représentent 3 coefficients à la même localisation spatiale (donc à la même sous-bande) de trois images successives et  $\text{Diff}_1, \text{Diff}_2$  sont les différences entre les coefficients. Ainsi nous pouvons construire des dictionnaires à partir des vecteurs des différences. Nous proposons qu'une séquence de trois images soit traitée de la façon suivante :

- 1- Chaque image subie une transformation en ondelette.
- 2- On calcule les différences  $\text{Diff}_1$  et  $\text{Diff}_2$
- 3- La première image reste codée WTSOM et est transmise suivant le protocole décrit au paragraphe 3.
- 4- Les différences  $\text{Diff}_1$  et  $\text{Diff}_2$  sont transmises de la même façon, seuls les dictionnaires changent.

La figure 7 illustre les améliorations apportées par cette méthode par rapport à la méthode précédente puisqu'elle permet d'obtenir un gain de PSNR de l'ordre de 3,5 dB.



(a) WTSOM 3D (PSNR moyen = 19.81 dB)



(b) WTSOM 3D avec diff. (PSNR moyen = 23.28 dB)

Fig. 7: Reconstruction de la même séquence codée avec WTSOM et avec WTSOM différence

Nous pouvons remarquer que la qualité se dégrade au fur et à mesure que l'image différence est éloignée de l'image de référence. En effet, si nous représentons l'évolution du PSNR des différentes séquences d'images représenté sur la figure 9, nous observons la cohérence des résultats concernant la baisse du PSNR en fonction de l'image traitée. Les améliorations possibles à envisager sont discutées au paragraphe 6.

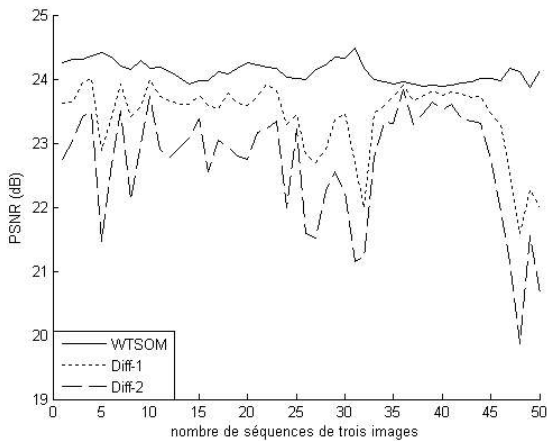


Fig. 8: PSNR des 3 images sur 50 séquences de la vidéo Forman

WTSOM : PSNR moyen=24,1 dB,  $\sigma = 0,0253$

Diff-1 : PSNR moyen=23,41 dB,  $\sigma = 0,558$

Diff-2 : PSNR moyen=22,68 dB,  $\sigma = 0,683$

Nous nous proposons maintenant d'analyser le comportement de ces algorithmes lors d'une transmission sur canal bruité.

## 5. Résultats sur canal gaussien

Dans cette partie, tous les résultats sont traités par la méthode WTSOM 3D avec différence, et le taux de compression est d'environ  $T_c = 35$ . Pour justifier l'intérêt de cette méthode, nous transmettons une séquence vidéo (figure 9a: extrait de la vidéo *Mom* originale Mom\_033, Mom\_034, Mom\_035). Nous construisons les dictionnaires dans les mêmes conditions que précédemment (§. 4) relativement au taux de compression choisi. En figure 9b., nous donnons le résultat de la reconstruction de cette séquence par WTSOM avec différence. Les séquences sont visuellement satisfaisantes avec un PSNR moyen de 31.38 dB. En figure 9c, nous transmettons la première séquence via un canal gaussien de TEB égal à  $2,3 \cdot 10^{-3}$ .



(a)



(b) WTSOM 3D avec différence, sans erreurs  
(PSNR moyen = 31.38 dB)



(c) WTSOM 3D avec différence, avec TEB =  $2,3 \cdot 10^{-3}$   
(PSNR moyen = 29.36 dB)

Fig. 9: Une séquence transmise sur un canal idéal (sans erreurs) et sur un canal gaussien:

(a) la séquence originale

(b) WTSOM 3D avec différence, sans erreurs,  
PSNR moyen = 31.38 dB

(c) WTSOM 3D avec différence, TEB =  $2,3 \cdot 10^{-3}$ ,  
PSNR moyen = 29.36 dB

La valeur de PSNR moyen diminue de 2dB pour ce TEB de l'ordre de  $2,3 \cdot 10^{-3}$ , ce qui prouve la robustesse de la méthode présentée et son intérêt pour le codage de vidéo. Les artefacts observés proviennent uniquement des erreurs localisées sur la sous bande  $LL_3$ . Il est à noter que contrairement à la vidéo Foreman, certaines séquences de la vidéo *mom* ont été utilisées dans l'élaboration des dictionnaires. Ce cas peut donc être considéré comme

valide d'un point de vue robustesse vis à vis des erreurs de transmission, mais pas d'un point de vue qualité de reconstruction .

## 6. Conclusion et perspectives

Dans cet article, nous avons proposés un nouvel algorithme de codage conjoint source canal WTSOM 3D appliqué au codage de séquence d'images de type visio conférence. Il est basé sur une quantification vectorielle de coefficients d'ondelette regroupés par séquences de 3 images. Une amélioration de cette méthode, basée sur le traitement de différences d'images, appelé WTSOM 3D avec différence est également proposée et permet une diminution significative de la distorsion résiduelle sur les images reconstruites. Enfin nous montrons que cette stratégie se prête bien à la transmission d'images sur canal bruité en présence d'un fort TEB.

Ces résultats encourageants vont nous conduire à affiner la méthode (apprentissage, nombre de trames considérées) ainsi que la définition des dictionnaires tant en nombre d'indices que de formes des vecteurs. Différentes stratégies de constructions d'images différences doivent également être abordées. Enfin, la connaissance conjointe d'une part de l'impact visuelle d'une erreur sur l'image reconstruite, et d'autre part de sa localisation doit permettre d'appliquer des algorithmes de traitement d'images afin de restaurer l'image reçue.

## 7. Bibliographie

- [1] Souhard B., Chatellier C., Olivier C. "On The Robustness of Joint Source/Channel Coding for Transmission Through an Ionospheric Channel", *HF2003, IEE Ninth International Conference on HF Radio Systems and Techniques*, Bath (UK), June 2003
- [2] Souhard, B. "Codage conjoint source canal : Application à la transmission d'images fixes sur canal ionosphérique", *Thèse de l'Université de Poitiers*, mars 2004.
- [3] Chatellier C., Bourdon P., Souhard B., Olivier C. "An efficient joint source channel coding scheme for image transmission through the ionospheric channel", *European Conference Propagation and Systems*, Brest (France), March 2005
- [4] 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
- [5] Antonini M., Barlaud M., Mathieu P., Daubechies I., "Image coding using wavelet transform" *IEEE Transactions of Image Processing, Vol. 1, Issue 2, pp 205-220*, April 1992.
- [6] Kohonen T. "Self Organization and associative memory". *Springer-Verlag*, 1994.
- [7] Boeglen H., Chatellier C. "On the robustness of a joint source-channel coding scheme for image transmission over non frequency selective Rayleigh fading channels", *2<sup>nd</sup> IEEE International Conference on Information & Communication Technology: From theory to applications*. Damas (Syria), April 2006