

# CODAGE MORPHOLOGIQUE MULTIREOLUTION DES ERREURS DE COMPENSATION DE MOUVEMENT

F. J. Hampson et J.-C. Pesquet \*

Laboratoire des Signaux et Systèmes,  
CNRS/UPS et GDR TdSI,  
ESE, Plateau de Moulon,  
91192 Gif-sur-Yvette Cedex  
e\_mail : hampson@lss.supelec.fr

## RÉSUMÉ

Dans les systèmes de compression de séquences d'image à base de compensation de mouvement, l'erreur de compensation de mouvement est porteuse d'informations nouvelles devant être codées. Dans cet article, nous proposons d'extraire les parties importantes de l'erreur à l'aide d'un algorithme de segmentation combinant une analyse multirésolution avec des opérateurs morphologiques à niveaux de gris. Ensuite nous employons une méthode adaptée aux images à basse résolution pour coder ces informations. Ce codage est optimisé de façon à minimiser la distorsion sous une contrainte de débit.

## ABSTRACT

In image sequence compression systems which rely on motion compensation, the displaced frame difference contains new information which must be coded. In this paper, we propose to extract the important parts of the displaced frame difference using a segmentation algorithm which combines a multiresolution analysis with grey scale morphological operators. Finally, we use a method adapted to low resolution images to code these parts. This coding method is optimized in order to minimize distortion for a bit rate constraint.

## 1. INTRODUCTION

La compression de séquences vidéo est généralement réalisée à l'aide d'une compensation du mouvement apparent entre les images successives [1]. Les régions découvertes et les imperfections de l'estimation produisent une erreur de compensation de mouvement (ECM) qui doit être codée. Cette dernière opération s'effectue le plus souvent en intra et peut se révéler coûteuse en terme de débit.

Les méthodes classiques de codage de l'ECM reposent sur une quantification scalaire ou vectorielle, éventuellement précédée d'une TCD. Leur inconvénient majeur est d'introduire des effets de blocs qui s'avèrent particulièrement gênants à bas débit. Pour parvenir à un codage plus performant, il est intéressant de chercher à mettre en évidence les structures présentes dans la scène considérée. La morphologie mathématique fournit des outils à la fois simples et efficaces pour accomplir cette tâche [2], [3], [4]. Dans [5] est proposée une façon originale d'extraire les "composantes principales" de l'ECM qui utilise des opérateurs morphologiques à niveaux de gris. Les régions ainsi extraites sont quantifiées et codées avec un codage de longueurs de plages ("run length coding") qui est combiné à un codeur arithmétique [6] pour atteindre de forts taux de compression. Dans cette communication, nous proposons une version multirésolution de cette méthode qui permet une meilleure prise en compte de la corrélation 2D des données traitées.

## 2. MÉTHODE DE SEGMENTATION ET D'EXTRACTION DES RÉGIONS

Pour extraire les structures présentes dans l'ECM, nous appliquons un opérateur morphologique à niveaux de gris suivi d'une décision. Nous avons étudié trois opérateurs morphologiques conçus pour des images d'erreur  $e(k, l)$ , i.e. des images ayant une valeur moyenne proche de zéro :

1. fermeture-ouverture appliquée directement sur l'image  $e(k, l)$  :

$$\Psi_n^1(e)(k, l) = \varphi_n \gamma_n(e)(k, l) ,$$

ou  $\varphi_n$  (respectivement  $\gamma_n$ ) représentent la fermeture (respectivement l'ouverture) morphologique avec un élément structurant  $B_n$ . La famille des éléments structurants utilisés,  $\{B_n, n \geq 1\}$ , est telle que  $n_1 < n_2 \Rightarrow B_{n_1} \subset B_{n_2}$  ;

2. décomposition de  $e(k, l)$  en parties claire et sombre et application d'une ouverture (resp. fermeture) sur la partie claire (resp. sombre).

$$e(k, l) = e_+(k, l) + e_-(k, l) - \epsilon ,$$

$$\Psi_n^2(e)(k, l) = \gamma_n(e_+)(k, l) + \varphi_n(e_-)(k, l) - \epsilon ,$$

où

$$e_+(k, l) = \begin{cases} e(k, l) & \text{si } e(k, l) > \epsilon \\ \epsilon & \text{sinon} \end{cases} ,$$

$$e_-(k, l) = \begin{cases} e(k, l) & \text{si } e(k, l) < \epsilon \\ \epsilon & \text{sinon} \end{cases} ;$$

\*Ce travail a été effectué dans le cadre du réseau européen MAN-ADIX (HCM).



3. ouverture sur la valeur absolue de  $e(k, l)$

$$\Psi_n^3(e)(k, l) = \text{sign}(e(k, l))\gamma_n(|e|)(k, l).$$

Ces trois opérateurs ont été comparés théoriquement et expérimentalement [5].

Après application de l'opérateur morphologique choisi, un seuillage adaptatif de  $N - 1$  seuils est appliqué sur l'image simplifiée  $e_s(k, l)$  de façon à la segmenter. Cette image simplifiée est quantifiée sur  $N$  niveaux de quantification choisis de façon à minimiser la distorsion dans les régions segmentées de l'ECM [7]. Une méthode d'optimisation itérative (algorithme du simplexe [8]) permet de trouver les valeurs des seuils qui minimisent la distorsion sous une contrainte de débit.

### 3. ANALYSE MULTIRÉSOLUTION ET SEGMENTATION DES COEFFICIENTS D'ONDELETTES

Dans cet article nous nous proposons d'effectuer une analyse multirésolution de l'ECM et ensuite de segmenter et coder les coefficients d'ondelettes selon la méthode décrite ci-dessus. La structure complète du codeur devient alors la suivante :

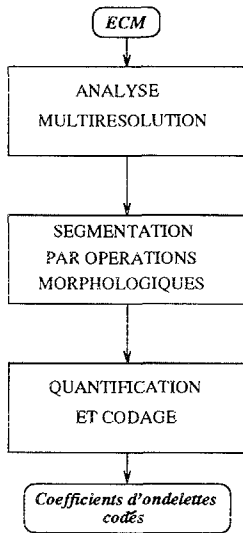


Figure 1 : Structure du codeur.

Cette analyse multirésolution préalable apporte deux avantages :

- Le codage des longueurs des plages implique un balayage monodimensionnel de l'image qui n'est pas justifié pour tout type d'image. Une décomposition en ondelettes séparables produit naturellement une analyse directionnelle de l'image. Il est ensuite possible de filtrer et coder les coefficients d'ondelettes suivant la direction dominante de la sous-bande à laquelle ils appartiennent.
- Un balayage adapté des coefficients d'ondelettes simplifiés permet, dans le cadre d'un codage par plages, d'augmenter la qualité de l'image codée pour un débit donné.

Dans les paragraphes suivants, nous allons décrire d'une manière plus détaillée les étapes d'analyse multirésolution, de segmentation et de codage utilisées pour coder l'ECM.

### 3.1. Analyse multirésolution

L'analyse multirésolution est effectuée à partir d'une décomposition redondante en trame (ou "frame"). Soit  $\{2^{-j/2}\psi(x/2^j - k), (j, k) \in \mathbb{Z}^2\}$  une base orthonormale d'ondelettes issue d'une analyse multirésolution standard de  $L^2(\mathbb{R})$  de fonction d'échelle  $\phi(t)$  [9]. Nous définissons les coefficients de la décomposition en trame sur  $J$  niveaux d'analyse comme suit :

$$\tilde{c}_j^p(k, l) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e(x, y) \psi_p \left( \frac{x-k}{2^j}, \frac{y-l}{2^j} \right) dx dy,$$

$$p \in \{1, 2, 3\}, \text{ si } j \in \{1, \dots, J-1\},$$

$$\text{ou } p \in \{0, 1, 2, 3\}, \text{ si } j = J.$$

$e(x, y)$  représentant dans cette équation un champ 2D de  $L^2(\mathbb{R}^2)$  et

$$\psi_0(x, y) = \phi(x)\phi(y), \quad \psi_1(x, y) = \psi(x)\phi(y),$$

$$\psi_2(x, y) = \phi(x)\psi(y), \quad \psi_3(x, y) = \psi(x)\psi(y).$$

Les coefficients de la décomposition en ondelettes orthogonales séparables de  $e(x, y)$  sont alors obtenus par sous-échantillonnage dyadique, soit  $c_j^p(k, l) = \tilde{c}_j^p(2^j k, 2^j l)$ . Notons que le calcul de  $\tilde{c}_j^p(k, l)$  peut être effectué à l'aide d'un algorithme à trou [10]. L'intérêt de la décomposition préconisée est double. D'une part, elle produit l'analyse directionnelle souhaitée et fournit une reconstruction aisée de l'image à partir de ses coefficients. D'autre part, la redondance introduite permet dans l'étape de simplification d'appliquer des opérations morphologiques avec un élément structurant de taille constante. Dans le cas d'un sous/suréchantillonnage d'une image, il apparaît en effet intuitif qu'il faut contracter/agrandir l'élément structurant pour assurer une simplification comparable à celle obtenue avant le sous/suréchantillonnage. Dans le cas d'un suréchantillonnage d'un facteur 2 par répétition de pixels, on peut montrer une équivalence pour la simplification par les opérations morphologiques ici considérées sous des conditions de taille de l'élément structurant. Par exemple, dans le cas d'un élément structurant de forme  $1 \times (n+1)$  ou  $(n+1) \times 1$ , on a :

$$\Psi_N^i(E)(2k, 2l) = \Psi_n^i(e)(k, l)$$

$$\text{si } N = 2n \text{ ou } N = 2n + 1,$$

où  $i \in \{1, 2, 3\}$ ,  $e$  est l'image d'origine et  $E$  est sa version suréchantillonnée.

La contrainte sur la taille de l'élément structurant pose un problème dans le cas d'une décomposition en base d'ondelettes, car il faut diminuer la taille de l'élément structurant lorsqu'on descend d'un niveau de résolution ce qui devient impossible pour un nombre élevé de niveaux d'analyse.

### 3.2. Segmentation des coefficients d'ondelettes

L'étape de segmentation des coefficients d'ondelettes est effectuée à l'aide des opérations de simplification décrites en 2. L'élément structurant de l'opérateur morphologique est choisi en fonction des coefficients à simplifier et du type de codage appliqué. Le codage monodimensionnel par longueurs de plages et la faible résolution spatiale des

images traitées nous amènent à choisir des éléments structurants  $B_n$  de la forme  $1 \times (n+1)$  (resp.  $(n+1) \times 1$ ). Ceux-ci simplifient l'image dans la direction verticale (resp. horizontale). Sur la Figure 2 sont données à titre d'exemple les directions de simplification pour une décomposition en trame d'ondelettes sur 3 niveaux de résolution.

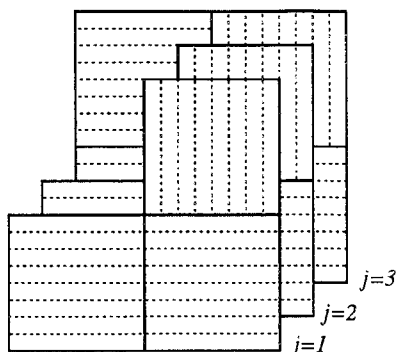


Figure 2 : Directions de simplification pour une décomposition en trame sur 3 niveaux de résolution.

Après l'étape de segmentation, les coefficients d'ondelettes sont sous-échantillonnés pour aboutir à une représentation non-redondante.

### 3.3. Codage

Les coefficients d'ondelette quantifiés sont codés avec un codage de longueurs de plages adapté aux signaux multi-niveaux suivi d'un codage arithmétique avec dictionnaire adaptatif. Ce codage conduit à de forts taux de compression, car dans les images de coefficients quantifiés le fond est important et seules subsistent quelques régions de petite taille. Le codage de longueurs de plages implique un balayage monodimensionnel des coefficients à coder. Nous avons étudié deux façons de réaliser ce balayage :

1. On balaye les sous-bandes suivant leur contenu directionnel. Les sous-bandes correspondant à un niveau de résolution donné sont successivement balayées. La direction de balayage coïncide avec celle de l'élément structurant utilisé lors de la simplification des coefficients.
2. On s'aperçoit que les emplacements des régions dans les sous-bandes d'un même niveau de résolution ne sont pas indépendants. Il est donc avantageux de les coder conjointement.

Pour chaque niveau de résolution  $j$ , on définit comme suit un masque binaire  $m_j$  de la taille des sous-bandes considérées :

$$m_j(k, l) = \begin{cases} 0 & \text{si } \forall p, c_j^p(k, l) = g_{fond} \\ 1 & \text{sinon} \end{cases},$$

où  $g_{fond}$  est le niveau de gris du fond. Ce masque est codé avec un codage de longueurs de plages classique suivi d'un codeur arithmétique à dictionnaire adaptatif. Il indique l'ensemble des emplacements des régions significatives dans les sous-bandes, au niveau de résolution  $j$ . Pour connaître le contenu des sous-bandes il suffit d'indiquer les niveaux de gris des pixels pour lesquels le masque prend la valeur 1.

On procède de la manière suivante :

- On crée une liste  $\mathcal{L}_j$  des niveaux de gris à coder.
- On code  $\mathcal{L}_j$  avec un codeur de longueurs de plages et un codeur arithmétique avec dictionnaire adaptatif.

Quand l'hypothèse de dépendance entre les régions des sous-bandes à un même niveau de résolution est vérifiée, cette seconde approche améliore le taux de compression.

## 4. RÉSULTATS

La méthode décrite ci-dessus a été appliquée à des séquences d'images au format QCIF. Nous présentons ici les résultats obtenus pour la séquence "Claire". La compensation de mouvement a été effectuée à l'aide d'une méthode de mise en correspondance de blocs  $8 \times 8$ , en tolérant un déplacement maximal de 5 pixels dans les directions horizontales et verticales. En raison de la faible résolution spatiale du type d'images traitées, l'élément structurant a été choisi de petite taille ( $2 \times 1$  pour un filtrage horizontal et  $1 \times 2$  pour un filtrage vertical). Pour les mêmes raisons, les réponses impulsionnelles des filtres d'analyse multirésolution sont courtes : nous utilisons ici des filtres de Daubechies de longueur 2 ou 4.

Le Tableau 1 donne les valeurs moyennes du rapport signal à bruit (RSB) crête sur les 30 premières images de la séquence dans les cas mono et multirésolution pour différents valeurs du nombre de niveaux de résolution  $J$  et de la longueur  $N$  des filtres. L'opération morphologique de simplification utilisée dans tous les cas est  $\Psi^3(\cdot)$ . Le débit pour le codage de l'ECM est au plus de 0.08 bit/pixel. Dans le cas de l'analyse multirésolution on compare les résultats obtenus avec les deux méthodes de balayage décrites en 3.3. On voit surtout que l'analyse multirésolution combinée avec le second type de balayage, permet d'améliorer le rapport signal sur bruit obtenu.

-	balayage 1	balayage 2
$N = 2 \quad J = 1$	37.8dB	38.4dB
$N = 4 \quad J = 1$	38.1dB	38.5dB
$N = 2 \quad J = 2$	38.0dB	38.4dB
$N = 4 \quad J = 2$	38.1dB	38.7dB
monorésolution	37.7 dB	

Tableau 1 : Comparaison des différentes méthodes en terme de RSB crête moyen.

Dans les résultats qui suivent, les paramètres  $N = 4$  et  $J = 1$  ont été retenus dans l'approche multirésolution. Sur la Figure 3 sont présentées à titre d'exemple les variations du RSB crête dans les cas mono et multirésolution. L'amélioration due à l'analyse multirésolution apparaît clairement pour toutes les images de la séquence. La Figure 4 indique le débit réel de codage de l'ECM dans le cas multirésolution. On voit que le débit reste stable sur toute la durée de la séquence.

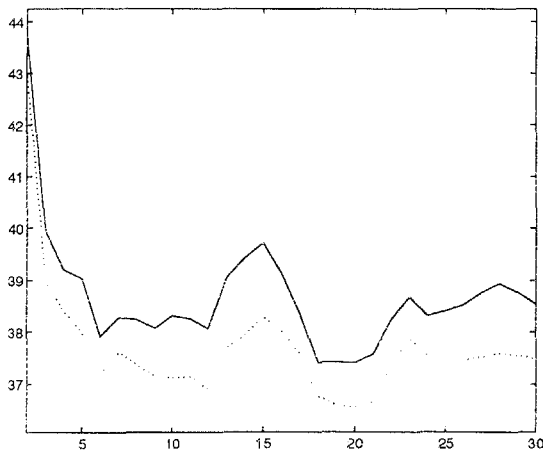


Figure 3 : Comparaison du RSB crête dans le cas multirésolution avec  $N = 4$ ,  $J = 1$  (—) et dans le cas monorésolution (···).

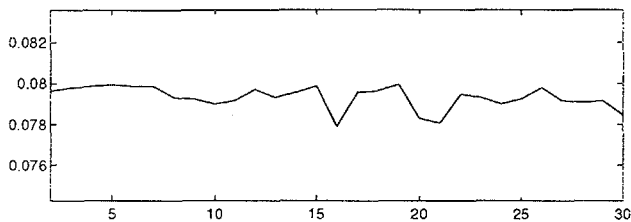


Figure 4 : Débit réel de codage de l'ECM obtenu dans le cas multirésolution avec  $N = 4$ ,  $J = 1$ .

La Figure 5 donne la version codée de la 23ème image de la séquence dans le cas multirésolution et la Figure 6 montre les coefficients d'ondelettes simplifiés et quantifiés pour cette même image.



Figure 5 : Codage de la 23ème image de la séquence dans le cas multirésolution  $N = 4$ ,  $J = 1$ .

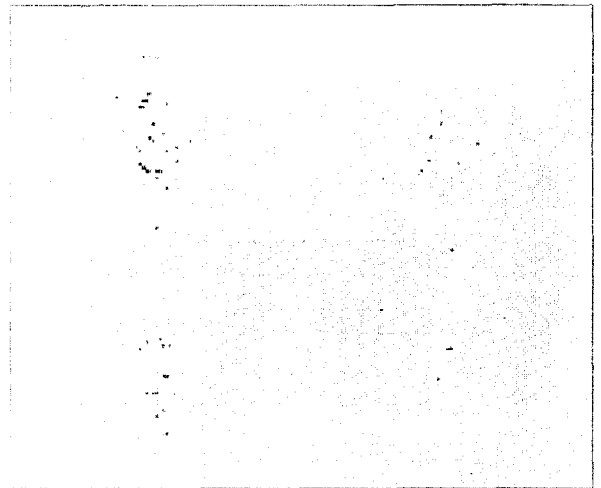


Figure 6 : Coefficients d'ondelettes segmentés lors du codage de la 23ème image ( $N = 4$ ,  $J = 1$ ).

Les résultats obtenus dans ces simulations démontrent l'intérêt d'une analyse en trame d'ondelettes avant la segmentation et le codage de l'ECM.

## 5. RÉFÉRENCES

- [1] G. Tziritas and C. Labit. *Motion analysis for image sequence coding*, volume 4 of *Advances in Image Communication*. Elsevier, 1994.
- [2] J. Serra. *Image analysis and mathematical morphology*, volume 2. Academic Press Ltd., 1988.
- [3] P. Salembier. Morphological multiscale segmentation for image coding. *Signal Processing*, 38:359 – 386, 1994.
- [4] W. Li and M. Kunt. Morphological segmentation applied to displaced frame difference coding. *Signal Processing*, 38:45 – 56, 1994.
- [5] F. J. Hampson and J.-C. Pesquet. A grey scale morphological approach to displaced frame difference coding. In *International Conference on Image Processing*. IEEE, Washington, USA, 1995.
- [6] I. H. Witten, R. M. Neal, and J. G. Cleary. Arithmetic coding for data compression. *Communications of the ACM*, 30(6):520–540, June 1987.
- [7] J. Max. Quantizing for minimum distortion. *IEEE Trans. Inform. Theory*, pages 7 – 12, March 1960.
- [8] J. A. Nelder and R. Mead. A simplex method for function minimization. *Computer Journal*, 7(4):308 – 313, Jan. 1965.
- [9] S. Mallat. Multifrequency channel decompositions of images and wavelet models. *ASSP*, ASSP-37:2091–2110, Dec. 1989.
- [10] M. J. Shensa. The discrete wavelet transform: wedding the à trou and Mallat algorithm. *SP*, 40:2464–2482, Oct. 1992.