



**CODEURS VIDEO DE FAIBLE COMPLEXITE A BASE DE CODAGE
EN SOUS-BANDES, DE COMPENSATION DE MOUVEMENT
HIERARCHISEE ET DE QUANTIFICATION VECTORIELLE.**

**G. Furlan, C. Galand
E. Lançon, C. Raimondo**

J. Menez

Centre d'études et recherches IBM

Le plan du bois
F-06610 La gaude

RÉSUMÉ

LASSY- Université de Nice-Sophia Antipolis
Equipe de l'URA-CNRS 1376
41, Bd Napoléon III
F-06041 Nice cedex

ABSTRACT

Dans cette communication, nous présentons deux types de codeurs vidéo de faible complexité. Le premier est un codeur hybride basé sur des techniques de codage prédictif et de codage en sous-bandes. Il permet d'obtenir des images de bonne qualité à 1.5 Mbps (MPEG/ISO) pour une complexité de l'ordre de 20 Mips. Le second met en oeuvre un procédé simplifié de compensation de mouvement par bloc et une méthode de quantification vectorielle ordonnée, conduisant à une bonne qualité visuelle sur des séquences de type visiophonique, pour un débit binaire de 64 Kbps (recommandation H.261/CCITT), avec une complexité inférieure à 50 Mips.

In this paper, we propose two low complexity video coders schemes. The first one is an hybrid coder based on an association of predictive coding and subband coding techniques. Good quality can be reached at a rate of 1.5 Mbps (MPEG/ISO) for a complexity of about 20 Mips. A simplified motion compensation method and an ordered vector quantization scheme are used in the second coder. In this case, a good quality can be obtained for "head and shoulders" sequences type at a bit rate of 64 Kbps (H.261/CCITT recommendation) for a complexity less than 50 Mips.

1. INTRODUCTION

Avec la rencontre de l'ordinateur, de l'image et du son, une aire nouvelle est en train de naître dans le domaine de la communication: la communication multimédia. Basée sur le "tout numérique", elle ouvre un formidable marché économique sous forme d'applications très diversifiées, comme par exemple, en ce qui concerne l'image, le visiophone, la vidéo-conférence, la HDTV, etc. Mais la représentation sous forme numérique d'images fixes ou animées nécessite des volumes considérables d'informations binaires, à transmettre et à stocker. Et si le développement de réseaux "haut débit" laisse augurer d'un avenir serein en ce qui concerne leur transmission, il reste toujours le problème du stockage. D'où le regain d'intérêt suscité par les techniques de compression, malgré l'apparition de normes et de standards, tels le JPEG de l'ISO pour l'image fixe, le H.261 du CCITT pour le visiophone et le MPEG de l'ISO pour le stockage d'images animées sur support numérique.

Dans cet article nous proposons des systèmes de compression vidéo qui mettent en oeuvre des techniques de codage en sous-bandes, de compensation de mouvement et de quantification vectorielle rapide, avec comme objectif essentiel la réduction du volume de calcul requis par leur mise en oeuvre. Après un rappel des techniques de réduction de la redondance intra-image, inter-images et de quantification vectorielle utilisées, nous présentons deux codeurs vidéo à complexité de calcul réduite. Le premier s'apparente plus particulièrement aux caractéristiques du standard MPEG à 1.5 Mbit/s, le second étant orienté vers une application visiophone à 64 Kbit/s (Recommandation H.261 du CCITT).

2. LES TECHNIQUES UTILISEES

Dans les séquences d'images, deux formes de redondance sont généralement présentes:

- la redondance intra-image,
 - la redondance inter-images,
- dont la réduction passe par l'utilisation de techniques plus ou moins spécifiques.

Réduction de la redondance intra-image [7]: la redondance intra-image est liée à la corrélation, et de façon plus générale à la dépendance statistique, entre pixels voisins d'une même image. Différentes méthodes se proposent de réduire ce type de redondance. Elles font appel soit à des techniques:

- de transformées, telle la DCT,
- de décompositions pyramidales ou de décompositions en sous-bandes,
- de prédiction linéaire mono ou bi-dimensionnelle.

Nous nous sommes plus particulièrement intéressés à la décomposition en sous-bandes à l'aide de bancs de filtres permettant de réaliser une décomposition / recombinaison parfaite ou quasi parfaite.

Il convient de remarquer qu'il ne suffit pas de prendre en compte la perfection du système de décomposition / recombinaison lorsqu'il s'agit de compression avec perte d'information car, dans ce cas, la plus grande partie de la distortion provient en général du bruit de quantification des sous-bandes après passage au travers du banc de filtres de recombinaison.

Réduction de la redondance inter-images [10]: cette redondance est liée au fait qu'entre deux prises d'image successives peu d'éléments ont changé.

Parmi les méthodes permettant de réduire cette redondance, les plus utilisées sont:



- les techniques de prédiction linéaire inter-images,
- la compensation de mouvement,
- et les combinaisons de ces deux méthodes.

Quantification [2] [10]: Si la quantification scalaire (quantificateurs Laplacien ou Gamma) est largement utilisée dans des systèmes de compression à débit moyen, on a de plus en plus recours à la quantification vectorielle pour atteindre des débits binaires inférieurs au bit/pixel. L'inconvénient de ce type de quantificateur est une complexité algorithmique importante, d'où l'intérêt que présente la recherche de méthodes de quantification vectorielle "rapide" et moins complexes.

3. CODEUR VIDEO EN SOUS-BANDES

Initialement proposé et développé pour la compression numérique du signal de parole [3], le codage en sous-bandes a depuis été étendu avec succès à la compression d'images fixes [4] [5] et de séquences d'images [6].

Codeur de G. Karlsson et M. Vetterli: En 1988, G. Karlsson et M. Vetterli [6] proposèrent un schéma de codage en sous-bandes adapté aux signaux vidéo, de nature tridimensionnelle, puisque présentant une dimension temporelle et deux dimensions spatiales. La figure 1 présente le schéma de codage en sous-bandes utilisé par Karlsson et Vetterli.

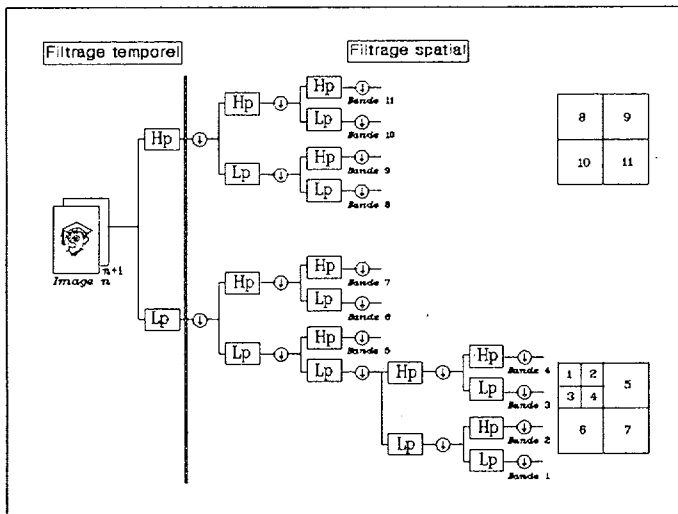


Figure 1. Décomposition en sous-bandes d'un signal tridimensionnel.

Ce codeur nous a permis d'obtenir sur les séquences Miss America et Table tennis, les résultats suivants:

Séquence	Bit/pixel	Débit (Mbit/s)	SNR (dB)
Miss America	0.9	1.4	30
Table tennis	1	1.6	23

Globalement les séquences obtenues à ces débits présentent une qualité qui n'est pas transparente, laissant notam-

ment apparaître une distortion sous forme d'un quadrillage de l'image. Celle-ci est liée à la largeur des bandes de transition des filtres utilisés [7] bien que la décomposition / recombinaison soit parfaite. Des dégradations sont également visibles au niveau des contours.

Modification: Si ce schéma offre un intérêt certain de par sa bonne résistance aux erreurs de transmission, il présente néanmoins l'inconvénient de ne pas réduire de façon efficace la redondance temporelle. En effet, si l'on considère une séquence répétitive comportant plusieurs fois la même image fixe, la décomposition temporelle en sous-bandes de cette séquence donnera naissance à deux séquences: - une "séquence passe-haut" ne comportant que des valeurs à zéro, car le mouvement entre deux images successives identiques est nul, - une "séquence passe-bas" identique à la séquence originale mais sous échantillonnée d'un facteur 2.

On peut donc constater que la redondance temporelle n'est pas correctement éliminée avec un tel procédé, puisqu'il est nécessaire de transmettre une image sur deux, même si N images successives de la séquence originale sont identiques.

Pour pallier cet inconvénient, nous avons remplacé le premier étage de la décomposition par un prédicteur linéaire à un pas, placé dans une boucle (MICD), afin de calculer la différence entre l'image à l'instant (t) et l'image reconstruite à l'instant (t-1). On obtient ainsi une image d'erreur de prédiction temporelle qui caractérise en fait le mouvement existant d'une image par rapport à l'autre. Il est à remarquer que la prédiction inter-images pourrait être effectuée non plus sur deux images successives mais sur trois ou plus, en remplaçant le prédicteur à un pas par un prédicteur à deux pas ou plus.

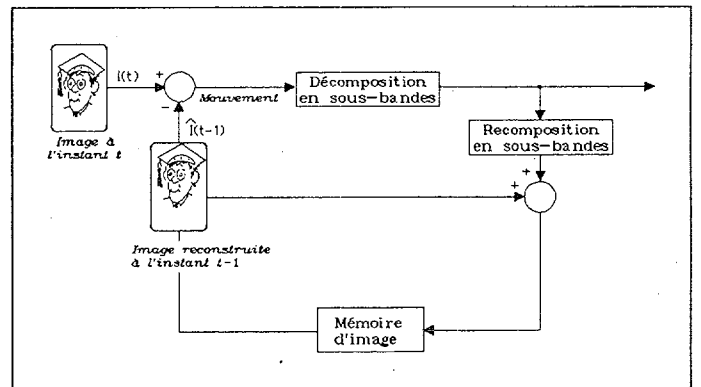


Figure 2. Codeur hybride MICD/SBC.

Cette modification nous a permis d'obtenir les résultats suivants:

Séquence	Bit/pixel	Débit (Mbit/s)	SNR (dB)	Séq. codée
Miss A.	0.8	1.3	37	Msbcs15
T. tennis	1	1.6	32	Tsbcs15

Au niveau de la réduction de la redondance temporelle, la suppression de la décomposition en deux sous-bandes au profit d'une boucle de prédiction, nous a permis d'accroître le rapport signal/bruit d'environ 7 dB par rapport aux résultats offerts par le codeur de G. Karlsson et M. Vetterli. Cette

méthode a également apporté une nette amélioration de la qualité visuelle des séquences.

Nous avons réalisé à partir des travaux de G. Karlsson et M. Vetterli [6], une estimation de la complexité de calcul associée à cette technique qui a montré qu'une vingtaine de Mips suffisaient à sa mise en oeuvre en temps réel [11].

Si le codeur que nous proposons, offre des séquences reconstruites de bonne qualité pour une faible complexité de calcul, il ne permet pas d'atteindre de faibles débits binaires de transmission, tels ceux requis par la recommandation H.261 du CCITT par exemple. Il faut alors faire appel à d'autres techniques que nous présentons dans le paragraphe suivant.

4. CODEUR VIDEO A 64 KBPS

Dans le schéma de codage précédent, nous avons présenté la méthode de réduction de la redondance temporelle par prédiction linéaire. De plus forts taux de compression peuvent être obtenus par l'utilisation d'algorithmes de détection d'activité et de compensation de mouvement [10] dans un système à prédiction. En effet, dans les séquences d'images de type "head & shoulders" (une ou plusieurs personnes présentées en buste face à une caméra fixe), utilisées en visiophonie, seules quelques zones varient à l'intérieur de chaque image, alors que l'essentiel de la scène reste inchangé. Les pixels, ou les zones de pixels, qui restent inchangés d'une image sur l'autre, présentent une information redondante qu'il est donc inutile de transmettre.

La structure d'un système de codage par compensation de mouvement est la suivante :

1. un **détecteur d'activité** qui distingue les blocs qui ont bougé. En effet il s'agit de classer les blocs en deux catégories : actifs - inactifs. Ceci permettra à l'estimateur de déplacement de ne travailler que sur un nombre restreint de blocs.
2. un **estimateur de déplacement** qui prédit les vecteurs "déplacement" des blocs actifs. Il s'agit ainsi de trouver la meilleure correspondance entre le bloc de référence contenu dans l'image à l'instant t et l'aire de recherche la mieux adaptée de l'image à l'instant $(t-1)$.
3. une **compensation de mouvement** qui permet de générer l'image à l'instant (t) . Cette fonction utilise la localisation des blocs actifs, la connaissance des vecteurs "déplacement" qui leur sont associés et l'image à l'instant $(t-1)$.

Compensation de mouvement hiérarchisée [9]: Cette technique, basée sur la décomposition en sous-bandes, permet de réduire le volume de calcul associé à la compensation de mouvement.

Il s'agit dans un premier temps de décomposer l'image originale $I(t)$ en 4 sous-bandes et de stocker l'image basse fréquence $B(t)$. Dans un second temps on étend la décomposition de l'image originale à 7 sous-bandes, et on stocke l'image basse fréquence $BB(t)$.

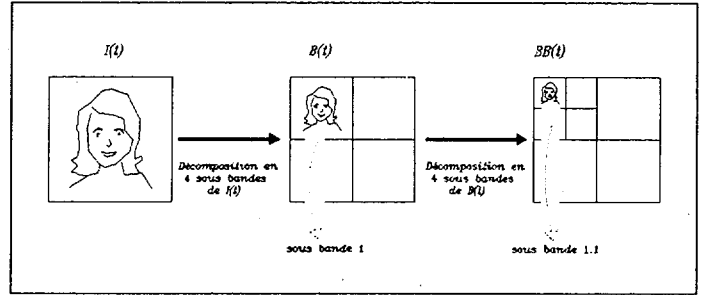


Figure 3. Principe de la décomposition

On effectue la détection d'activité et une première estimation de déplacement entre les images de sous-bandes $BB(t)$ et $BB(t-1)$. Puis on raffine l'estimation de déplacement sur des résolutions croissantes de l'image ($B(t)$ et $I(t)$). On construit alors l'image compensée en mouvement à l'instant t , $\hat{I}(t)$, à partir de l'image $\hat{I}(t-1)$, des coordonnées des blocs actifs et des vecteurs "déplacement". Il nous faut maintenant quantifier l'erreur de prédiction de compensation de mouvement.

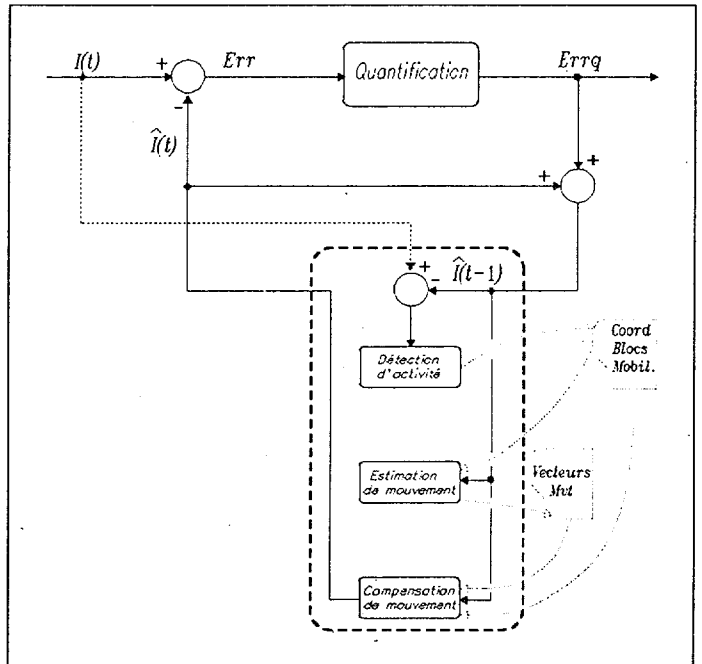


Figure 4. Schéma du codeur proposé

Quantification vectorielle ordonnée [10] [11]: Afin de déterminer le vecteur du répertoire le plus proche du vecteur à coder, on n'effectue plus une recherche systématique ("full-search") parmi tous les vecteurs représentatifs, mais on réalise la comparaison sur un sous-ensemble des vecteurs du répertoire.

La quantification vectorielle ordonnée s'effectue en deux étapes:

- **le pré-traitement**: il n'intervient pas dans l'évaluation de la complexité de l'algorithme et consiste à réordonner tous les vecteurs C_i du répertoire par distance croissante d_i par rapport à un vecteur V_{ref} . On a donc introduit la notion de relation d'ordre entre les vecteurs représentatifs du répertoire.

- **le traitement**: il s'agit de déterminer le vecteur du répertoire ordonné le plus proche du vecteur à coder. On commence par calculer la distance d entre le vecteur à coder et V_{ref} . Ensuite on cherche à déterminer l'indice i tel que $d_{i+1} < d < d_i$. A partir de l'indice i on recherche dans le voisin-



nage immédiatement à gauche s'il existe un vecteur C_g plus proche du vecteur à coder que C_i au sens de la minimisation du critère des moindres carrés. On recherche selon le même principe, dans le voisinage immédiatement à droite de $i + 1$, s'il existe un vecteur C_r plus proche du vecteur à coder que C_{i+1} . Un test, basé sur le principe de l'inégalité triangulaire, nous permet de déterminer à quel moment il faut arrêter la recherche à gauche ou à droite. Il ne reste plus qu'à choisir le vecteur C_g , issu de la recherche à gauche et le vecteur C_r , issu de la recherche à droite, celui qui minimise le critère des moindres carrés par rapport au vecteur à coder.

Cette technique est optimale dans le sens où on obtient 100% des vecteurs identiques à ceux que l'on aurait obtenu si l'on avait effectué une recherche systématique ("full-search").

La réduction de complexité est de 55 à 80% suivant la séquence traitée. Ces chiffres correspondent à un codage à 64 Kbps utilisant 256 vecteurs de taille 16.

5. CONCLUSION

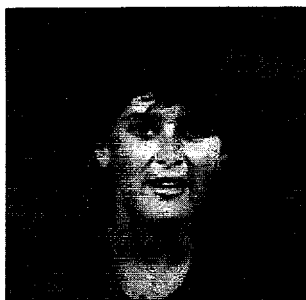
Dans cet article nous avons proposé deux systèmes de compression vidéo performants.

Le premier vise des applications de type MPEG et permet d'obtenir, grâce à l'utilisation de techniques de codage en sous-bandes, des séquences de très bonne qualité visuelle à 1.5 Mbit/s, ceci quel que soit le taux d'activité qu'elles présentent.

Le second vise plus particulièrement des applications de type visiophone et donc des séquences présentant un faible taux d'activité. Il permet d'obtenir, grâce à l'utilisation de techniques de compensation de mouvement hiérarchisée et de quantification vectorielle ordonnée des séquences de bonne qualité visuelle à 64 Kbps pour une complexité de calcul réduite.

Le tableau ci-dessous résume les performances de ces deux systèmes.

Séquence	Débit	Complexité	SNR (dB)
Miss A.	64 Kbps	45 Mips	36
Miss A.	1.5 Mbps	20 Mips	37
Ten.table	64 Kbps	23 Mips	27
Ten.table	1.6 Mbps	20 Mips	32



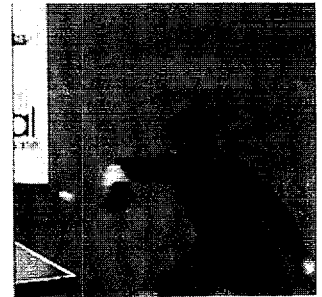
Miss américa originale



Msbcb15



Table tennis originale



Tsbc15



Débit = 64 Kbps Format = 180x144 pixels



BIBLIOGRAPHIE

- [1] R. Forchheimer & T. Kronander, "Image coding - From waveforms to animation", IEEE transactions on ASSP, Vol.37, No 12, Décembre 1989
- [2] R.M. Gray, "Vector quantization", IEEE ASSP Magazine, Vol.1, Avril 1984
- [3] C. Galand, "Codage en sous-bande: théorie et application à la compression numérique du signal de parole", Thèse d'état, Université de Nice, Mars 1983
- [4] M. Vetterli, "Multi-dimensionnal subband coding: some theory and algorithms", Signal Processing No 6, 1984
- [5] J.W. Woods & S.D. O'Neil, "Subband coding of images", IEEE transactions on ASSP, Vol.34, No 5, Octobre 1986
- [6] G. Karlsson & M. Vetterli, "Three dimensionnal sub-band coding of video", Conf. Icassp, pp.1100-1103, 1988
- [7] G. Furlan, "Contribution à l'étude et au développement d'algorithmes de traitement du signal en compression de données et images", Thèse de doctorat, Université de Nice - Sophia Antipolis, Avril 1990
- [8] A. Gersho & B. Ramamurthi, "Image coding using vector quantization", IEEE Proceedings ICASSP, Avril 1982
- [9] A. Von Brandt, "Motion estimation and subband coding using quadrature mirror filters", Eurasip 1986
- [10] C. Raimondo, C. Galand, E. Goubault, E. Lançon & J. Menez, "Low bit-rate hybrid coding using hierarchical motion compensation and low complexity vector quantization", Conf. Icassp, Toronto, Mai 1991
- [11] C. Raimondo, "Contribution à l'étude et au développement de méthodes de compression numérique de séquences d'images", Thèse de doctorat, Université de Nice - Sophia Antipolis, Mai 1991