

QUANTIFICATION VECTORIELLE ALGEBRIQUE ET ARBORESCENTE : METHODE D'ACTUALISATION DU DICTIONNAIRE. APPLICATION AU CODAGE VIDEO.

*Vincent Ricordel †, Moncef Gabbouj ‡ et Claude Labit **

†MS-GESSY/ISITV, Université de Toulon et du Var,
Av. Georges Pompidou BP56, 83162 La Valette du Var, FRANCE

‡DMI/SPL, Tampere University of Technology,
P.O. Box 553, 33101-Tampere, FINLANDE

* IRISA/INRIA Rennes, Campus de Beaulieu
35042 Rennes Cedex, FRANCE

ricordel@univ-tln.fr, gabbouj@cs.tut.fi, labit@irisa.fr

Résumé

L'objet de cet article est d'introduire une nouvelle méthode de quantification vectorielle (QV¹) adaptative pour la compression à bas débits de séquences d'images numériques. Nos travaux précédents [13, 14, 12] décrivent un schéma de QV algébrique et arborescente (QVAA) basée sur une technique d'emboîtements successifs de réseaux tronqués, et aboutissant à une QV multi-étages. A présent nous explorons les aptitudes de ce QVAA au codage adaptatif via un schéma mettant à profit sa structure arborescente. Le dictionnaire est alors composé de deux parties : une "souche" permanente constituée hors-ligne relativement à une large gamme de séquences d'images. De branches ajoutées pour chaque nouveau type de séquences à coder.

Notre QV adaptatif prenant place au sein d'un codeur de type MPEG, nous analysons les résultats expérimentaux obtenus lors du codage de sources réelles.

1 INTRODUCTION

Les codages prédictifs et en sous-bandes [11] sont les deux familles de méthodes incontournables pour la compression de séquences d'images numériques. Car un codage

¹QV signifiera aussi bien Quantification Vectorielle que Quantificateur Vectoriel

hybride vise précisément à ne transmettre ni l'information prédictible, ni l'information imperceptible par le système visuel humain (SVH). La figure 1 montre alors le codeur générique au sein duquel prend place notre QVAA adaptatif. Remarquez que nous ne considérons pas l'aspect de la quantification de l'information de mouvement obtenue après estimation-compensation [16], mais que celui de la QV des erreurs de prédiction transformées.

Les étapes de décorrélation inter-intra images conduisent à une mise en forme du signal à quantifier, car au sein de chaque sous-images les données sont classées en fonction de directions fréquentielles privilégiées et éventuellement à différentes échelles si l'analyse est multirésolution. Une QV séparée de chaque sous-image est appropriée afin de tirer profit des corrélations demeurant entre les coefficients transformés [3], et la stratégie d'allocation binaire est produite telle que, tenant compte des propriétés du SVH, les "basses fréquences" reçoivent plus de bits (mais ce gain escompté est subjectif). La modélisation de la distribution statistique d'une sous-image est communément faite à l'aide d'une fonction Gaussienne Généralisée dont l'étroitesse souligne les corrélations existantes entre les coefficients [3]. Cependant la source vectorielle obtenue est toujours non-stationnaire [6, 1]. Seule une QV adaptative [9, 7], intégrant un procédé d'actualisation rapide des représentants du dictionnaire, est capable de s'adapter aux changements statistiques de la source au fur et à mesure du codage des séquences d'images.

2 RESUME DE L'ETUDE PRECEDENTE

2.1 Contexte : quantification vectorielle algébrique

La QV algébrique (QVA) [5, 3] a été introduite avec succès afin de réduire la complexité inhérente aux techniques de QV de type LBG [8]. L'encodage repose sur des opérations d'arrondi et de changements d'échelle, et est donc indépendant de la taille du dictionnaire. Il n'y a plus de recherche exhaustive à effectuer au sein de ce dernier, et pratiquement aucune norme n'est calculée. Du fait de la structure prédéfinie des réseaux algébriques, il n'y a pas à transmettre de dictionnaire et aucune procédure d'apprentissage n'est requise pour sa construction. Cependant un réseau ne peut quantifier optimalement qu'une source uniforme et, du fait de sa taille infinie, il doit être tronqué afin d'indexer les code-vecteurs. Le choix de la métrique L_2 [10, 2] (resp. L_1 [5, 10, 3]) permet d'effectuer une troncature de forme hypersphérique (resp. hypercubique) et de compter les points. Toutefois le QVA ainsi construit n'est optimal que pour la quantification de sources symétriques ayant une distribution Gaussienne (resp. Laplacienne) telle qu'elles recouvrent exactement le dictionnaire. Cette modélisation restrictive de la source est à la base de méthodes judicieuses pour achever le QVA. Toutefois ces conditions favorables disparaissent si l'on considère le codage à bas débit de sources complexes. Pour répondre à ces limites, nous avons proposé la QVAA .

2.2 Quantification vectorielle algébrique et arborescente

Considérant les réseaux algébriques pour lesquels Conway et Sloane ont déterminé des algorithmes de quantification rapide [4], la QVAA repose sur une technique d'emboîtements successifs de réseaux tronqués de même nature. Dans [13, 14, 12], nous détaillons les différentes étapes de sa conception avec : la troncature du réseau et l'emboîtement, la mise en place d'un QV multi-étages, le découpage du dictionnaire arborescent résultant et la détermination du réseau optimal relativement à cette méthode. La détection et le traitement des probables vecteurs hors-norme sont aussi abordés.

La QVAA offre des solutions originales aux limites de la QVA avec une découpe de l'espace adaptée à la distribution de sources complexes et fonction d'un critère débit-distorsion, un indexage basé sur la structure arborescente du dictionnaire, et une détection et un traitement simple des vecteurs marginaux. Le dictionnaire obtenu automatiquement est particulièrement adapté au codage de sources différentielles.

2.3 Méthode d'allocation binaire

Le problème de l'allocation binaire survient lorsqu'il faut partager les ressources binaires entre les différentes sous-images [15]. Nous nous proposons de présenter sommairement la solution apportée dans le cas particulier de la QVAA.

Il s'agit précisément de minimiser la distorsion D sous la contrainte que le débit global R demeure inférieur à un seuil R_d . La transformée est considérée orthogonale et il y a M sous-images, nous voulons donc :

$$\min D = \min \sum_{j=0}^{M-1} d_{j,i} \text{ avec } R = \sum_{j=0}^{M-1} r_{j,i} \leq R_d$$

A chaque sous-bande j , un ensemble de quantificateurs $q_{j,i}$ potentiels est associé (chaque $q_{j,i}$ est une configuration possible de l'arbre du QVAA), $d_{j,i}$ est la distorsion correspondant au débit $r_{j,i}$. Pour une combinaison de M quantificateurs (cad un QVAA particulier pour chaque sous-image) nous obtenons un point (R, D) dans l'espace distorsion vs. débit, et l'ensemble des N^M combinaisons produit un nuage de points (la figure 2 montre un exemple). Le problème devient la détermination sur l'enveloppe convexe de ce nuage, du point dont le débit est juste inférieur à R_d . Afin de réduire la complexité opératoire le Multiplicateur de Lagrange λ est introduit, il faut à présent résoudre :

$$\min(D + \lambda.R) \iff \sum_{j=0}^{M-1} \min_{q_{j,i}} (d_{j,i} + \lambda.r_{j,i})$$

La complexité décroît car la réduction de distorsion est effectuée séparément pour chaque sous-bande. La forme générale de l'algorithme est :

1. L'enveloppe convexe relative à chacune des sous-images est calculée. Pour la QVAA, elle est obtenue directement en faisant croître l'arbre de quantification.

2. Le point sur l'enveloppe convexe globale est déterminée, son débit est juste inférieur à R_d .

Pour cette dernière étape nous appliquons la méthode introduite par Shoham [15]. Elle repose sur le calcul de valeurs singulières de λ (cad les pentes des droites passant au travers les points consécutifs de l'enveloppe convexe). Aussi, à partir d'un premier point de l'enveloppe, par les calculs successifs des valeurs singulières, nous obtenons l'enveloppe globale. La figure 2a illustre un résultat expérimental avec des QVAA. Une défaut apparaît car de larges écarts subsistent entre des points de cette enveloppe globale. Aussi nous modifions l'algorithme afin d'accéder aux "quantificateurs optimaux" (voir la figure 2), cad les points juste au dessus de l'enveloppe convexe. Nous procédons en deux étapes :

1. L'algorithme précédent permet de déterminer les deux points de l'enveloppe convexe encadrant R_d .
2. A partir de ces deux points, l'algorithme de Shoham [15] est mis en oeuvre à nouveau afin d'obtenir une portion locale d'enveloppe convexe .

La courbe intitulée "quantificateurs intermédiaires" de la figure 2b est tracée en calculant les portions d'enveloppes convexes locales entre chaque bipoints de l'enveloppe convexe globale. On peut constater qu'un grand nombre de quantificateurs optimaux sont atteints.

3 CONCEPTION DU QV ADAPTATIF

Le dictionnaire du QVAA adaptatif est construit, par apprentissage, en deux parties :

1. Une racine permanente conçue hors-ligne et relativement à une séquence d'apprentissage de grande taille représentative d'une large gamme de séquences d'images.
2. De branches ajoutées pour chaque nouveau type de séquence à coder. Ces branches forment la seule information à transmettre au décodeur afin d'actualiser le dictionnaire.

Etant donné que l'allocation binaire nécessite une étape d'élagage de l'arbre de quantification, il y a quatre phases dans l'algorithme (voir le schéma de la figure 3) :

1. Construction d'une racine de grande taille.
2. Première allocation binaire pour achever la racine (le seuil R_d correspondant est choisi très bas).
3. Addition de branches de grandes tailles.
4. Allocation binaire finale pour l'arbre complet avec le débit binaire cible.

Pour réactualiser le dictionnaire seules les étapes 3 et 4 doivent être mises en oeuvre. La méthode pour faire croître le dictionnaire arborescent est décrite dans [13].

4 RESULTATS EXPERIMENTAUX

Le codeur (voir figure 1) est basé MPEG car les outils mis en oeuvre sont une mise en correspondance de blocs pour l'estimation du mouvement, et une DCT pour la transformée. La taille des blocs est 2×2 , aussi quatre sous-images sont-elles constituées (voir la figure 4). Les simulations sont faites avec des séquences d'images QCIF, et concernant les temps CPU indiqués, la machine cible est une SparcStation 20 (75 Mhz).

Les images de quatre séquences différentes sont utilisées pour la construction de la racine, mais pour l'ajout de branches, l'apprentissage est effectué à partir des images d'une séquence unique (voir le tableau 3 pour des résultats numériques typiques). Cette même séquence est ensuite encodée (voir le tableau 1). Le temps CPU d'encodage d'une image est d'environ 1.8 s.

A titre de comparaison le tableau 4 montre les résultats relatifs à la construction d'un dictionnaire d'un QVAA classique, et le tableau 2 les résultats correspondant lors de l'encodage de la séquence.

5 CONCLUSION

Le schéma de QVAA adaptatif offre, relativement aux résultats de nos précédents travaux, une plus grande précision dans l'allocation binaire, et une meilleure régularité concernant la qualité de reconstruction des séquences d'images codées-décodées. La somme d'information à transmettre pour réactualiser le dictionnaire est restreinte, et les temps CPU de construction ou de mise à jour du dictionnaire sont courts.

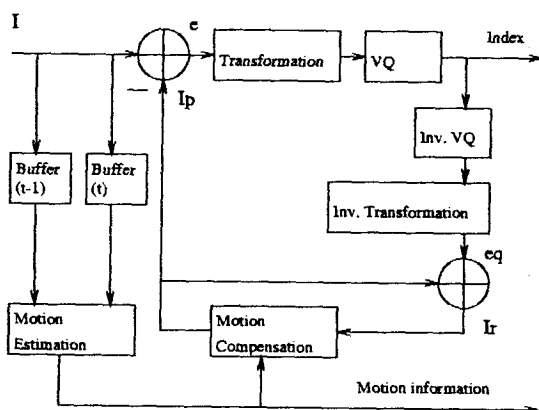


FIG. 1: Codeur hybride. I : image source ; I_p : image prédite ; I_r : image reconstruite ; e : erreurs de prédiction ; e_q : erreurs de prédiction quantifiées.

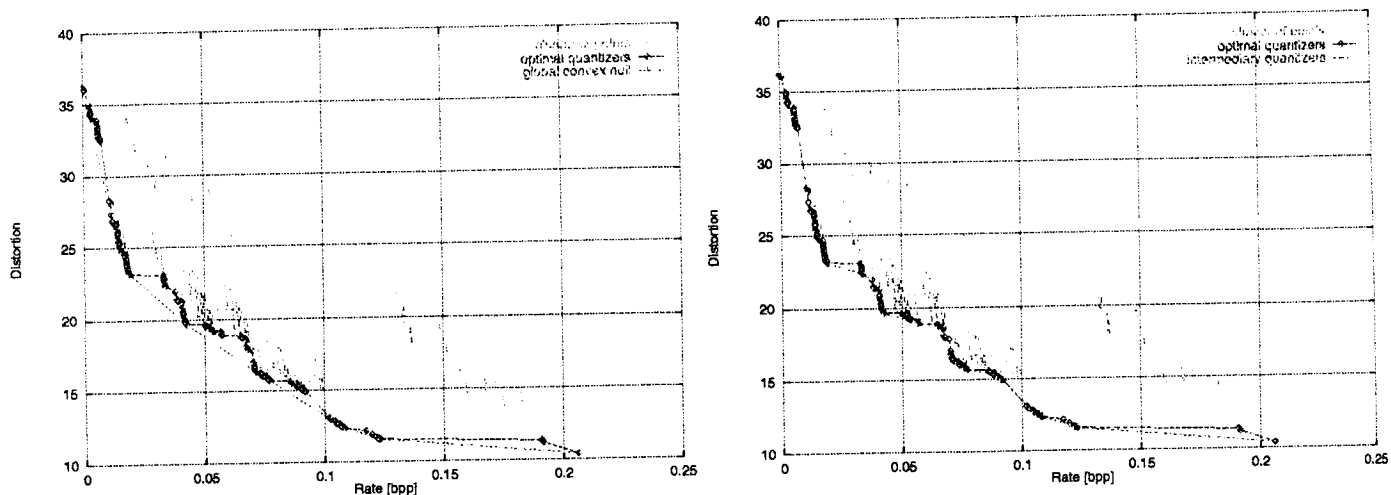


FIG. 2: [a,b] : Nuages de points (R, D) et enveloppes calculées pour l'allocation binaire.

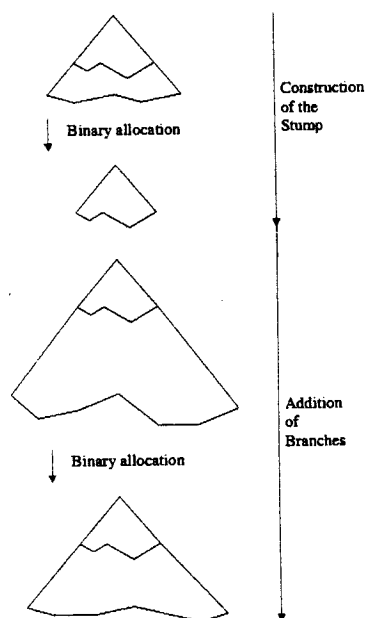


FIG. 3: QVAA adaptatif : les 4 étapes de la construction du dictionnaire.

□	1 bpp	□□	.5 bpp
	A		B
□□	5 bpp	□□□□	2 bpp
	C		D

FIG. 4: Dictionnaire : forme des vecteurs, entropies maximales avant allocation binaire et étiquettes des sous-images.

nom de la séquence	nombre d'images	entropie moyenne [bpp]	PSNR moyen [dB]
"Miss America"	150	0.273	38.03
"Salesman"	180	0.170	36.66
"Carphone"	180	0.257	32.65

TAB. 1: Codage de séquences avec le QVAA adaptatif.

nom de la séquence	nombre d'images	entropie moyenne [bpp]	PSNR moyen [dB]
"Miss America"	150	0.134	36.87
"Salesman"	180	0.177	36.58
"Carphone"	180	0.381	34.12

TAB. 2: Codage de séquences avec le QVAA classique.

construction de la racine à l'aide de 4 séquences							
étiquette de la sous-bande	avant allocation binaire			après allocation binaire ($R_d = 0.1\text{bpp}$)			
	taille séquence d'apprentissage [images]	temps CPU construction [s]	nombre de représentants	entropie du dictionnaire [bpp]	nombre de représentants	entropie du dictionnaire [bpp]	taille du dictionnaire [octets]
A	4	4.93	55	0.698	20	0.332	238
B	12	6.15	106	0.213	26	0.020	259
C	12	6.22	133	0.231	34	0.027	315
D	200	69.43	1982	0.083	92	0.003	810

addition de branches à l'aide de la séquence "Salesman"							
étiquette de la sous-image	avant allocation binaire			après allocation binaire ($R_d = 0.2\text{bpp}$)			
	taille séquence d'apprentissage [images]	temps CPU construction [s]	nombre de représentants	entropie du dictionnaire [bpp]	nombre de représentants	entropie du dictionnaire [bpp]	taille information additionnelle [octets]
A	5	7.18	48	0.954	20	0.337	0
B	10	5.97	104	0.205	104	0.205	681
C	10	6.11	161	0.318	86	0.253	448
D	150	52.92	1348	0.078	92	0.002	0

TAB. 3: Construction du dictionnaire avec le QVAA adaptatif : résultats numériques.

étiquette de la sous-bande	avant allocation binaire			après allocation binaire ($R_d = 0.20\text{bpp}$)			
	taille séquence d'apprentissage [images]	temps CPU construction [s]	nombre de représentants	entropie du dictionnaire [bpp]	nombre de représentants	entropie du dictionnaire [bpp]	taille du dictionnaire [octets]
A	5	7.15	27	0.964	21	0.420	244
B	10	5.48	137	0.258	37	0.033	336
C	10	5.50	148	0.281	148	0.281	1358
D	150	46.75	1248	0.087	529	0.066	4995

TAB. 4: Construction du dictionnaire avec le QVAA classique et à l'aide de la séquence "Salesman" : résultats numériques.

Références

- [1] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies. Image coding using wavelet transform. *IEEE Transactions on Image Processing*, 2, avril 1992.
- [2] M Barlaud. *Wavelets in image communication*. Elsevier, New York, 1994.
- [3] M. Barlaud, P. Solé, T. Gaidon, M. Antonini, and P. Mathieu. Pyramidal lattice vector quantization for multiscale image coding. *IEEE Transactions on Image Processing*, 3(4) :367–381, juillet 1994.
- [4] J.H. Conway and Sloane N.J.A. Fast quantizing and decoding algorithms for lattice quantizers and codes. *IEEE Transactions on Information Theory*, 28(2) :227–232, mars 1982.
- [5] T.R. Fisher. A pyramid vector quantizer. *IEEE Transactions on Information Theory*, 32(4) :568–583, juillet 1986.
- [6] T.R. Fisher. Geometric source coding and vector quantization. *IEEE Transactions on Information Theory*, 35(1) :137–145, janvier 1989.
- [7] J.E. Fowler. *Adaptive Vector Quantization for the Coding of Nonstationary Sources*. PhD thesis, The Ohio State University, 1996.
- [8] A. Gersho and R.M. Gray. *Vector Quantization and Signal Compression*. Kluwer Academic Publishers, Boston, 1992.
- [9] M. Goldberg, P.R. Boucher, and S. Schlien. Image compression using adaptative vector quantization. *IEEE Transactions on Communications*, 34(2) :180–187, février 1986.
- [10] D.G. Jeong and J.D. Gibson. Uniform and piecewise uniform lattice vector quantization for memoryless gaussian and laplacian sources. *IEEE Transactions on Information Theory*, 39(3), mai 1993.
- [11] A.N. Netravali and B.G. Haskell. *Digital pictures : Representation and Compression*. Plenum Press, New York, 1988.
- [12] V. Ricordel. *Etude de schémas de quantification vectorielle algébrique et arborescente. Application à la compression de séquences d'images numérique*. PhD thesis, Université de Rennes I, 1996.
- [13] V. Ricordel and C Labit. Vector quantization by packing of embedded truncated lattices. In *Proc. of International Conference on Image Processing ICIP*, volume 3, pages 292–295. Washington DC, USA, octobre 1995.
- [14] V. Ricordel and C Labit. Tree-structured lattice vector quantization. In *Proc. of European Signal Processing Conference EUSIPCO*, volume 2, pages 731–734. Trieste, Italie, septembre 1996.
- [15] Y. Shoham and A. Gersho. Efficient bit allocation for an arbitrary set of quantizers. *IEEE Transactions on Acoust. Speech Signal Processing*, 36(9) :1445–1453, septembre 1988.
- [16] G. Tziritas and C. Labit. *Motion analysis for image sequence coding*, volume 4 of *Advances in image communication*. Elsevier Science Publishers, Londres, juillet 1994.