

Codage de vecteurs mouvement par compétition de prédicteurs spatio-temporels dans le standard H.264

G. Laroche^{1,2}

J. Jung¹

B. Pesquet-Popescu²

¹MAPS/MDG/SIA

France Telecom R&D
38-40 rue du Général Leclerc,
92794 Issy Les Moulineaux

² Signal and Image Proc. Department

ENST Paris
46 rue Barrault, 75634 Paris

{guillaume.laroche, joelb.jung}@orange-ft.com

{beatrice.pesquet}@enst.fr

Résumé

Le nouveau standard vidéo H.264/MPEG4-AVC permet une réduction significative du débit par rapport à ses prédécesseurs. Les performances obtenues sur le codage de la texture ainsi que la compensation de mouvement sub-pixelique ont contribué à augmenter la proportion du débit de l'information allouée au mouvement. Le nouvel objectif de l'ITU-T est de concevoir un codec vidéo réduisant le débit de 50% par rapport au standard H.264. Ce prochain codec augmentera certainement la proportion de l'information de mouvement. Par conséquent la réduction du débit de cette information devient un sujet de recherche essentiel pour le codage vidéo. Dans cet article, une méthode par compétition de prédicteurs spatio-temporels est proposée. La sélection s'effectue à l'aide d'un critère débit-distorsion prenant en compte la nouvelle information liée au mouvement. Cette méthode tire partie des redondances temporelles des champs de vecteurs, non exploitées par le médian spatial présent dans le standard. Les réductions de débit obtenues par rapport au codec H.264/MPEG4-AVC atteignent 20% pour des séquences complexes.

Mots clefs

Codage de vecteurs mouvement, H.264, prédiction spatio-temporelle, compétition, critère débit-distorsion.

1 Introduction

Le récent standard de l'ITU-T/ISO-IEC, H.264 [1], appelé MPEG4-AVC par l'ISO/IEC, obtient un gain de compression significatif par rapport à ses prédécesseurs H.263 et MPEG-4 part 2. Ce gain provient de l'amélioration des outils existants et de l'introduction de nouveaux outils dont les plus importants sont : les multiples prédicteurs Intra, les partitions variables des macroblocs, la compensation de mouvement au $\frac{1}{4}$ de pixel, le "deblocking filter" contenu dans la boucle de codage et le codage arithmétique adapté au contexte (CABAC). De plus un travail conséquent a été entrepris sur le codec de référence [2]. Basé sur un critère

de débit-distorsion (RD), il offre ainsi les choix optimaux parmi la multitude des modes de codage possibles. Cette réduction efficace du débit affecté au codage de la texture entraîne une baisse du débit total de H.264 mais augmente la proportion de l'information de mouvement dans ce débit. En effet, pour les bas débits cette proportion atteint 40%.

Le Video Coding Expert Group (VCEG/ITU-T SG16 Q6) atteindra probablement son objectif de normaliser un codec obtenant un gain de compression de 50% par rapport à la référence H.264. Ce prochain standard augmentera certainement la proportion de l'information de mouvement, à l'aide d'une compensation de mouvement beaucoup plus fine et d'une forte diminution du débit des résiduels de bloc. Nous avons donc porté notre attention sur la réduction de cette information de mouvement.

Ce problème a déjà été abordé dans la littérature. Dans cet article seules les méthodes de codage sans perte des vecteurs mouvement, beaucoup plus répandues, sont décrites. Toutefois notons qu'il existe des méthodes de codage avec pertes [3]. L'exploitation de redondances temporelles par prédiction entre deux champs de vecteurs a déjà été entreprise dans [4]. Cette méthode donne de bons résultats pour des séquences contenant des mouvements complexes. Cependant la prédiction temporelle est moins efficace qu'une prédiction spatiale seule lorsqu'un ensemble de vidéos représentatives est utilisé. Dans [5] un prédicteur spatial et un prédicteur temporel sont employés, mais le choix entre ces deux prédicteurs dépend uniquement de la valeur des prédicteurs, ce qui ne constitue pas une méthode par compétition. Dans [6], une sélection entre des prédictions spatiales et spatio-temporelles est proposée, mais uniquement au niveau *slice*. Cette approche est une méthode par compétition. En effet, le meilleur prédicteur, parmi un ensemble de prédicteurs, est sélectionné à l'aide d'un critère d'efficacité de codage et l'indice correspondant à ce prédicteur est codé. Dans [7] une compétition au niveau vecteur est explicitée. Cependant les redondances temporelles ne sont pas exploitées.

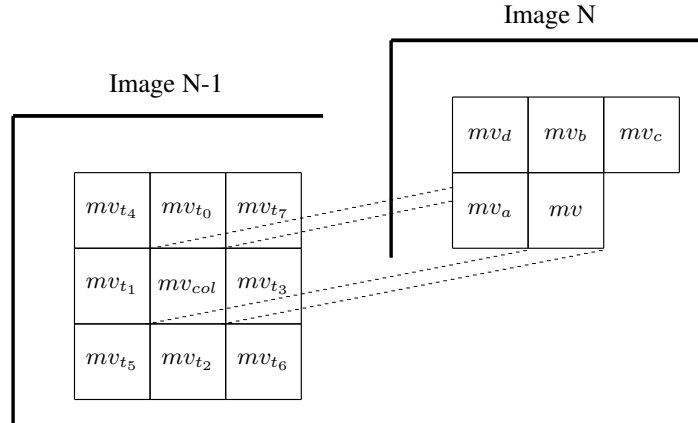


Figure 1 – Localisation des vecteurs spatiaux et temporels utilisés pour la prédiction.

Dans cet article nous proposons deux nouvelles techniques pour améliorer ces méthodes. Dans un premier temps, une compétition de prédicteurs spatio-temporels de vecteurs mouvement est introduite, avec la modification du critère de débit-distorsion associée. Puis l'introduction d'un vecteur conditionnel spatio-temporel pour le mode SKIP est explicité. L'ensemble de ces modifications a été implémenté dans le codec de référence du standard H.264.

Cet article est organisé de la manière suivante : un résumé sur le codage des vecteurs mouvement dans H.264 est présenté dans la section 2. Les modifications apportées à ce codage et la proposition d'un nouveau vecteur pour le mode SKIP sont décrites dans la section 3. Enfin la section 4 commente brièvement l'impact des méthodes proposées sur la complexité, et présente les gains de compression situés en moyenne à 4% pour une qualité équivalente, au meilleur profil d'H.264.

2 Etat de l'art du codage des vecteurs mouvement

Tout d'abord décrivons les procédés non-normatifs pour la sélection des vecteurs mouvement, implémentés dans le codec de référence de H.264. Le standard H.264 applique un codage prédictif des vecteurs mouvement, et le résiduel de ce vecteur ε_{mv} est donné par la formule suivante :

$$\varepsilon_{mv} = mv - p \quad (1)$$

dans laquelle mv est le vecteur mouvement, p le médian de 3 vecteurs voisins représentés dans la Figure 1 (mv_a, mv_b, mv_c). Cependant si mv_b n'est pas disponible, la valeur de mv_b est égale à la valeur du vecteur mv_d . Si un ou plusieurs des vecteurs voisins ne sont pas disponibles, p est égal, en fonction des disponibilités, à mv_a ou mv_b ou mv_c ou 0. Le meilleur compromis entre la qualité et le débit est obtenu en minimisant le critère de débit-distorsion :

$$J = D + \lambda R \quad (2)$$

où D est la distorsion calculée dans le domaine spatial ou transformé et λ est une constante positive. R est le débit représentant l'ensemble des composantes du débit [8] :

$$R = R_r + \lambda_m R_m + \lambda_o R_o + \lambda_{mv} R_{mv} \quad (3)$$

où R_r est le débit du résiduel de bloc (luminance+chrominance), R_m le débit du mode de codage, R_{mv} le débit du résiduel de vecteur mouvement et R_o le débit des autres composantes (entête, structure de bloc, bit de bourrage, quantificateur). λ_m, λ_o et λ_{mv} sont des coefficients de pondération dépendant du pas de quantification. Les différentes façons d'estimer la distorsion et le débit sont décrites dans [9]. En terme de complexité de calcul, cette sélection est intensive, mais optimale au sens du compromis débit-distorsion. Cependant elle engendre un champ de vecteurs "non-naturels" (ne correspondant pas aux mouvements réels des objets). Ce processus de sélection est entrepris pour chaque partition de bloc (16x16, ..., 4x4), pour chaque image de référence, pour le pixel entier, le $\frac{1}{2}$ et le $\frac{1}{4}$ de pixels.

Le mode SKIP est un cas particulier du codage Inter. Pour un macrobloc utilisant ce mode, le codeur ne transmet aucun résiduel de bloc, aucun vecteur mouvement, ni aucun indice d'image de référence. La seule information envoyée est le mode de codage (SKIP). Le vecteur mouvement associé à ce codage correspond au vecteur prédicteur d'un Inter 16x16, hormis le cas où mv_a ou mv_b ne sont pas disponibles. Dans ce cas le vecteur est égal à 0.

La Figure 2 représente l'évolution, en fonction du pas de quantification, des débits des différentes composantes de l'Equation 3, pour le profil *High* de H.264. A bas débit, l'information de mouvement R_{mv} représente la plus importante composante du débit total et atteint 38%.

Cette observation a motivé nos recherches basées sur la diminution de l'entropie des résiduels de vecteurs mouvement. Cette diminution utilise une sélection et un codage conjoint optimal au sens débit-distorsion, avec une exploitation adaptative des redondances spatiale et temporelle, à l'aide d'une méthode par compétition.

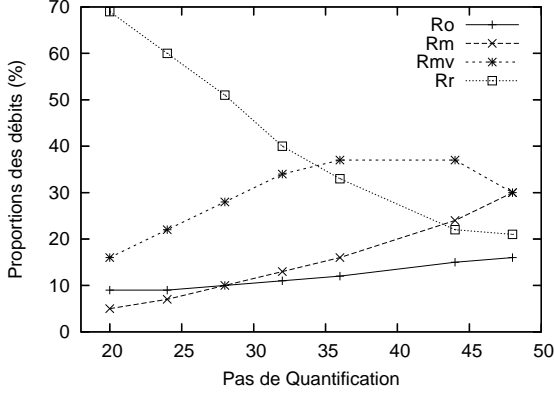


Figure 2 – Proportions des débits en fonction du pas de quantification pour la séquence Foreman.

3 Codage des vecteurs mouvement par compétition

Cette section présente nos travaux sur la sélection des prédicteurs des vecteurs mouvement, en comparaison avec la méthode du standard H.264 décrite dans la section 2.

3.1 Sélection du meilleur prédicteur

Lorsqu’un codage sans perte des vecteurs mouvement est utilisé, l’efficacité de codage dépend essentiellement des performances du prédicteur. Une méthode de codage par compétition implique plusieurs prédicteurs possibles pour un vecteur mouvement, et permet de sélectionner le meilleur. Nous avons donc défini un ensemble \mathcal{P} de prédicteurs, incluant des prédicteurs spatiaux, temporels et spatio-temporels. Les prédicteurs spatiaux sont les vecteurs voisins mv_a, mv_b, mv_c, mv_d de la Figure 1 et le médian de la norme H.264 $mv_{H.264}$. Les prédicteurs temporels sont le *collocated* mv_{col} (le vecteur mouvement utilisé pour le codage du bloc ayant la même position dans l’image précédente) et 2 médians temporels mv_{tm5} et mv_{tm9} définis par :

$$mv_{tm5} = \text{median}\{mv_{col}, \{mv_{t_j}, 0 \leq j \leq 4\}\} \quad (4)$$

$$mv_{tm9} = \text{median}\{mv_{col}, \{mv_{t_j}, 0 \leq j \leq 8\}\} \quad (5)$$

Les prédicteurs spatio-temporels que nous utilisons sont la combinaison de vecteurs spatiaux et temporels, par exemple le vecteur mv_{spt} défini ci-dessous :

$$mv_{spt} = \text{median}\{mv_{col}, mv_{col}, mv_a, mv_b, mv_c\} \quad (6)$$

Notons que ce médian donne une plus grande importance au vecteur *collocated* qu’aux vecteurs mv_a, mv_b, mv_c . Un mode de prédiction i et un résiduel ε_{mv_i} sont maintenant associés à chaque prédicteur $p_i \in \mathcal{P}$:

$$\varepsilon_{mv_i} = mv - p_i, \forall i \in \{1, \dots, n\} \quad (7)$$

où n est le nombre de prédicteurs de \mathcal{P} .

Le mode i doit être transmis dans le bitstream, tout comme le résiduel ε_{mv_i} . Le coût de cette nouvelle information (mode i) n’est pas négligeable (en moyenne 3.5% du débit et 12.5% du débit de l’information allouée au mouvement pour $n = 2$). L’efficacité de cette méthode de codage par compétition dépend toutefois du compromis entre ce coût additionnel et le gain obtenu par une plus grande précision de prédiction.

La sélection du vecteur mouvement de l’Equation 3 est remplacée par l’Equation 8 et le débit du résiduel de vecteur mouvement R_{mv} par $R_{mv/mm}$. $R_{mv/mm}$ contient le coût de la prédiction (ε_{mv_i}) et le coût du mode de prédiction (i).

$$R = R_r + \lambda_m R_m + \lambda_o R_o + \lambda_{mv/mm} R_{mv/mm} \quad (8)$$

$R_{mv/mm}$ est donné par :

$$R_{mv/mm} = \min\{\zeta(\varepsilon_{mv_i}) + \zeta(i)\}_{i \in \{1..n\}} \quad (9)$$

où $\zeta(x)$ est le coût associé à la donnée x dans le bitstream. Notons que les modes de prédiction sont codés à l’aide du codage CABAC. Un élément clé de notre méthode est la capacité du décodeur à “deviner” dans certains cas le mode de prédiction utilisé. Le codeur simule le processus de décodage pour déterminer quel mode peut être deviné et selon le cas, encode ou non le mode. Pour deviner un mode, les informations à la disposition du décodeur sont : la valeur du résiduel ε_{mv_i} , l’ensemble des prédicteurs possibles \mathcal{P} , le codage des blocs voisins (spatialement et temporellement). En pratique le mode de prédiction peut être deviné dans 27% des cas lorsque $n = 2$.

3.2 Modification du vecteur mouvement pour le mode SKIP

Comme nous l’avons expliqué dans la section 2, le mode SKIP est très efficace. Le choix de ce mode signifie qu’il est plus intéressant, au sens du critère débit-distorsion, de n’envoyer aucun résiduel de bloc ou de vecteur. Ce mode est largement utilisé, notamment dans les séquences à fond fixe. Notre objectif est par conséquent d’augmenter le plus possible l’occurrence de ce mode. La seule solution pour augmenter le nombre de modes SKIP, sans modifier le critère débit-distorsion, est de modifier son vecteur mouvement par un vecteur qui permet d’obtenir une distorsion plus faible. Le vecteur mouvement de ce mode correspond à un saut, en fonction des disponibilités, entre les valeurs : $\text{median}(mv_a, mv_b, mv_c)$, mv_a, mv_b, mv_c ou 0. Nous avons modifié ce processus en définissant un ordre privilégié de prédicteurs spatiaux et temporels. Les sauts entre les différents prédicteurs dépendent des disponibilités des données permettant de calculer ces prédicteurs. Ces disponibilités dépendent de la position du bloc et du codage (Inter/Intra) des vecteurs voisins. Le décodeur est capable de reproduire le même comportement, sans recevoir de nouvelles informations.

L’ordre des sauts que nous avons défini est donné dans le Tableau 1. Cet ordre est le suivant : 1- le médian spatial

Ordre	Prédicteur	Disponibilités
1	$median(mv_a, mv_b, mv_c)$	mv_a, mv_b, mv_c
2	mv_{tm9}	$mv_{t_j}, \forall j < 8, mv_{col}$
3	mv_{tm5}	$mv_{t_j}, \forall j < 4, mv_{col}$
4	mv_{col}	mv_{col}
5	mv_a	mv_a
6	mv_b	mv_b
7	mv_c	mv_c
8	0	

Tableau 1 – Ordre des prédicteurs pour le vecteur mouvement du mode SKIP.

est sélectionné si mv_a, mv_b, mv_c sont disponibles, 2, 3- sinon le médian temporel avec 9 puis 5 composantes est retenu si tous les vecteurs permettant le calcul sont disponibles, autrement 4- le vecteur *collocated* est choisi, puis 5- mv_a , 6- mv_b , 7- mv_c , et finalement 8- la valeur '0'. Cet ordre a été défini dans le but de privilégier des compromis entre des vecteurs. Ainsi les trois premiers prédicteurs de cet ordre sont des médians. De plus, des tests effectués sur un ensemble représentatif de séquences ont montré que le médian spatial est plus efficace que chacun des deux médians temporels.

Ce codage des vecteurs mouvement et cette modification du mode SKIP nous ont permis d'obtenir des gains de compression pour une qualité équivalente à la référence. Ces gains sont analysés et décrits dans la section suivante.

4 Résultats expérimentaux et analyse de complexité

Les expériences ont été effectuées dans le codec de référence JM10.0 H.264 [2]. Nous avons utilisé le profil *High* (Fr-Ext) avec une fenêtre de recherche de 32×32 , l'option RD-optimisation (RdOpt=1) et le codage entropique CABAC. Cette configuration correspond à la meilleure qualité possible avec les outils normatifs de H.264 et les décisions non-normatives les plus efficaces implémentées dans le récent JM, excepté les images B et les images de référence multiples. Nos modifications ne sont pas encore implémentées pour ces deux options. L'ensemble des séquences choisies est composé de 10 séquences en résolution QCIF, 10 en CIF, et 5 en SD de 100 images chacune, avec des contenus et mouvements variés. Nous avons utilisé les pas de quantification 30, 36, et 42. Comme les résultats obtenus pour les différentes résolutions sont de même ordre de grandeur, nous détaillerons uniquement les gains pour les 10 séquences CIF.

4.1 Impact sur la complexité

Le codec de référence de H.264 permet d'utiliser tous les profils et d'extraire de nombreuses statistiques du codage obtenu. Ce code C n'est pas optimisé ni en terme de com-

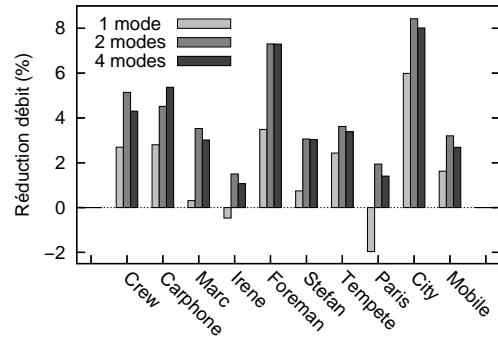


Figure 3 – Réduction du débit pour les configurations 1, 2 et 4 modes.

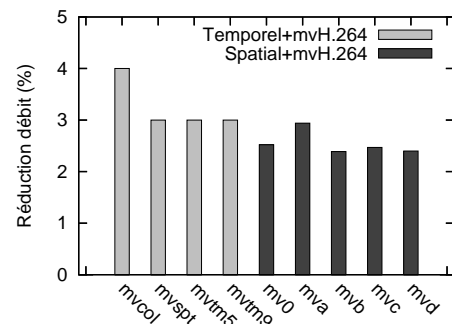


Figure 4 – Réduction du débit pour chaque combinaison de $mv_{H.264}$ et d'un autre prédicteur.

plexité de calcul ni en terme de gestion de mémoire. Par conséquent, une mesure quantitative de la complexité ne permettrait pas une évaluation correcte de cette complexité. Nous soulignerons donc uniquement l'impact de nos modifications sur l'algorithme. Cet impact dépend du nombre de prédicteurs de l'ensemble \mathcal{P} . En effet, lorsque seul le prédicteur $mv_{H.264}$ est calculé dans le codec de référence, les n prédicteurs de \mathcal{P} sont calculés. De plus pour chaque vecteur candidat de chaque fenêtre de recherche, l'ensemble des résiduels ε_{mv_i} de l'Equation 7 et le débit $R_{mv/mm}$ de l'Equation 9 sont calculés et évalués. L'impact de la modification pour le mode SKIP est négligeable car elle ne modifie pas l'algorithme de recherche. En termes de gestion de mémoire, les prédictions temporelles que nous avons définies requièrent le stockage des vecteurs mouvement et des modes utilisés pour le codage de l'image précédente.

4.2 Nombre de prédicteurs

Les résultats de notre méthode de codage par compétition dépendent du nombre et du type de prédicteurs utilisés. Nos expériences faites pour 1 ($mv_{H.264}$), 2 ($mv_{H.264} + mv_{col}$) ou 4 ($mv_{H.264} + mv_a + mv_{col} + mv_{tm9}$) prédicteurs avec la modification pour le mode SKIP, montrent que le meilleur compromis est obtenu pour 2 prédicteurs. Les résultats de cette expérience sont illustrés dans la Figure 3. Le gain moyen obtenu sur le débit total pour ces 3 configurations

est respectivement 1.8%, 4.2%, et 3.9%. Il est utile de noter que la réduction du débit est plus importante pour 2 prédicteurs, excepté pour une séquence (Carphone). Evidemment la réduction du débit des vecteurs mouvement (R_{mv}) est forte pour 4 modes de prédictions, mais le compromis avec la transmission du mode de codage ($R_{mv/mm}$) entraîne des résultats légèrement plus faibles. Une méthode adaptative, permettant de choisir des ensembles de prédicteurs différents selon des critères statistiques ou de caractéristiques locales, devrait augmenter le gain de compression.

Pour sélectionner le second prédicteur, nous avons associé le médian $mv_{H.264}$ à l'ensemble des prédicteurs décrits précédemment. Les résultats donnés en Figure 4 montrent que la combinaison de $mv_{H.264}$ avec un prédicteur temporel donne de meilleurs résultats que l'association de $mv_{H.264}$ avec un prédicteur spatial. En effet, deux prédicteurs spatiaux ont la plupart du temps des valeurs similaires. Ainsi les résiduels de vecteurs engendrés ont un coût équivalent.

4.3 Configuration avec 2 prédicteurs

La meilleure configuration sélectionnée est la combinaison d'un prédicteur spatial, le médian de la norme H.264, et d'un prédicteur temporel, le vecteur *collocated*. De plus le nouvel ordre de saut pour le vecteur du mode SKIP, défini dans la section 3.2, est employé. Tous les résultats qui suivent sont donnés pour cette configuration.

Réduction du débit de l'information de mouvement. La Figure 5(a) illustre le gain obtenu par la nouvelle information de mouvement ($R_{mv/mm}$) par rapport au coût de l'information de mouvement du standard H.264 R_{mv} . La réduction moyenne est d'environ 10%. Cette diminution est plus élevée pour les bas débits et les séquences contenant du mouvement. Elle s'explique par le fait que l'information de mouvement représente une large part du débit total, comme on peut le voir dans la Figure 2.

Augmentation du nombre de modes SKIP. La Figure 5(b) représente le pourcentage d'augmentation du nombre de modes SKIP, dont la moyenne se situe à 6%. On observe que cette augmentation n'est pas liée aux pas de quantification (QP), mais aux types des séquences. Les séquences à fond fixe (Marc, Irene, Paris) ne profitent pas de l'ordre donné dans la section 3.2. Ainsi pour ce type de séquences la valeur 0 devrait avoir une forte priorité dans l'ordre de choix. Il est aussi intéressant de noter que le gain moyen obtenu par la seule modification du mode SKIP (sans compétition sur les vecteurs) est situé à 1.8% comme le montre la Figure 3. Cette moyenne est certes faible, mais atteint 2.8% pour les séquences contenant plus de mouvement. Ce résultat est important pour une modification apportant une complexité de calcul aussi faible. De plus une adaptation de l'ordre des prédicteurs pour le mode SKIP, basée sur le type de séquence, devrait accroître les gains.

Réduction du débit total. Le Tableau 2 représente le pourcentage de sélection de chacun des deux prédicteurs.

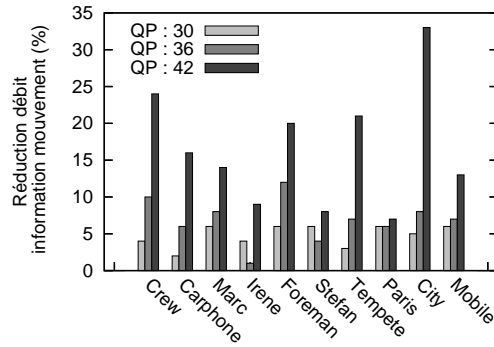
QP	Prédicteur spatial $mv_{H.264}$	Prédicteur Temporel mv_{col}	Prédicteurs égaux (même valeur)
30	62%	38%	16%
36	56%	44%	15%
42	46%	54%	17%
Moyenne	55%	45%	16%

Tableau 2 – Pourcentage de sélection des prédicteurs.

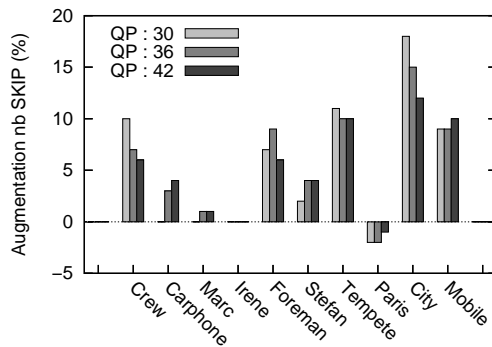
On constate que le prédicteur temporel mv_{col} est sélectionné en moyenne dans 45% des cas. Ce pourcentage exclut les cas où les deux prédicteurs sont égaux (en moyenne 16% des cas). Sachant que la sélection a été effectuée au sens du critère débit-distorsion, ce pourcentage élevé confirme l'utilité du prédicteur temporel. Une caractéristique intéressante est la corrélation entre le pourcentage de prédicteurs mv_{col} sélectionnés et les pas de quantification. Notons enfin que le pourcentage de mv_{col} varie de 24% à 62% en fonction des séquences et des pas de quantification. La Figure 5(c) montre la réduction du débit pour chaque séquence à chaque pas de quantification. On notera le gain positif pour toutes les séquences et tous les pas de quantification, même pour les séquences contenant peu de mouvement, avec une perte de PSNR moyenne de 0.04 dB, par rapport au standard H.264. La qualité visuelle reste équivalente. Le gain moyen sur le débit est d'environ 4% et atteint un maximum de 20%. Evidemment, l'augmentation est plus faible pour les séquences contenant peu de mouvement (e.g. séquences de type visiophonie) car le mode SKIP est déjà largement utilisé. Pour les séquences avec un mouvement rapide et complexe les gains sont plutôt élevés. Finalement les séquences avec un mouvement global et constant et combiné à de hautes fréquences spatiales (comme City), tirent le plus grand avantage de la prédiction temporelle, tandis que le médian spatial échoue. La réduction du débit est bien sûr liée aux pas de quantification, comme nous l'avons vu dans la Figure 2. En effet la proportion de l'information de mouvement dans le débit augmente avec le pas de quantification.

5 Conclusion

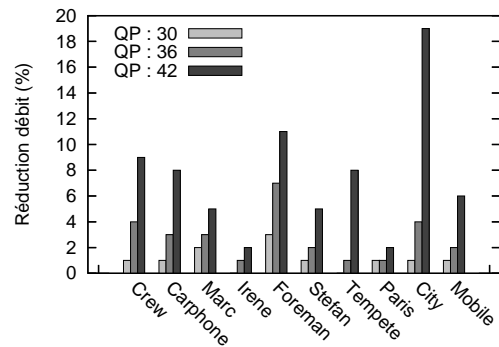
Dans cet article, une méthode de codage par compétition de prédicteurs des vecteurs mouvement est proposée. Les deux prédicteurs, spatial et temporel, sont sélectionnés via un critère débit-distorsion prenant en compte le débit des résiduels et le mode de la prédiction. De plus, une modification est proposée pour augmenter le nombre de macroblocs encodés en mode SKIP. Ces deux techniques combinées et implémentées dans le codec de référence JM10.0 H.264 fournissent un gain systématique de compression (4% en moyenne et atteint 20% avec une dégradation de PSNR négligeable) avec une très légère augmentation de complexité. Il est prévu d'implémenter ces modifications



(a) Réduction du débit de l'information de mouvement.



(b) Augmentation du nombre de modes SKIP.



(c) Réduction du débit total.

Figure 5 – Résultats expérimentaux pour 2 prédicteurs.

pour les images B et les images de référence multiples afin d'accroître encore ce gain.

Références

- [1] ITU-T Recommendation H.264 et ISO/IEC 14496-10 AVC. Advanced video coding for generic audiovisual services. version 3 : 2005.
- [2] H.264/AVC software coordination. K. Suehring, <http://iphome.hhi.de/suehring/tml/>.
- [3] L.A. Da Silva Cruz et J.W. Woods. Adaptive motion vector quantization for video coding. Dans *IEEE ICIP*, volume 2, pages 867–870, Octobre 2000.
- [4] J. Yeh, M. Vetterli, et M. Khansari. Motion compensation of motion vectors. Dans *IEEE ICIP*, volume 1, pages 574–577, Octobre 1995.
- [5] M.C. Chen et A.N. Willson. A spatial and temporal motion vector coding algorithm for low-bit-rate video coding. Dans *IEEE ICIP*, volume 2, pages 791–794, Octobre 1997.
- [6] A.M. Tourapis, F. Wu, et S. Li. Direct mode for bipredictive slices in the H.264 standard. *IEEE Trans. on CSVT*, 15(1), Janvier 2005.
- [7] S. Deuk Kim et J. Beom Ra. An efficient motion vector coding scheme based on minimum bitrate prediction. *IEEE Trans. on Image. Proc.*, 8(8) :1117–1120, Août 1999.
- [8] G.J. Sullivan et T. Wiegand. Rate-distortion optimization for video compression. *IEEE Signal Proc. Mag.*, pages 74–90, 1998.
- [9] K.P. Lim, G. Sullivan, et T. Wiegand. *Text Description of JM Reference Encoding Methods and Decoding Concealment Methods*. JVT-N046 contribution, Hong-Kong, Janvier 2005.