

Adaptation orientée contenu pour le codage par compétition de prédicteurs de mouvement

Guillaume Laroche^{1,2} Joël Jung¹ Béatrice Pesquet²

¹ Orange-France Telecom R&D, 38-40 rue du G. Leclerc, 92794 Issy Les Moulineaux

² ENST Paris, 46 rue Barrault, 75014 Paris, France

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

{beatrice.pesquet}@enst.fr

Résumé

Le standard de compression vidéo H.264/MPEG4-AVC permet une réduction significative du débit par rapport à ces prédécesseurs. Ces gains proviennent en partie de l'amélioration de la compensation de mouvement avec une résolution sub-pixellique plus fine et l'utilisation de tailles variables de bloc. Ces outils ont contribué à l'augmentation de la proportion de l'information de mouvement dans le débit total. Dans nos travaux précédents nous avons proposé une méthode de compétition de prédictions de vecteurs mouvement afin d'améliorer le codage de cette information. Cette méthode a été adoptée par le Video Coding Experts Group et intégrée dans le JM KTA. Dans cet article nous proposons une adaptation d'ensembles de prédicteurs pour le mode Skip orientée contenu afin d'une part de réduire l'information liée à la compétition exhaustive des prédicteurs et d'autre part d'offrir des prédicteurs adaptés au contenu.

Mots clefs

H.264, mode Skip, JM KTA, méthode de codage par compétition, critère débit distorsion, vecteur mouvement.

1 Introduction

Le standard de compression vidéo ITU-T SG16-Q6 H.264 [1] du Video Coding Experts Group (VCEG) appelé aussi ISO/IEC JTC1/SC29/WG11 MPEG-4 AVC par le Moving Picture Experts Group (MPEG) commence à être largement adopté par les industriels grâce aux gains significatifs obtenus par rapport à ses prédécesseurs H.263 et MPEG-4 partie 2. Ces gains proviennent de la multiplication des possibilités de codage pour un macrobloc (MB) avec les multiples prédicteurs Intra, les partitions variables des MBs et le 1/4 de pixel pour la compensation de mouvement. Parallèlement, un important travail a été accompli sur la sélection du meilleur codage parmi l'ensemble des possibilités offertes par H.264. Cette

sélection optimale, non normative, implémentée dans le codec de référence [2], est basée sur un critère débit distorsion (RD) [3].

Actuellement le Joint Video Team (JVT) regroupant les deux groupes VCEG et MPEG, est focalisé sur la fin de la normalisation de l'annexe scalable (H.264-SVC) et sur le codage multivues (H.264-MVC). Cependant les activités classiques sur l'amélioration du codage vidéo n'ont pas cessé. Afin d'encourager de nouvelles contributions, il a été décidé au 26ième meeting VCEG la création d'un logiciel, nommé KTA (Key Technical Area) [4], regroupant l'ensemble des outils proposés depuis la fin de la normalisation d'H.264 qui apporte des gains jugés significatifs par le groupe VCEG.

Le JM KTA est basé sur le codec de référence de la norme H.264 JM11.0 [2] et contient 5 nouveaux outils : le 1/8 de pixel pour l'estimation et la compensation de mouvement [5], l'Adaptive Interpolation Filter (AIF) [6] pour améliorer le codage sub-pixellique, l'Adaptive Prediction Error Coding [7] permettant la sélection entre les transformées DCT standards et un mode de codage sans transformée, l'Adaptive Quantization Matrix Selection [8] pour sélectionner la meilleure matrice de quantification pour les résiduels de texture transformée et enfin la méthode de compétition de prédiction de vecteurs mouvement (MVComp) que nous avons proposée dans [9] et [10]. Cette méthode permet de mettre en compétition un ensemble de prédicteurs à la fois pour les prédictions de vecteurs mouvement et pour le mode Skip. De plus il est possible de définir, slice par slice, l'ensemble des prédicteurs à utiliser pour chacune des deux compétitions. Ainsi nous avons proposé dans [11] une première adaptation afin d'améliorer l'efficacité de codage obtenue avec MVComp.

Dans cet article nous proposons une adaptation automatique de l'ensemble des prédicteurs à utiliser pour le mode Skip basée sur le contenu. Chaque MB est classé soit dans une zone contenant du mouvement soit dans une zone ne contenant pas ou peu de mouvement. Pour chaque zone on affecte un ensemble de prédicteurs qui évolue

slice par slice en fonction de statistiques. L'idée est d'affecter un nombre limité et efficace de prédicteurs dans les zones contenant peu ou pas de mouvement où le mode Skip est souvent sélectionné dans la référence et d'attribuer un nombre plus élevé de prédicteurs pour les zones contenant du mouvement afin d'augmenter la sélection du mode Skip.

Cet article est organisé de la manière suivante : une description du mode de codage Skip de la norme H.264 ainsi que la méthode de codage par compétition de prédicteurs pour ce mode sont décrits dans la première section. La section 2 présente la méthode proposée avec la classification des MBs et la mise à jour de chacun des deux ensembles de prédicteurs utilisés. Enfin la section 3 présente les résultats expérimentaux situés en moyenne à 8,9% de réduction de débit par rapport à la norme H.264.

2 Etat de l'art

2.1 Sélection et codage du mode Skip dans la norme H.264

Le mode Skip est un mode particulier du codage Inter (codage exploitant les redondances temporelles entre les images). Il réalise une prédiction temporelle pour laquelle aucune information n'est transmise au décodeur : aucun résiduel de bloc ni de vecteur mouvement ni d'indice de référence frame, excepté le mode de codage utilisé pour le MB (mode Skip). Le décodage d'un MB codé en mode Skip s'effectue en recopiant le bloc prédicteur provenant de la première image de référence compensée en mouvement avec le vecteur mouvement prédicteur [1].

Le vecteur mouvement prédicteur pour le mode Skip est le médian spatial des vecteurs mouvement mv_a , mv_b et mv_c , utilisés pour le codage des trois blocs voisins. Si l'un de ces vecteurs mouvement n'existe pas (bordure de l'image, codage Intra) ou si les vecteurs ne pointent pas tous sur la même image de référence, alors le vecteur mouvement prédicteur pour le mode Skip est égal à mv_a ou mv_b ou mv_c ou mv_d ou 0. De plus si mv_a ou mv_b est égal à 0, la valeur du prédicteur est égale à 0.

Dans le codec de référence de la norme H.264, la sélection du meilleur mode de codage parmi l'ensemble des possibilités de codage se fait à l'aide de la minimisation du critère de débit distorsion :

$$J = D + \lambda R \quad (1)$$

où D est la distorsion, λ le multiplicateur de Lagrange dépendant du pas de quantification (QP) utilisé et R le débit de l'ensemble des composantes. Les débits sont calculés pour chaque mode au bit près et en particulier pour le débit du mode Skip :

$$J_{SKIP} = D_{SKIP} + \lambda_m R_m \quad (2)$$

où D_{SKIP} est la distorsion introduite par le mode Skip, λ_m est le multiplicateur de Lagrange pour les modes de codage et R_m le coût en nombre de bits du mode de codage Skip. Ce coût est généralement inférieur à 1 bit par MB et donc négligeable comparé aux autres modes de codage.

2.2 Codage par compétition pour le mode Skip

Dans ce paragraphe nous décrivons la méthode [10] dont le groupe VCEG a jugé les résultats significatifs pour être intégrée dans le JM KTA. La méthode de codage par compétition de prédicteurs de vecteur mouvement pour le mode Skip consiste à utiliser un ensemble de N prédicteurs à la place du seul prédicteur médian de la norme H.264. La sélection du meilleur prédicteur se fait de manière exhaustive à l'aide du critère débit distorsion. Ainsi l'Eq. 2 devient:

$$J_{Skip} = \min \{ J_{skip}^i \}_{i=1..N} + \lambda_m R_m \quad (3)$$

avec

$$J_{skip}^i = D_i + \lambda_m \zeta(i) \quad (4)$$

où D_i est la distorsion du prédicteur numéro i et $\zeta(i)$ est le coût en nombre de bits de l'indice i à transmettre. Si tous les prédicteurs sont égaux, l'indice i n'est pas transmis car le décodeur est capable de tester l'égalité de ces prédicteurs. De plus si 2 ou plusieurs prédicteurs sont égaux entre eux, l'indice de chaque prédicteur est réordonné afin que seuls les prédicteurs ayant des valeurs distinctes soient considérés dans l'ensemble des prédicteurs à utiliser.

2.3 Analyse de la sélection du mode Skip

Le mode Skip est un mode de codage très performant, son coût en nombre de bits pour un MB est inférieur à 1 bit. La sélection de ce mode signifie qu'il est plus intéressant au sens du critère débit distorsion de ne transmettre aucun résiduel de bloc ou de vecteur mouvement. Ce mode de codage a été créé pour des zones ayant une forte probabilité d'obtenir un résiduel de vecteur et un résiduel de texture égal à zéro. Ce mode est donc souvent sélectionné dans les zones sans mouvement ou avec des mouvements constants. Cette hypothèse nous a été confirmée par notre étude de la répartition spatiale des modes de codages de plusieurs séquences. Le mode Skip est donc plus souvent sélectionné pour les zones statiques ou avec peu de mouvement que pour les zones contenant beaucoup de mouvement.

3 Adaptation orientée contenu d'ensembles de prédicteurs pour le mode Skip

Dans [11] nous avons montré qu'une adaptation de l'ensemble de prédicteurs utilisés pour la compétition de prédicteurs pour le mode Skip basée sur des statistiques de codage permet d'accroître les gains. Dans cet article nous proposons une amélioration de cette méthode avec l'utilisation de 2 ensembles de prédicteurs, l'un pour les MBs des zones ayant peu ou pas de mouvement et l'autre pour les MBs des zones ayant du mouvement. De plus les deux ensembles évoluent indépendamment. Cette section détaille la méthode proposée.

3.1 Adaptation d'un ensemble de prédicteurs

La mise à jour de chaque ensemble se fait slice par slice au codeur et l'ensemble de prédicteurs utilisé pour chaque classe est transmis au décodeur. L'ensemble courant pour une classe correspond donc aux statistiques du slice précédent. Pour chaque MB tous les J_{skip}^i de tous les prédicteurs, appartenant ou non à l'ensemble utilisé, sont calculés. Notons $RdCount^i$ le nombre de fois où le prédicteur p_i remplit la condition :

$$J_{skip}^i \leq J_{min} \quad (5)$$

où J_{min} est le coût minimum du critère débit distorsion obtenu pour tous les modes de codages Inter et Intra. Ce critère est efficace car il représente le nombre de fois où le prédicteur p_i est meilleur ou équivalent au codage sélectionné pour le MB. Pour sélectionner l'ensemble des prédicteurs à utiliser pour une classe définissons :

$$MaxRdCount = \max\{RdCount^i\}_{\forall i < N} \quad (6)$$

où N est le nombre de prédicteurs total. Au début de la mise à jour aucun prédicteur n'appartient à l'ensemble. Un prédicteur p_i est ajouté dans l'ensemble des prédicteurs si :

$$(MaxRdCount) \leq \delta \times RdCount^i \quad (7)$$

Ceci signifie que le prédicteur p_i est ajouté si le nombre de fois où il minimise le meilleur des codages est proche, en fonction de la constante δ , du maximum de fois où un prédicteur a minimisé l'ensemble des codages. La constante $\delta \in [1; +\infty[$; si elle est égale à 1, seul le prédicteur qui a maximisé le nombre de minimisations de codage est sélectionné. Si δ est égal à $+\infty$, l'ensemble des prédicteurs va être utilisé.

3.2 Liste des prédicteurs

Nous avons défini 11 prédicteurs pour l'adaptation :

- $mv_{H.264}$, le médian de la norme H.264 [1] ;
- mv_a, mv_b, mv_c les vecteurs utilisés pour le codage des blocs voisins ;
- mv_{extspa} , vecteur légèrement différent de $mv_{H.264}$ qui retourne le médian de mv_a, mv_b, mv_c si les trois vecteurs sont disponibles sinon mv_a si disponible sinon mv_b sinon mv_c sinon 0 ;
- $mv_{col}, collocated$, le vecteur utilisé pour le codage du bloc ayant la même position que le bloc courant dans l'image précédente.
- $mv_{sa}, mv_{sb}, mv_{sc}$, les vecteurs voisins mis à l'échelle en fonction des distances temporelles parcourues par le vecteur courant et les vecteurs voisins.
- mv_{jf} le vecteur mouvement du bloc dans l'image précédente à la position donnée par $mv_{H.264}$.

Le nombre de prédicteurs peut sembler élevé mais beaucoup d'entre eux donnent la même valeur dans la plupart des cas. Par exemple $mv_{H.264}$ et mv_{extspa} ont la même valeur si tous les vecteurs sont disponibles et qu'ils ont la même image de référence. De même, pour les vecteurs mv_a, mv_b, mv_c et les vecteurs $mv_{sa}, mv_{sb}, mv_{sc}$ qui ont respectivement la même valeur si les vecteurs originaux pointent sur la première image de référence. mv_{col} et mv_{jf} ont aussi la même valeur si $mv_{H.264}$ est égal à 0. Par conséquent l'utilisation conjointe de $mv_{H.264}$ et mv_{extspa} ou de mv_{col} et mv_{jf} etc. n'implique pas une augmentation élevée du coût de l'indice du prédicteur. De plus lorsqu'une rupture du mouvement (rupture spatiale ou temporelle du mouvement) apparaît, ces couples de prédicteurs ont des valeurs différentes. C'est justement dans ce cas que la multiplication des possibilités de codages est intéressante.

3.3 Classification des macroblocs

Chaque MB est classé soit dans la classe **S** représentant une zone statique avec peu ou pas de mouvement, soit dans une classe **M** représentant une zone avec du mouvement. La classe **S** correspond aux zones où le mode Skip a une forte probabilité d'être sélectionné et la classe **M** représente les zones où le Skip a une plus faible probabilité d'être sélectionné.

La classification des MBs doit être calculable au décodeur pour ne pas introduire une nouvelle information à transmettre. Nous avons donc choisi d'utiliser les deux images de référence précédentes qui sont disponibles au décodeur pour calculer ce critère. Ce critère est la somme des valeurs absolues (*Sum of Absolute Differences*, ou SAD) du MB *collocated* de la première image de

référence et du MB *collocated* de la seconde image de référence. Si ce SAD est inférieur à un seuil, le MB courant est classé dans la classe *S*, sinon dans la classe *M*. A noter que les adaptations de l'ensemble des prédicteurs des deux classes sont indépendantes. Ainsi, chacune des classes a ses propres *RdCount*¹, son propre *MaxRdCount* et son propre δ , afin que les prédicteurs et le nombre de prédicteurs sélectionnés dans chaque ensemble *S* et *M* correspondent aux caractéristiques de la zone. En effet, afin que la classe *S* contienne peu de prédicteurs, le δ pour cette classe est proche de 1 et le δ de la classe *M* lui est supérieur afin qu'il contienne plus de prédicteurs.

4 Résultats expérimentaux

Les expériences ont été effectuées avec le profil Baseline dans le JM KTA 1.1 [4]. Nous avons utilisé les conditions de tests [12] recommandées par le groupe VCEG excepté les pas de quantification. En effet nous ciblons les bas débits où le nombre de MB encodés avec le mode Skip est plus élevé. Les conditions de test sont les suivantes : codage entropique CAVLC, une seule image codée en Intra (la première), 4 images de référence, une fenêtre de recherche pour l'estimation de mouvement 32x32 et 4 QPs 27 32 37 42. L'ensemble des gains en réduction de débit présenté dans ce papier sont calculés avec la métrique de Bjontegaard [13], qui permet de calculer une distance moyenne entre des courbes de débit distorsion.

4.1 Augmentation de l'occurrence du mode Skip

La Figure 1 représente le pourcentage d'augmentation du nombre de MB encodés avec le mode Skip pour la méthode de compétition de vecteurs mouvement MVComp et la méthode proposée dans la section 3.

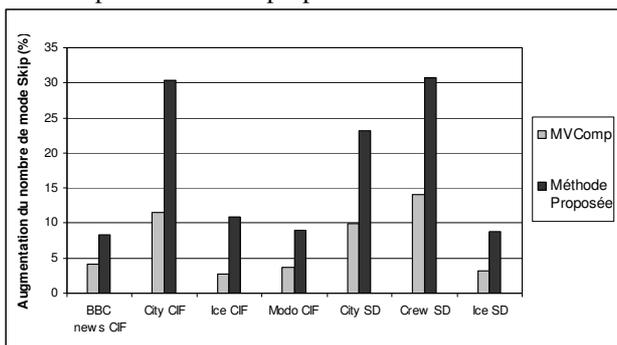


Figure 1 – Augmentation du nombre de macroblocs encodés avec le mode Skip pour MVComp et la méthode proposée

Les résultats de la méthode MVComp sont donnés avec l'utilisation de deux prédicteurs pour la compétition de prédicteurs pour le mode Skip mv_{extspa} et mv_a qui est la meilleure configuration obtenue en moyenne. A noter que pour les deux méthodes, la compétition sur les vecteurs

mouvement est également activée. Le pourcentage d'augmentation du nombre de MBs skippés est corrélé avec le type des séquences. Le pourcentage d'augmentation est plus faible pour les séquences ayant un fond fixe car le nombre de MBs skippés est élevé. Pour toutes les séquences la méthode proposée obtient un pourcentage plus de 2 fois supérieur au pourcentage obtenu avec MVComp.

4.2 Analyse des deux ensembles de prédicteurs

La Figure 2 représente le nombre moyen de prédicteurs pour chacun des deux ensembles de prédicteurs, pour chaque séquence, pour tous les QPs. En moyenne l'ensemble *S* contient 2,8 prédicteurs et l'ensemble *M* contient 8,8 prédicteurs. Ces nombres sont évidemment liés aux constantes δ fixés à 1,05 pour l'ensemble *S* et 1,6 pour l'ensemble *M*.

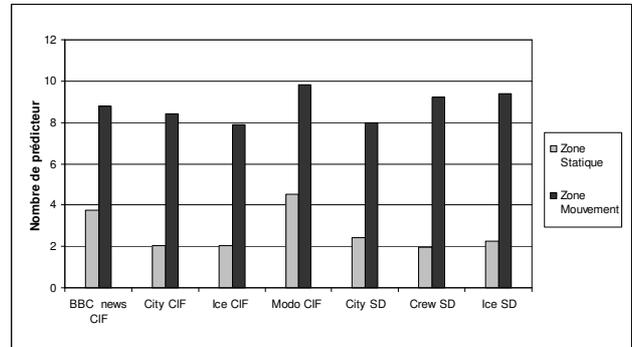


Figure 2 – Nombre moyen de prédicteurs utilisé pour chacun des ensembles de prédicteurs

La Figure 3 représente le nombre moyen de bits nécessaire à la transmission de l'index du prédicteur pour le mode Skip pour la méthode MVComp décrite dans [10] et pour chacune des zones de la méthode proposée. Dans cette figure le coût moyen est donné séquence par séquence pour tous les QPs.

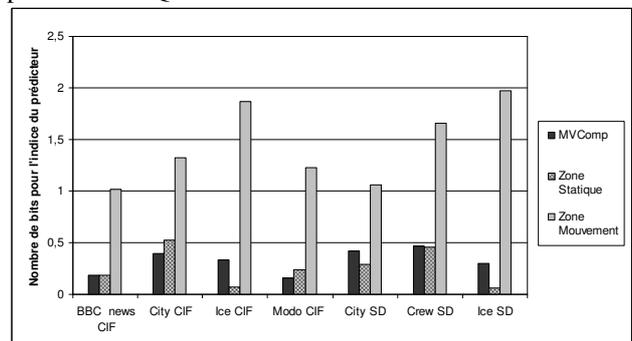


Figure 3 - Nombre de bits moyen par macrobloc nécessaire au codage de l'index du prédicteur

Nous pouvons remarquer que le nombre de bits nécessaire pour le codage des indices des prédicteurs de la zone sans mouvement est proche de celui de la méthode MVComp

avec deux prédicteurs contrairement au codage des indices des prédicteurs de l'ensemble M qui est largement supérieur avec l'utilisation de plus de 8 prédicteurs en moyenne.

4.3 Gain sur le débit total

La Figure 4 représente le gain sur le débit total, par rapport à la référence H.264, pour la méthode proposée et MVComp pour chaque séquence en moyenne pour tous les QPs. Notons que, de même que pour tous les autres résultats présentés précédemment, la compétition sur les prédictions des vecteurs mouvement est aussi activée.

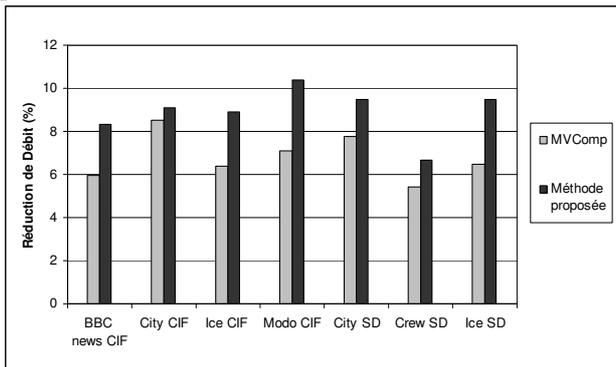


Figure 4 – Réduction du débit pour la méthode MVComp et la méthode proposée en moyenne sur les 4 QPs

La réduction de débit pour la méthode MVComp est en moyenne de 6,8% et de 8,9% pour la méthode proposée. Le gain est systématique pour l'ensemble des séquences que l'on a utilisées. Le plus mauvais résultat que nous avons obtenu est pour la séquence *Soccer SD* avec une augmentation de débit en moyenne de 0,9% par rapport à MVComp mais reste inférieur au débit de la référence H.264. Notons que le gain de la méthode proposée par rapport à MVComp est en moyenne de 2.25%. Les gains semblent être liés aux types des séquences. Les séquences à fond fixe comme *Modo* et *Ice* obtiennent de plus fortes augmentations. Le coût élevé de l'indice des prédicteurs pour l'ensemble M a été compensé par l'augmentation du nombre de MBs sautés dans les zones avec du mouvement.

5 Conclusion

Dans cet article, une méthode d'adaptation d'ensembles de prédicteurs orientée contenu pour la compétition de prédicteurs pour le mode Skip est proposée. Chaque macrobloc est classé soit dans une classe dédiée aux zones statiques, soit dans une classe dédiée aux zones en mouvement. Chacune des deux classes a son propre ensemble de prédicteurs, déterminé indépendamment à l'aide de statistiques basées sur le critère débit distorsion. Les ensembles de prédicteurs sont transmis à chaque slice. Cette méthode donne des gains systématiques sur l'ensemble des séquences présentées par rapport à nos

travaux précédents déjà adoptés par le groupe VCEG et intégrés dans le JM KTA. Les gains moyens comparés à la norme H.264 sont de 8.9%.

Références

- [1] ITU-T. Recommendation H.264 and ISO/IEC 14496-10 AVC, "Advanced video coding for generic audiovisual services" version 3: 2005.
- [2] H.264 software coordination, K. Suehring, <http://iphome.hhi.de/suehring/tml/>.
- [3] K. Lim, G. Sullivan, and T. Wiegand, Text Description of JM Reference Encoding Methods and Decoding Concealment Methods, JVT-N046 contribution, Hong-Kong, January 2005.
- [4] KTA software coordination, Y. Vatis, <http://www.tnt.uni-hannover.de/~vatis/kta/>.
- [5] T. Wedi, "1/8-pel motion vector resolution for H.26L," ITU-T VCEG, Portland, USA, Proposal Q15-K-21, August 2000
- [6] Y. Vatis, B. Edler, D. Thanh Nguyen, J. Ostermann, "Two-dimensional non-separable Adaptive Wiener Interpolation Filter for H.264/AVC", ITU-T SGI 6/Q.6 Doc. VCEG-Z17, Busan, April 2005.
- [7] M. Narroschke, H. G. Musmann, "Adaptive prediction error coding in spatial and frequency domain with a fixed scan in the spatial domain", ITU-T SG16/Q.6 Doc. VCEG-AD07, Hangzhou, October 2006.
- [8] A. Tanizawa, T. Chujoh, "Adaptive Quantization Matrix Selection on KTA Software", ITU-T SG16/Q.6 Doc. VCEG-AD06, Hangzhou, October 2006.
- [9] G. Laroche, J. Jung, and B. Pesquet-Popescu, "A spatio-temporal competing scheme for the rate-distortion optimized selection and coding of motion vectors," in Proc. European Signal Processing Conf., Florence, Italy, September 2006.
- [10] J. Jung, G. Laroche, "Competition-Based Scheme for Motion Vector Selection and Coding", VCEG Contribution VCEG-AC06, Klagenfurt, July 2006.
- [11] J. Jung, G. Laroche, B. Pesquet-Popescu, "RD optimized competition scheme for efficient motion prediction", Invited Paper, VCIP 2007, SPIE Electronic Imaging, January 28-February 1st, 2007, San Jose, CA, USA
- [12] T. K. Tan, G. J. Sullivan, and T. Wedi, "Recommended simulation common conditions for coding efficiency experiments," ITU-T VCEG, Nice, Input / Discussion VCEG-AA10, October 2005.
- [13] G. Bjontegaard, "Calculation of average PSNR differences between RD-curves," ITU-T VCEG, Texas, USA, Proposal VCEG-M33, April 2001.