H.264/MPEG-4 AVC, un nouveau standard de compression vidéo

F. Loras[1] J. Fournier[2]

1 France Telecom R&D / DMI / SGV (Laboratoire de Service de visiophonie de Groupe et de Visioconférence)

2 France Telecom R&D / DIH / HDM (Laboratoire Hyperlangages et Dialogues Multimédia)

France Telecom R&D,
39 avenue du Général Leclerc 92794 Issy Les Moulineaux Cedex 9 France

{frederic.loras, jerome.fournier }@francetelecom.com

Résumé

L'objectif de ce document est de présenter une nouvelle norme de compression vidéo. Une première partie présente l'historique de la compression vidéo à travers les organismes de normalisation. Les principes de base de la compression vidéo sont présentés. Une seconde partie présente la prochaine norme de codage vidéo H.264 / MPEG-4 « AVC », qui promet des résultats significativement meilleurs que les normes de codage actuelles.

Mots clefs

Compression vidéo, H.264, MPEG-4 « AVC », JVT, H.26L, UIT-T, ISO/IEC.

1 Introduction

La normalisation joue un rôle essentiel dans le succès des technologies numériques. En effet, le choix de ces technologies, en termes d'efficacité, d'interopérabilité et de pérennité, permet d'assurer un large déploiement des services et produits associés à un coût optimal pour les constructeurs et les utilisateurs.

Au niveau international, les deux organismes les plus actifs pour la normalisation des systèmes de compression vidéo sont l'UIT-T et l'ISO/IEC. Les travaux techniques de l'ISO/IEC sont menés au sein du groupe MPEG (Motion Picture Experts Group) qui a défini les standards MPEG-1, MPEG-2 et MPEG-4 pour des applications aussi variées que la télévision ou le multimédia. En parallèle des activités de MPEG, le groupe vidéo de l'UIT-T s'intéresse principalement à la définition de recommandations techniques destinées aux applications de visiophonie et de visioconférence (normes H.261 et H.263).

Actuellement, ces deux organismes travaillent à l'élaboration d'une norme commune (H.264 / MPEG-4 « AVC ») dont les performances attendues devraient permettre de réduire de moitié le débit de transmission ou de stockage pour une qualité visuelle équivalente aux normes précédentes.

2 Historique de la compression vidéo

Les codeurs vidéo actuellement normalisés (familles H.26x ou MPEG-x) utilisent tous le même principe de base dont la connaissance aide à comprendre à la fois leurs potentialités et limites d'utilisation. Cette partie décrit les principaux standards normalisés à l'UIT-T et à l'ISO/IEC.

2.1 Les normes de l'UIT-T

Première norme vidéo approuvée en 1993, **H.261** [1] vise les applications de visiophonie pour le réseau RNIS à des débits multiples de 64 kbit/s. Les formats d'image traités sont le QCIF (144x176 pixels) et le CIF (288x352 pixels). La fréquence image de base est 29.97 Hz mais peut être réduite (sous-multiples). Le schéma de codage (Figure 1) est constitué de plusieurs modules : l'estimation et la compensation de mouvement, le calcul des résidus, la transformée espace-fréquence (DCT[1]), la quantification et le codage entropique. Le codeur intègre également un décodeur qui effectue les opérations inverses. Pour la compensation de mouvement chaque MB peut être affecté d'un vecteur de mouvement dont les dimensions horizontales et verticales ne peuvent excéder +/- 15 pixels. Le vecteur de mouvement estimé sur la luminance pointe un MB inclus dans l'image précédemment reconstruite. Les chrominances sont prédites par le même vecteur mouvement dont les

[1] DCT : Discrete Cosine Transform

coordonnées ont été divisées par 2 et arrondies à l'entier le plus proche. Cette prédiction correspond à la construction d'une image P. Le filtrage, optionnel, est un filtre spatial à deux dimensions non récursif travaillant au niveau pixel sur des blocs de taille 8x8 intervenant dans la boucle de codage. Le flux vidéo est structuré en quatre niveaux : image, groupe de blocs (GOB), macrobloc[2] (MB) et enfin bloc[3].

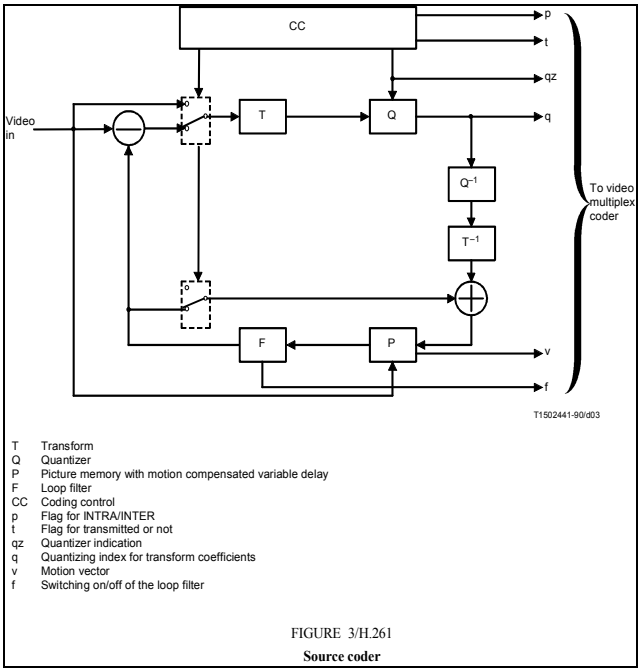


Figure 1 – *Schéma de codage H.261*

H.263 [2] est une norme de codage vidéo pour la communication vidéo à très bas débit dont la première version fut adoptée en 1995. Elle vise les applications de visiophonie et de visioconférence sur RTC et RNIS. Cette norme repose sur les principes mis en place par la recommandation H.261. Les formats d'images sont des multiples et sous-multiples du CIF (352x288 pixels). Le codeur H.263 a la possibilité d'effectuer des compensations en mouvement avec une précision au demi-pixel, ce qui améliore grandement la qualité de la vidéo en réduisant fortement les effets dits "moustiques". L'utilisation d'image B est désormais possible.

La version 2 de la recommandation H.263 (1998), souvent appelée **H.263+** [3], met en oeuvre douze options supplémentaires et permet désormais de définir des formats et fréquences d'image personnalisés. Les caractéristiques de vidéo (Taille, fréquence) sont transmises dans le flux vidéo. Les options ajoutées améliorent fortement la qualité et la robustesse aux erreurs.

La dernière version de H.263 (2000), appelée **H.263++** [4], ajoute trois options et une spécification à la version antérieure. Outre l'amélioration en termes de qualité et de taux de compression, elle prend mieux en compte la transmission vidéo temps réel sur des réseaux à qualité de service non garantie (IP et mobiles).

2.2 Les normes de l'ISO/IEC

Première norme de compression vidéo développée par l'ISO/IEC, **MPEG-1** [5] vise une qualité équivalente au VHS (format SIF ou CIF) à un débit de 1,5 Mbps. Cette norme a été construite sur la base de H.261 dont elle reprend les principes en les améliorant : compensation de mouvement au ½ pixel, images de types I (Intra), P (Prédite) ou B (Bidirectionnelle), quantification optimisée par l'utilisation de matrices de quantification, prédiction spatiale du coefficient DC pour les images I. MPEG-1 n'est plus guère utilisée aujourd'hui si ce n'est en compression du son avec le format MP3 pour le stockage de la musique.

La norme **MPEG-2** [6] a été définie pour les applications liées à la TV numérique, à la fois au niveau professionnel (production audiovisuelle, etc.) et au niveau du grand public (diffusion vers les postes TV). Elle reprend les principes de MPEG-1 en ajoutant les outils indispensables pour les applications télévisuelles : traitement des formats entrelacés, optimisation des outils MPEG-1 (dynamique des vecteurs mouvement, etc.), scalabilité visant la compatibilité TV/TVHD. Ce standard a été adopté par le consortium DVB (Digital Video Broadcasting) pour les services de TV numérique par voie hertzienne terrestre (DVB-T) et satellite (DVB-S). Il est également utilisé comme format de codage du DVD (Digital Video Disc).

MPEG-4 [7] est une norme générique de compression destinée à la manipulation d'objets multimédia. Elle permet le codage d'une grande variété de format vidéo (taille, résolution, fréquence image) mais aussi le codage d'objets vidéo de forme arbitraire, d'images fixes (codage par ondelettes) ainsi que d'objets synthétiques 3D. De ce fait, cette norme adresse une large gamme d'applications audiovisuelles allant de la visioconférence à la production audiovisuelle en passant par le streaming sur Internet ou encore la réalité virtuelle distribuée. Concernant le codage de la vidéo traditionnelle, MPEG-4 combine les outils de MPEG-2 et H.263 ainsi que de nouveaux outils lui conférant une plus grande efficacité en compression tels que des modes de scalabilité plus adaptés à une transmission sur réseaux à débit fluctuant ainsi qu'une plus grande robustesse aux erreurs de transmission.

[2] Un macrobloc est un bloc carré de 16x16 pixels

[3] Un bloc est un carré de 8x8 pixels

3 La norme H.264 / MPEG-4 «AVC»

3.1 Historique

H.26L a été le nom provisoire de la future norme de codage en cours d'élaboration au sein de l'UIT-T. H.26L est fondé sur une architecture identique aux précédentes normes de codage (compensation en mouvement, calcul de l'erreur de prédiction, transformée fréquentielle, quantification, codage entropique). Les premiers tests de la norme ont montré de très bonnes performances pour les applications classiques de vidéo comme pour des applications de haute qualité (Cinéma numérique).

Les groupes de travail à l'UIT-T et à MPEG ont approuvé le rapprochement de leurs équipes vidéo pour la définition commune d'un nouveau standard de compression. Cette décision a conduit les groupes à fusionner sous le nom de JVT (Joint Video Team) le 6 décembre 2001. Le but de cette nouvelle entité est de standardiser un codec vidéo dont la base est H.26L. Les travaux, commencés en 1998 à l'UIT-T, devraient aboutir à un standard international en mars 2003 (Figure 2).

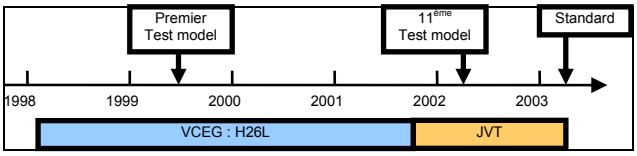


Figure 2 – *Historique de H.264 / MPEG-4 « AVC »*

3.2 Spécificité de la norme

H.26L n'apporte pas de rupture technologique par rapport aux normes de codage vidéo actuelles. Les différences se situent à plus petite échelle sur chaque partie du principe général de codage (prédiction, transformée, quantification, etc.).

3.2.1.1 Les types d'image

On retrouve les mêmes types d'images que dans les normes précédentes (I, P ou B). En revanche, H.26L apporte un nouveau type d'image, les « SP[4] frame », qui servent à coder la transition entre deux flux vidéo. Elles permettent, sans envoyer d'images intra coûteuses en débit, de passer d'une vidéo à une autre. Ces Images peuvent être utilisées dans des contextes variés comme la transmission sur réseaux à QoS[5] non garantie.

3.2.1.2 Les macroblocs « intra »

Les coefficients des macroblocs intra sont prédits en privilégiant certaines directions spatiales. Huit directions de prédiction ainsi que deux méthodes de codage existent aujourd'hui. Ces améliorations permettent d'obtenir des taux de compression comparables à JPEG 2000 pour la compression d'images fixes.

3.2.1.3 La compensation de mouvement

Le processus de compensation de mouvement diffère des normes précédentes en proposant une grande variété de formes et de tailles de blocs (16x16, 16x8… 8x4…4x4) avec une précision pouvant aller jusqu'au ¼ de pixel. H.26L utilise alors entre 8 et 36 types de blocs pour la compensation de mouvement alors que les précédents standards n'en utilisaient que de un à trois. Cette caractéristique permet de s'adapter plus finement au contenu spatial et en mouvement des images.

Une autre nouveauté importante est que l'image à coder peut être prédite à partir de plusieurs images de référence. Cette information est définie au niveau des macroblocs. H.26L mutualise ainsi les fonctionnalités de MPEG-4 (prédiction au ¼ de pixel) et de H.263+ (utilisation possible de plusieurs images de référence, mais seulement au niveau Slice).

3.2.1.4 La transformée fréquentielle

Elle a les mêmes propriétés que la DCT. Contrairement aux autres standards qui utilisent des blocs 8x8, elle opère sur des blocs 4x4. La transformée inverse de H.26L est définie de manière exacte (précision entière) pour éviter les divergences dues aux arrondis. Dans certains cas, pour diminuer les coefficients isolés, la transformée est faite en deux passes avec une extension de la taille des blocs à 2x2 pixel.

3.2.1.5 La quantification

Les pas de quantification définis dans la norme s'incrémentent d'une valeur de 12,5% et leur dynamique est augmentée puisque les valeurs vont de 1 à 52. Dans les normes vidéo précédentes, le pas de quantification augmente par pas fixes, ce qui entraîne des zones inaccessibles pour certains quantificateurs. De plus, afin d'obtenir de meilleurs résultats visuels, la quantification de la chrominance est plus fine que celle de la luminance.

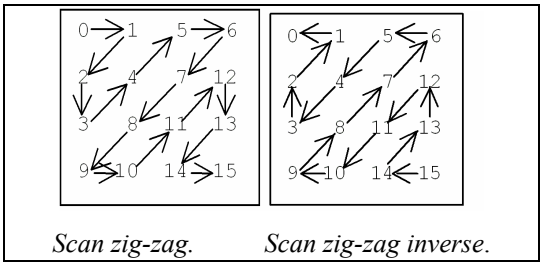
3.2.1.6 Le filtrage de boucle

H.26L intègre également un filtre de boucle qui améliore l'efficacité en compression et la qualité visuelle des séquences vidéo en gommant certains effets indésirables du codage tels que les effets de bloc.

3.2.1.7 Le parcours des données Il existe dans H.26L deux modes de parcours des coefficients transformés (zigzag) pour concentrer l'information utilisée. Le deuxième mode de parcours autorise notamment le parcours du macrobloc en sens inverse pour fonctionner avec un codage entropique adaptatif.

[4] SP : Switch Picture

[5] QoS : qualité de service



Scan zig-zag. *Scan zig-zag inverse.*

Figure 3 - *table des scans de coefficient dans un bloc*

3.2.1.8 Le codage entropique

Le codage entropique peut-être réalisé de trois manières différentes dans H.26L. Une première méthode utilise une table universelle de mot de code (UVLC - *Unified Variable Lenght Coding*). Cette table est utilisée pour coder la plupart des éléments de synchronisation comme les en-têtes. Les deux autres méthodes sont utilisées pour coder presque tous les autres éléments syntaxiques (coefficients, vecteurs mouvements). Il s'agit d'une part d'un codage VLC adaptatif au contexte (CAVLC - *Context Adaptative Variable Length Coding*) et d'autre part d'un codage arithmétique adaptatif contextuel [8] (CABAC - *Context Adaptative Binary Arithmetic Coding*).

3.2.1.9 L'adaptation au réseau

Les algorithmes de H.26L sont conceptuellement divisés en deux couches : une première couche de codage vidéo (VCL - *Video Coding Layer*) qui s'occupe de représenter efficacement le contenu vidéo et une couche d'adaptation au réseau (NAL - *Network Adaptation Layer*) qui s'intéresse plus particulièrement à une mise en forme des données vidéo adaptée au support de transmission.

3.3 L'efficacité en compression

Afin d'anticiper l'émergence d'un nouveau standard, la communauté MPEG a lancé un appel à proposition fin 2000 [9] dont l'objectif principal était de démontrer l'existence éventuelle de solutions de codage supérieures à MPEG-4 vidéo. Parmi les cinq solutions présentées (UIT-T SG16 Q.6 (H.26L), NTT DoCoMo, daViCo, Bosch GmbH et DynaPel), quatre étaient basées sur la version courante de H.26L qui était alors en cours de normalisation à l'UIT-T.

Pour départager les différentes solutions, six conditions de test ont été définies et ont permis d'évaluer les codecs[6] à deux formats d'image (QCIF et CIF) et à des débits compris entre 32 kbps et 1024 kbps (Tableau 1). Les séquences compressées étaient ensuite évaluées sur une échelle de qualité continue à 5 items : 1-Mauvais, 2-Médiocre, 3-Assez bon, 4-Bon, 5-Excellent (Figure 4).

[6] Codec : Codeur et décodeur

Tests d'efficacité en compression						
Classe de débit	Low			Medium		
Séquences	Foreman News Container Tempete			Bus Mobile and Calendar Flower Garden Tempete		
Résolution	QCIF (176x144)		CIF (352x288)	CIF (352x288)		
Débit (kbps)	32	64	128	256	512	1024
Framerate (fps)	10	15	15	15	30	30
Condition	A	B	C	D	E	F

Tableau 1 – *conditions de test*

C'est la version « MPEG-4 Advanced Simple Profile » qui a été utilisée pour ces tests. Les performances en efficacité en compression des séquences MPEG-4 et H.26L ont largement été améliorées par l'implémentation d'un algorithme « débit/distorsion » au niveau des logiciels de référence [10].

Les tests subjectifs menés [11] par la FUB (Fondazione Ugo Bordoni) ont montré que la solution H.26L de l'UIT-T est significativement meilleure que MPEG-4 vidéo, le gain en PSNR étant en moyenne de 2dB. Quand les comparaisons sont possibles (même séquence vidéo, même nombre d'images par seconde et même format d'image), on constate que H.26L permet de gagner un facteur deux en compression pour une performance équivalente à MPEG-4. En fait, cette comparaison directe des performances entre H.26L et MPEG-4 à des débits différents n'a pu être réalisée que pour un format CIF et aux débits de 512 kbps et 1024 kbps. Des études complémentaires seront donc nécessaires afin valider que ce facteur deux existe à tous les débits et indépendamment du type de contenu.

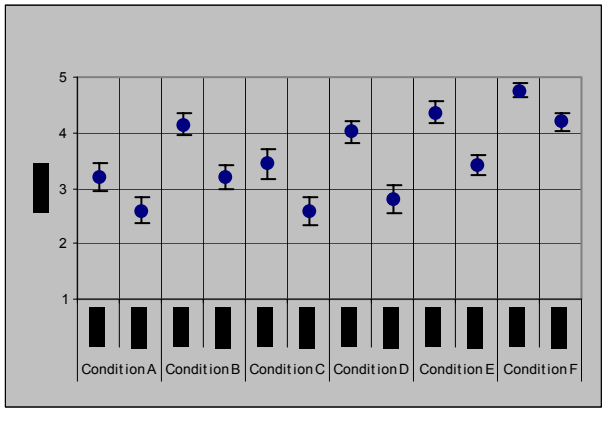


Figure 4 - *résultats sur l'ensemble des séquences*

3.4 Perspectives

Les améliorations algorithmiques apportées par H.264/MPEG-4 « AVC » permettent d'obtenir un gain en qualité significatif par rapport aux solutions de codage actuelles. Les tests subjectifs montrent que cette future norme a le potentiel pour fournir une qualité visuelle équivalente aux solutions actuelles à un débit deux fois plus faible !! Cependant, pour exploiter pleinement ce potentiel, le gain en compression se fait au détriment d'une complexité beaucoup plus importante (mémoire, MIPS[7]). Côté décodeur, la complexité semble équivalente aux solutions actuelles (MPEG-4, H.263) si l'on fait abstraction des images de référence multiples qui demandent une taille mémoire plus importante (n images au lieu d'une). En revanche, le codeur augmente le nombre de prédictions possibles et le délai de codage (blocs de taille variable, images de référence multiples). Dans ce contexte, il est donc essentiel d'anticiper les implications de l'arrivée de cette nouvelle technologie au niveau des services potentiels.

Dans un premier temps, H.264/MPEG-4 « AVC » ne pourra pas répondre de façon optimale à toutes les applications. En effet, certaines d'entre elles seront pénalisées par la complexité de la norme en comparaison avec les standards existants. Ainsi, les applications imposant des restrictions au niveau de la taille mémoire au codage ou au décodage (terminaux mobiles, etc.) pour des raisons de coût ne pourront bénéficier totalement de l'amélioration en qualité. En effet, aujourd'hui le codage d'une image au format QCIF (176x144) demande 30 secondes avec un Pentium IV cadencé à 1,7 GHz !! Dans un premier temps, l'apport des outils innovants de cette nouvelle norme se fera donc principalement dans les applications « offline » telles que la VOD (vidéo à la demande) ou encore le stockage de films sur DVD ou CD-ROM.

Cependant les applications software fonctionnant sur des postes de travail nous poussent à croire qu'une apparition prochaine de codecs temps réel est possible. En effet, sur les Pc du marché (Pentium IV à 2 GHz) on est à environ 3,5 fois le temps réel pour coder une image CIF dans un profil de base de H.264. Le gain obtenu est d'environ 40% par rapport à l'existant et les premiers travaux d'amélioration annoncent déjà des implémentations temps réel.

[7] Millions d'instructions par seconde

Références

[1] H.261 : Video codec for audiovisual services at p x 64 kbit/s. Recommandation H.261 à l'UIT-T. Mars 1993.

[2] H.263 : Video coding for low bit rate communication. Première Recommandation H.263 à l'UIT-T. Mars 1996.

[3] H.263+ : Video coding for low bit rate communication. Deuxième recommandation H.263 à l'UIT-T. Février 1998.

[4] H.263++ : H.263 Annex U, V, W and X. Compléments de la recommandation H.263 à l'UIT-T. Janvier 2000.

[5] Standard MPEG-1 : ISO/IEC 11172-2, Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s.

[6] Standard MPEG-2 : ISO/IEC 13818-2, Information Technology – Generic coding of moving pictures and associated audio information.

[7] Standard MPEG-4 : ISO/IEC 14496-2, Information Technology – Coding of Audio-Visual Objects.

[8] Detlev Marpe, Heiko Schwarz, Gabi Blättermann, Guido Heising, Thomas Wiegand, Context-Based Adaptive Binary Arithmetic Coding in JVT/H.26L, IEEE International Conference on Image Processing (ICIP), Rochester, Septembre 2002.

[9] Call For Proposals On New Tools For Video Compression Technology - ISO/IEC/JTC/SC29/ WG11/N4065, Singapour, Mars 2001.

[10] Heiko Schwarz, Thomas Wiegand, An Improved MPEG-4 Coder Using Lagrangian Coder Control, ITU-T/SG16/Q6/VCEG-M49, Austin, Avril 2001.

[11] Preliminary Results of Subjective Assessment of Responses to Video Call for New Tools to Further Improve Coding Efficiency, ISO/IEC/JTC/SC29/ WG11/N4240, Sydney, Juillet 2001.