

Comment ça marche ?

Les transmissions radio-numériques

19 – Le codage de l'image animée (2)

Par le radio-club F6KRRK

Après avoir vu le principe général MPEG de codage des images animées, nous allons poursuivre avec les codages MPEG-2 et la suite. Se référer à l'article précédent pour les définitions de certains acronymes. Noter que cette série n'a d'autre ambition que de donner une idée générale des processus employés.

Le codage vidéo MPEG-2 (applications broadcast)

Les normes européennes qui couvrent la télédiffusion sont le résultat des travaux démarrés en 1991 et menés par le groupe de travail ELG devenu DVB. Son objectif principal était la définition d'une norme de télévision numérique pour diffusion par satellite, câble ou réseau terrestre.

MPEG-2 constitue la norme pour le codage de source du système défini par le DVB. Elle peut être définie comme une boîte à outils de compression plus complexe que MPEG-1, dont elle peut être considérée comme un "sur-ensemble" car elle en reprend tous les outils en y ajoutant d'autres. De plus la norme prévoit la compatibilité ascendante (un décodeur MPEG-2 doit pouvoir décoder des trains élémentaires à la norme MPEG-1).

MPEG-2 comporte cinq profils qui déterminent le jeu d'outils de compression utilisé et quatre niveaux définissant la résolution de l'image. Pour ceux-ci, leur signification est la suivante :

- Le niveau "low" (bas) correspond à la résolution de MPEG-1.
- Le niveau "main" (principal) correspond à la résolution 4:2:0 "normale" (jusqu'à 720×576).
- Le niveau "high-1440" (haut-1440) est destiné à la TVHD (jusqu'à 1440×1152).
- Le niveau "high" (haut) est optimisé pour la TVHD au format 16/9 (jusqu'à 1920×1152).

Concernant les profils, on retiendra le profil "main" qui correspond au meilleur compromis qualité/taux de compression et le profil "high" prévue pour la télédiffusion HDTV. Il y a compatibilité ascendante entre les profils.

La combinaison "main profil at main level" (MP @ ML) avait été retenue en Europe pour les applications de télédiffusion numérique grand public.

Le codage vidéo MPEG-4

MPEG-4, également appelé "ISO/CEI 14496", est une norme de codage d'objets audiovisuels spécifiée par le "Moving Picture Expert Group" (MPEG).

La norme MPEG-4 spécifie d'abord des techniques pour gérer le contenu de scènes comprenant un ou plusieurs objets audio-vidéo. Contrairement à MPEG-2 qui visait

uniquement des usages liés à la télévision numérique (diffusion DVB et DVD), les usages de MPEG-4 englobent toutes les nouvelles applications multimédias comme le téléchargement et le streaming sur Internet, le multimédia sur téléphone mobile, la radio numérique, les jeux vidéo, la télévision et les supports "haute définition".

La norme spécifie de nouveaux codecs audio et vidéo et enrichit les contenus multimédia, en ajoutant de nouvelles applications comme le VRML (étendu), la prise en charge de la gestion des droits numériques et de plusieurs types d'interactivités.

H.264 ou MPEG-4 AVC

À l'origine, l'UIT-T lança le projet H.26L en 1998 afin de créer une nouvelle architecture de codec ayant pour but un gain en efficacité de codage d'un rapport au moins égal à 2 par rapport aux standards existants (MPEG-2, H.263 et MPEG-4 Part 2). Un autre but était de créer une interface simple pour pouvoir adapter le codec aux différents protocoles de transport (commutation de paquets et de circuits). Le codec a été développé en s'assurant qu'il serait transposable sur plates-formes à un coût raisonnable, c'est-à-dire en tenant compte des progrès réalisés par l'industrie des semi-conducteurs en matière de design et de procédés.

En 2001, l'UIT-T et MPEG décidèrent d'un commun accord de créer le Joint Video Team (JVT) dans le but de standardiser le codec ensemble et de l'adapter aux différents besoins de l'industrie (vidéophonie, streaming, télévision, mobile). En effet, les applications traditionnellement visées par l'UIT-T concernent les bas débits (vidéophonie, mobile), applications pour lesquelles H.26L était optimisé, alors que les membres de MPEG désiraient l'adapter à d'autres formats (télévision, HD). Des outils algorithmiques comme le support de l'entrelacé ont été ajoutés et une réduction de la complexité a été accomplie.

Le nom H.264 provient de la famille de normes vidéo H.26x définies par l'UIT. Cependant, ce codec a été développé dans le cadre du MPEG, l'UIT-T se contentant de l'adopter ensuite et de l'éditer en son sein. Dans le cadre du MPEG, le sigle "AVC" (*Advanced Video Coding*) fut choisi par analogie avec le codec audio "AAC" MPEG-2 part 7 qui avait été nommé ainsi pour le différencier du codec audio MPEG-2 part 3 (le fameux "MP3"). La norme est habituellement appelée H.264/AVC ou MPEG-4 PART 10 (ou les deux accolées pour souligner l'héritage commun). Il existe un précédent dans l'élaboration d'une norme de codage vidéo commune entre le MPEG et l'UIT-T avec "MPEG-2" et "H.262" qui sont identiques. La première version de la norme UIT-T "H.264" a été approuvée en mai 2003. Le JVT a ensuite travaillé sur le concept d'extensibilité en élaborant une extension à la norme H.264 : les spécifications "SVC" (Scalable Video Coding), puis la norme "HEVC" (*High Efficiency Video Coding*).

H.264/AVC (MPEG-4 Part 10) comprend de nombreuses techniques nouvelles qui lui permettent de compresser beaucoup plus efficacement les vidéos que les normes précédentes (H.261, MPEG-1, MPEG-2, MPEG-4 Part2/ASP) et fournit plus de flexibilité aux applications dans un grand nombre d'environnements réseau. Ces fonctionnalités principales comprennent :

"Prédiction Inter-trame" (*Inter-frame prediction*) ou "prédiction temporelle" :

- L'estimation et la compensation de mouvement peuvent être effectuées par rapport à plusieurs images de référence déjà codées. Le choix de l'image de référence intervient au niveau macrobloc et sous-macrobloc. Ceci permet d'utiliser dans certains cas jusqu'à 32 images de référence et jusqu'à 4 références différentes pour un même macrobloc. Si cette fonctionnalité permet des améliorations modestes au niveau du débit et de la qualité dans la plupart des scènes, dans celles contenant des flashes

rapides et répétitifs ou avec des scènes qui réapparaissent fréquemment, il permet une réduction du débit réellement significative.

- Une compensation de mouvement pouvant utiliser 7 tailles de blocs différentes (16×16, 16×8, 8×16, 8×8, 8×4, 4×8, 4×4) permet une segmentation très précise de zones se déplaçant.
- Une précision au quart de pixel pour la compensation de mouvement, permettant une description très précise du déplacement des zones en mouvement. Pour la chrominance, la précision de la compensation de mouvement se fait même au huitième de pixel.
- Une compensation de mouvement pondérée en (*Weighted Prediction*) par des poids et des décalages permettant au codeur de construire des prédictions s'adaptant au changement de luminance et de chrominance de la scène courante.

"Intra prédiction" ou Prédiction spatiale.

"Adaptation des transformées discrètes" :

- Une transformée entière est effectuée sur des blocs de taille 4×4 pixels (proche de la DCT classique). Pour les nouveaux profils issus des extensions FRExt, une transformée supplémentaire de taille 8×8 a été ajoutée.
- Une "transformée de Hadamard" effectuée sur les coefficients moyens de la transformée spatiale primaire pour obtenir encore plus de compression là où l'image est adoucie.

"Plusieurs codeurs entropiques" :

- CABAC (*Context-Adaptive Binary Arithmetic Coding*) : Il s'agit d'un codage entropique qui produit d'excellents résultats en terme de compression au détriment d'une grande complexité.
- CAVLC (*Context-adaptive Huffman Variable Length Coding*) : Il s'agit d'un codage adaptatif de type Huffman à longueur variable pour les tables de coefficients de transformation. Moins complexe que CABAC, CAVLC est plus efficace que les méthodes habituellement utilisées jusqu'à présent pour coder les coefficients.

"Codage à longueur variable" (*Variable length coding*) pour de nombreux éléments de syntaxe non codés par CABAC ou CAVLC, considéré comme du code exponentiel "Golomb".

"Filtrage anti-blocs" (deblocking filter), effectué dans la boucle de codage et opéré sur les blocs 4×4, permettant de réduire les artefacts caractéristiques du codage avec transformation en bloc.

"Deux modes d'entrelacement" :

- Codage en Picture-Adaptive Frame-Field (PAFF ou PicAFF) : L'encodeur choisit d'adapter le codage de la trame courante soit de façon progressive (une seule trame), soit par entrelacement (deux champs ou fields).
- Codage en Macroblock-Adaptive Frame-Field (MBAFF) : Même principe que pour le PAFF sauf que l'application ne s'effectue pas au niveau de la trame mais au niveau des macroblocs 16x16.

"Couche d'abstraction réseau" (NAL : Network Abstraction Layer), définie pour permettre l'usage de la même syntaxe vidéo dans de nombreux environnements réseau. Ceci inclut des possibilités telles que des paramètres de séquence (SPS : *Sequence Parameter Set*) et d'image (PPS : *Picture Parameter Set*) qui offrent plus de robustesse et de flexibilité que les conceptions antérieures.

"Tranches de commutation" (appelées SP et SI) qui permettent à un codeur de diriger un décodeur pour que ce dernier puisse s'insérer dans un flux vidéo entrant, ceci permet du streaming vidéo à débit variable et un fonctionnement en « trick mode » (mode truqué). Quand un décodeur saute au milieu d'un flux vidéo en utilisant cette technique, il peut se

synchroniser avec les images présentes à cet endroit malgré l'utilisation d'autres images (ou pas d'images) comme références préalables au déplacement.

"Ordonnement flexible des macroblocs" (FMO : *Flexible Macroblock Ordering, alias "slice groups"*) et **"l'ordonnement arbitraire des tranches"** (ASO : *Arbitrary Slice Ordering*) sont des techniques de restructuration de l'ordonnement des régions

fondamentales de l'image (macroblocs). Typiquement utilisées pour améliorer la résistance aux erreurs et aux pertes, ces techniques peuvent également être utilisées à d'autres fins.

"Partitionnement des données" (DP : *Data Partitioning*) donne la possibilité de séparer les éléments de syntaxe d'importance plus ou moins élevée dans différents paquets de données. Ceci permet d'appliquer un niveau de protection inégal (UEP : *Unequal Error Protection*) aux erreurs en fonction de l'importance des données et d'améliorer ainsi la fiabilité du flux.

"Tranches redondantes" (RS : *Redundant Slices*) : elles permettent d'améliorer la résistance aux erreurs et aux pertes en permettant au codeur de transmettre une version additionnelle de tout ou partie de l'image dans une qualité moindre qui pourra être utilisée si le flux principal est corrompu ou perdu.

"Processus automatisé simple de prévention" contre la création accidentelle de faux codes de démarrage. Il s'agit de séquences binaires spéciales qui sont placées au sein des données, permettant un accès aléatoire au flux de données ainsi qu'une resynchronisation en cas de perte temporaire du flux.

"Informations supplémentaires d'amélioration" (SEI : *Supplemental Enhancement Information*) et des **informations d'état qualitatif de la vidéo** (VUI : *Video Usability Information*) sont des informations supplémentaires qui peuvent être insérées dans le flux pour améliorer son usage dans un grand nombre d'applications.

"Des images auxiliaires" peuvent être utilisées pour des usages tels que le mixage par alpha channel.

"La numérotation des images" permet la création de sous-séquences (permettant une scalabilité temporelle par l'inclusion optionnelle d'images supplémentaires entre d'autres images) ainsi que la détection et la **dissimulation de la perte d'images entières** (qui peuvent se produire en cas de perte de paquets réseau ou d'erreurs de transmission).

"Le comptage de l'ordre des images" permet de conserver l'ordre des images et du son dans des images décodées isolément des informations de minutage (ce qui permet à un système de transporter, contrôler et/ou changer l'information de minutage sans affecter le contenu des images).

Profils du H.264/AVC

Le standard inclut les six ensembles de caractéristiques suivants, qui sont appelés des *profils*, chacun ciblant une classe d'applications précise :

Baseline Profile (BP) : Principalement pour les applications à bas-coût qui utilisent peu de ressources, ce profil est très utilisé dans les applications mobiles et de visioconférence.

Main Profile (MP) : Prévu à l'origine pour les applications grand public de diffusion et de stockage, ce profil a perdu de l'importance quand le profil *High* a été ajouté avec le même objectif.

Extended Profile (XP) : Prévu pour la diffusion en flux (*streaming*) des vidéos, ce profil a des capacités de robustesse à la perte de données et de changement de flux.

High Profile (HiP) : C'est le profil principal pour la diffusion et le stockage sur disque, en particulier pour la télévision haute définition (adopté pour les disques HD DVD et Blu-ray ainsi que pour la télévision numérique française HDTV).

High 10 Profile (Hi10P) : Ce profil va au-delà des applications grand public et s'appuie sur le profil High en ajoutant jusqu'à 10 bits de précision par pixel.

High 4:2:2 Profile (Hi422P) : Le profil principal pour les applications professionnelles, il s'appuie sur le profil High 10 en ajoutant le support pour la quantification 4:2:2 jusqu'à 10 bits par pixel.

Le "High profile" est en passe de devenir la norme universelle pour les transmissions numériques d'images animées sur tous les supports : diffusions terrestre et satellite, disques vidéo et Internet (streaming).

Quand on voit que l'article ci-dessus ne concerne que l'introduction au standard MPEG-4 AVC, on comprendra que l'on ne poursuit pas avec les détails de mise en œuvre. Dans le prochain "Comment ça marche", on abordera (enfin) les modes numériques radioamateurs.

La Rubrique "Comment ça marche" est une activité collective du radio-club F6KRK (<http://www.f6krk.org>). Pour une correspondance technique concernant cette rubrique : f5nb@orange.fr.

Bibliographie

Nous nous sommes servi de l'ouvrage "La TELEVISION NUMERIQUE" par Hervé Benoit (DUNOD).

Par ailleurs, nous avons beaucoup emprunté à Wikipedia, consultée sur le sujet "MPEG-4".