

MPEG-4

Dès le milieu des années 1990, on pouvait estimer que le MPEG-2 était un vaste succès commercial et qu'il offrait la meilleure solution technologique du moment en ce qui concerne la compression. En 1993, au moment des premiers développements de la nouvelle norme MPEG-4, l'idée originelle était sans doute moins de perfectionner les technologies appliquées dans MPEG-2 et de réduire les débits que d'apporter des fonctionnalités innovantes, notamment liées à l'interactivité au travers d'une nouvelle forme de description de l'image vidéo s'appuyant sur la notion d'objet. C'était la particularité phare de MPEG-4, celle qui était mise en avant. Cette nouvelle norme apparaissait alors plus comme un mode de représentation multimédia que comme un standard de codage comme MPEG-1 et 2.

Au début des années 2000, une inflexion s'est produite. A la suite des recherches menées par différents organismes de normalisation, une nouvelle technologie de compression des données vidéo (H264) était développée. L'objectif était de réduire de moitié les débits utilisés avec MPEG-2. Concrètement, cela devait permettre, non seulement de diffuser de la vidéo sur des supports qui ne le permettait pas jusqu'alors (Internet, ADSL...) mais aussi de multiplier le nombre de programmes diffusés dans un même canal.

Les premières orientations de la norme ont été provisoirement reléguées au second plan et n'ont guère été suivies de réalisations commerciales concrètes. Seules quelques démonstrations matérialisant les recherches effectuées par les sociétés phares (Envivio par exemple) ont été présentées dans les salons spécialisés.

Aujourd'hui c'est cette deuxième piste, la réduction des débits, qui est essentiellement mise en avant. Tous les recherches et tous les développements industriels, toutes les démonstrations vont dans ce sens. H264 sera utilisé dans la télévision numérique terrestre, la haute définition, les nouveaux DVD mais aussi sur ADSL. Il est vrai que les progrès réalisés sont spectaculaires. Nous y reviendrons. Lorsque les savoir-faire seront parfaitement maîtrisés, lorsque la technologie sera telle que l'on pourra réaliser des processeurs plus performants et suffisamment rapides, alors les techniques de codage des objets ressortiront de leurs cartons et les développements reprendront leur cours dans le sens de l'interactivité.

Par rapport à ses devancières MPEG-1 et MPEG-2 dont le champ d'application se limitait pour chacun à un seul domaine, MPEG-4 a introduit une véritable révolution par l'étendue du champ couvert. Ce nouveau standard se plaçait en effet à la convergence de trois mondes : la télévision, l'informatique et les télécommunications.. Au fur et à mesure, la norme a élargi progressivement ses domaines d'activité et constitue désormais un standard très ambitieux et très complexe visant de nombreuses applications : diffusion TV, postproduction, télévision numérique terrestre (TNT), télévision haute définition, DVD, téléphonie, visiophonie, streaming, télésurveillance... mais aussi des applications plus confidentielles comme celles qui reposent par exemple sur les capacités de gestion de personnages virtuels.

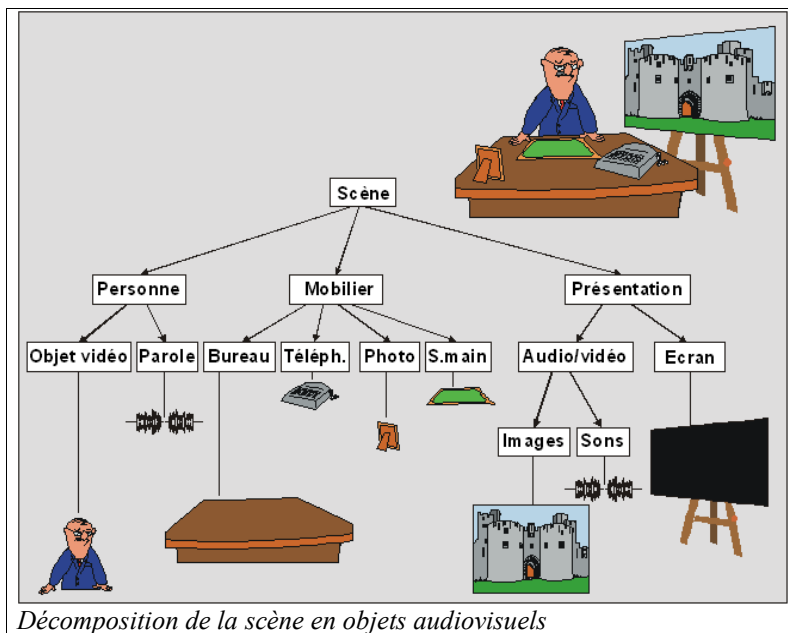
MPEG-4 a été conçu initialement pour les bas débits. Aujourd'hui, sa plage d'utilisation s'est étendue et MPEG-4 est utilisable dans une très large gamme de débits : de quelques Kb/s pour les bas débits jusqu'à une centaine de Mb/s et plus pour les développements les plus récents.

Les premières orientations : une approche orientée “objets”

La description des images vidéos :

Le principe adopté dans MPEG-4 est celui d'une description autonome du contenu : l'image n'est plus codée dans sa globalité, c'est à dire en la considérant indépendamment de son contenu comme une surface rectangulaire de X sur Y pixels, mais elle apparaît comme une composition réalisée avec différents “objets” audiovisuels. Ces objets peuvent être de différentes natures : image fixe (par exemple, le décor du fond, un tableau), objets vidéos (le personnage sans le décor), objets audios (la voix de la personne, le fond musical)... Ils peuvent être naturels ou de synthèse, en 2D ou en 3D. Un texte, un graphique, un son synthétique... sont également considérés comme des objets. Chacun de ces objets est codé indépendamment des autres, et isolé du fond ou de l'environnement

Une hiérarchisation des contenus est également établie : les objets “basiques” correspondant à une même entité (par exemple l'image d'un personnage et sa voix) peuvent être réunis et créer un nouvel objet composite comprenant l'ensemble des attributs. De tels regroupements vont, non seulement faciliter le travail de conception des auteurs mais également permettre la manipulation des contenus par les utilisateurs. La notion d'objet peut théoriquement s'appliquer également au sein même de la vidéo. La génération de ces différents objets (par mask ou shape) ne fait pas partie du format mais relève de solutions de post-production de type "alpha channel". Cette technique se révèle aujourd'hui assez délicate à utiliser et n'a pas encore fait l'objet de développements importants.



Le langage de MPEG-4 qui permet la description et la modification dynamique d'une scène s'appelle BIFS (Binary Format for Scenes). Il est très largement issu du VRML (Virtual Reality Modeling Language) pour la description et l'interactivité des objets 3D auquel il ajoute certaines fonctionnalités.

Ce découpage sous la forme d'objets se retrouve bien évidemment au niveau des flux de données : à chacun des objets va correspondre un flux élémentaire.

Cette séparation des flux offre au moins deux avantages :

- En premier lieu, au niveau du codage de chacun des objets. Elle permet d'appliquer pour chacun d'entre eux la technique de compression la mieux adaptée à sa nature et

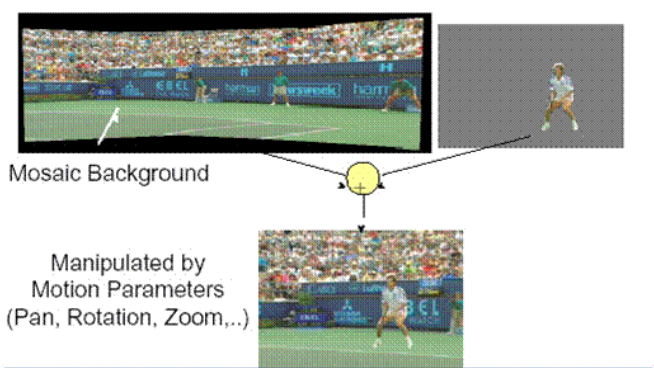
les technologies de transmission les plus efficaces, notamment en termes de qualité de service (des documents textes, des tableaux... ne peuvent tolérer aucune perte de données mais les débits peuvent être faibles, ce qui n'est pas le cas pour la vidéo qui peut s'accommoder de quelques pertes ou d'erreurs de transmission mais pour lesquels les débits sont élevés). MPEG-4 permet de synchroniser des flux venant de différents serveurs : il devient possible de diffuser des textes ou des photos synchronisées sur la vidéo avec toute la qualité requise pour chacun d'eux.

- En second lieu, chacun de ces objets retrouvant sa propre autonomie au niveau du terminal usager, on peut imaginer toutes les formes d'interactivité possibles. MPEG4 a été directement conçu pour le Rich Media. Avec ce standard, une scène audiovisuelle n'est pas figée et il devient possible d'agir (si cela a été autorisé par le concepteur du programme) sur les différents composants : déplacer un objet, changer son apparence géométrique ou acoustique voire le supprimer, modifier ses attributs (son ou langue, texture d'une surface, paramètres d'un objet synthétique), changer le point de vue et d'écoute du spectateur (comme s'il se déplaçait devant la scène). Qui plus est, le terminal étant muni d'une voie retour, le spectateur peut intervenir dans le déroulement de la présentation et provoquer des actions (par la sélection de zones actives).

Ces mêmes flux devraient également pouvoir contenir un certain nombre d'informations associées à des objets : renseignements sur le contenu ("metadonnées"), droits d'auteurs (identification, protection et gestion des droits).

La norme prévoit que certains de ces objets peuvent ne pas être transmis et être stockés en permanence dans le récepteur ou encore être transmis une fois pour toutes en début de séquence ou de manière progressive. La notion de "sprite" correspond à un objet vidéo persistant pendant la séquence vidéo et qui peut être plus large que la vidéo affichée. Typiquement, un décor ou un arrière plan. Les "sprites" sont susceptibles d'améliorer spectaculairement la compression.

Sprites



Certains objets peuvent n'être transmis qu'une fois et stockés dans le récepteur : ici le décor.

Encadré :

On a entendu parfois dire que MPEG 4 était fondé sur QuickTime. Qu'en est-il ? Lorsqu'il s'est agi de définir un container c'est à dire un format de fichier, le groupe de travail MPEG-4 s'est d'abord inspiré du format correspondant de QuickTime. Ce container ne constituait qu'une partie de la norme. Ultérieurement, ce format a d'ailleurs été modifié, supprimant la compatibilité. La version 6 de QuickTime a intégré nativement le support de MPEG 4.

Un flux unique, un format de fichier

Tous ces flux sont multiplexés en un flux unique pour le transport, ce qui signifie également dans un cadre purement informatique, un seul fichier qui contient la totalité des informations.

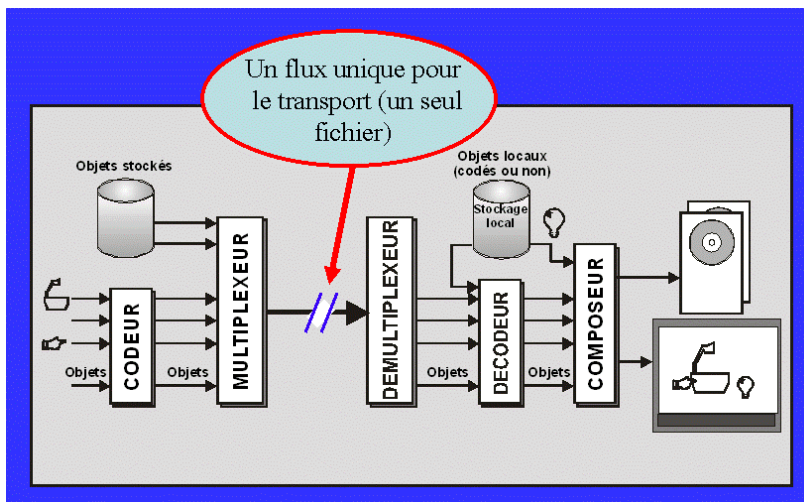


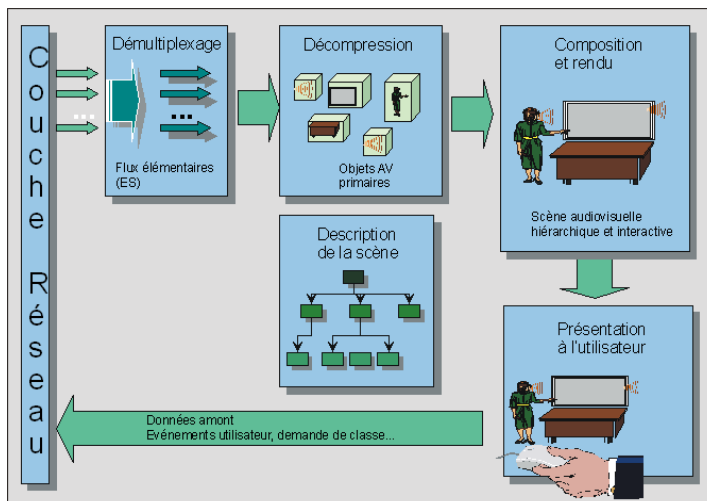
Schéma de transmission et de reconstitution de la scène

Encadré :

La couche FlexMux (Flexible Multiplexing) procède au regroupement dans le même flux des flux élémentaires présentant des points communs (par exemple, ceux présentant les mêmes exigences en matière de QoS). Elle s'appuie sur les spécifications DMIF (Delivery Multimedia Integration Framework). DMIF traite de l'interface entre les applications et le réseau de transport, et de l'établissement des canaux de communication en fonction des QoS nécessaires pour chaque flux élémentaire. La couche TransMux (Transport Multiplexing) gère les services de transport sur le réseau externe : (RTP) UDP/IP, (AAL5)/ATM, MPEG-2 TS... Seule l'interface avec cette couche est décrite dans la norme MPEG-4. (le transport ne fait pas partie de la norme).

C'est seulement au niveau du décodeur final que les différents éléments sont recomposés et synchronisés pour restituer l'image finale. En clair, le terminal de l'utilisateur reçoit un programme en kit. Il reconstitue ce programme en n'utilisant que les éléments qui lui sont nécessaires ou qui sont compatibles avec son terminal.

De MPEG 2 à MPEG 4, on est passé d'une approche statique de la composition de l'image à une approche dynamique qui ne se matérialise désormais qu'au stade du décodage.



Synoptique du terminal de réception

H264 : une avancée considérable dans les techniques de compression

MPEG-4 : les bases sont dans MPEG-2

La numérisation d'une image vidéo génère un débit très élevé difficilement conciliable dans la majorité des cas avec les technologies de diffusion d'aujourd'hui. Il est donc nécessaire de procéder à une réduction du volume des données numériques. Cette opération doit s'effectuer tout en conservant la meilleure qualité possible pour les images. C'est la compression.

Encadré

Pour réaliser cette opération, les systèmes MPEG s'appuient sur de nombreuses technologies notamment celles qui exploitent les redondances :

- redondances spatiales :
Dans les grandes plages uniformes à l'intérieur d'une image, on exploite les similitudes entre des zones voisines et on ne re-encode pas les points qui sont sensiblement identiques.
- redondances temporelles
Dans une séquence vidéo, les différences entre deux images successives sont minimales : on ne code que les changements.
- Redondances subjectives :
Il est inutile de coder les détails fins que l'œil ne voit pas.

Les « coupes » sont drastiques mais l'œil s'en accommode. Les résultats sont spectaculaires et permettent en diffusion aujourd'hui de réduire les données de plus de 95% ! Mais les résultats sont variables en fonction de la nature même des images.

Pour l'essentiel, les principes de la compression étaient restés inchangés ces dernières années. Comme ses prédécesseurs MPEG 1 et 2, MPEG 4 comporte pour le traitement des images animées quatre phases principales : division de l'image en blocs de pixels ; exploitation intra-bloc de la redondance spatiale via une transformation mathématique (la DCT - Discrete Cosinus Transform), quantification et codage entropique ; exploitation de la redondance temporelle inter-bloc en ne transmettant que les seuls changements via des techniques de prédiction et de compensation de mouvement ; deuxième élimination des redondances spatiales en codant les blocs résiduels (différence entre originaux et prédicts).

Sur un plan strictement vidéo, le premier successeur de MPEG-2 fut MPEG-4 Part 2 publié en 1999 (mais sans lui être compatible !). Cette nouvelle norme a permis quelques améliorations par rapport à MPEG-2 mais sans procurer un gain énorme de la compression (entre 15 et 20 % selon un document de l'UER). Ce n'est plus le cas avec les innovations radicales apportées avec les dernières évolutions de la norme. Elles constituent une véritable rupture technologique en permettant d'emblée, on le verra, une réduction de 50% de la bande passante nécessaire à qualité égale par rapport aux premières versions de MPEG 4.

H264, MPEG-4 AVC ou MPEG-4 Part 10

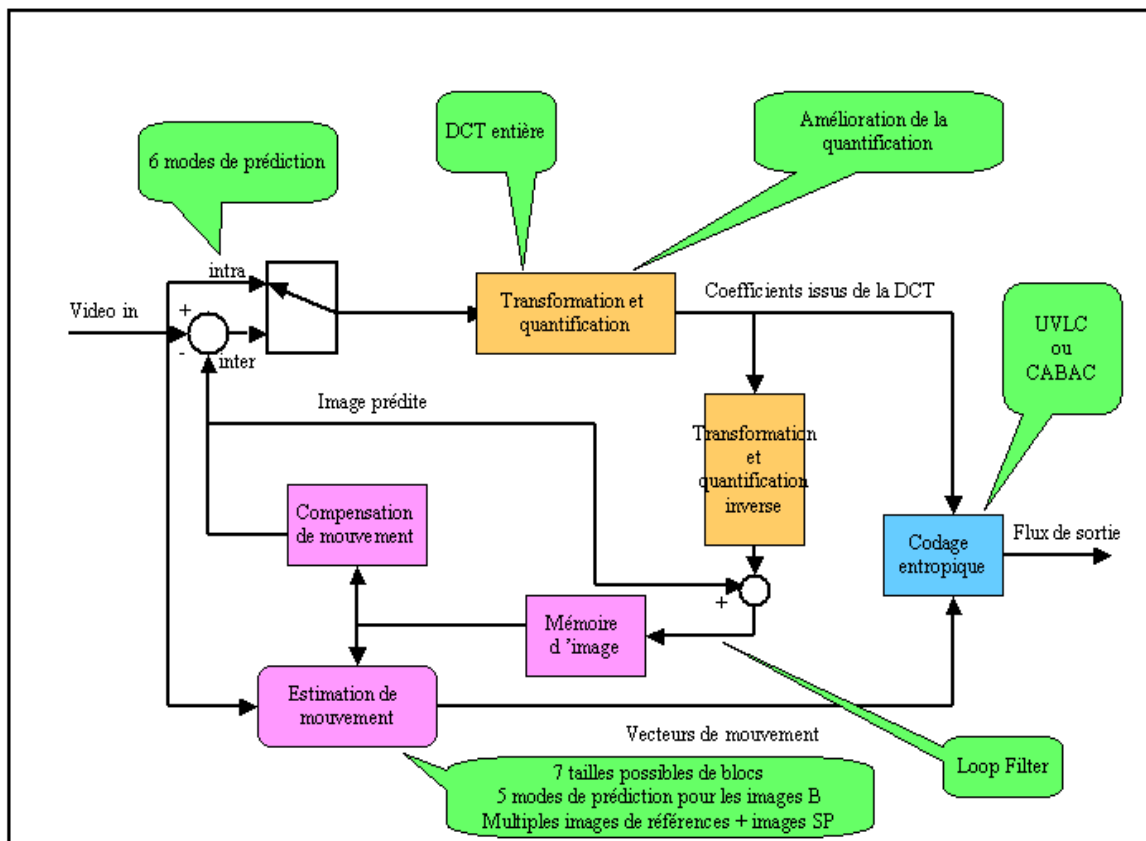
Ces nouveaux développements sont issus des recherches d'un groupe de travail commun aux deux groupes de normalisation les plus actifs au niveau international : ITU (International Telecommunication Union) et ISO (International Standards Organisation). Ces deux groupes ont unis leurs efforts pour développer un nouveau système de codage et le normaliser, d'où la triple appellation que l'on trouvera dans la littérature spécialisée : MPEG-4 AVC (Advanced Video Coding) ou MPEG 4 part 10 pour l'ISO et H264 pour ITU. C'est tout récemment (en sept 2004) que les derniers développements de ces nouveaux standards ont été définis et approuvés.

MPEG 4 part 10 conserve les quatre phases principales de traitement de la DCT classique mais elle améliore chacune d'entre elles grâce à l'introduction de nouvelles techniques plus sophistiquées. Elle raffine ainsi le codage intra en introduisant une prédiction spatiale à 6 modes. En plus des traditionnelles images I, B et P, la future norme introduit deux nouveaux modes de prédiction des macroblocs, les modes de transition SI et SP destinés à favoriser la commutation entre flux de débits variables, les accès aléatoires ou les visionnements accélérés. MPEG 4 part 10 propose également la sélection de multiples images de référence qui servent à définir SP.

Mais c'est surtout en matière d'estimation de mouvement que sont réalisés les gains en efficacité les plus marquants. MPEG 4 part 10 innove avec un système adaptatif fondé sur 7 types de blocs de pixels regroupant au choix 4x4, 4x8, 8x4, 8x8, 8x16, 16x8, 16x16 pixels pouvant être utilisés pour affiner l'estimation de mouvement, occasionnant un gain d'efficacité de 15% par rapport au bloc unique 16x16. Le recours à une compensation de mouvement précise au quart de pixel (et 1/8 ultérieurement) augmente la précision de celle-ci et procure un gain de 20%.

En matière de transformation, MPEG 4 part 10 a recours à une transformation spatiale DCT entière, approximation de la DCT classique qui réduit les défauts lors de l'opération inverse.

Le codage entropique (optimisant la longueur des codes, les plus courts étant affectés aux symboles les plus fréquents) utilise soit classiquement des Universal Variable Length Codes (UVLC) soit le nouveau et sophistiqué Context-based Adaptive Binary Arithmetic Coding (CABAC) qui tient compte des corrélations possibles entre les symboles d'encodage en se fondant sur un modèle probabiliste. Ce qui procure une amélioration de l'ordre de 10% surtout pour les bas débits.



H264 apporte des améliorations à tous les stades de l'encodage

Des gains spectaculaires en qualité

Comme ses devancières, la norme MPEG 4 ne spécifie pas l'encodage (ce qui laisse toute latitude aux industriels pour améliorer leurs codecs) mais définit seulement la sémantique du décodage (garantissant ainsi la compatibilité pour les utilisateurs). Une particularité qui permet une forte marge d'évolution en tirant parti des progrès constants en matière de traitement informatique, à l'instar de MPEG-2 qui a vu diviser par un facteur proche de 3 le débit nécessaire pour transmettre la même qualité d'image entre 1996 et 2003.

La compression MPEG-2 atteint aujourd'hui ses limites. Ce n'est pas le cas avec MPEG-4 qui est une technologie toute récente et qui sera donc sujette, à l'image de ce qui c'est déroulé pour MPEG-2, à de nombreuses améliorations au fil du temps.

On annonce souvent que MPEG-4 AVC apporte aujourd'hui un gain de 50 % en terme de débit par rapport à MPEG-2. Ce chiffre doit cependant être quelque peu nuancé, d'une part parce que la notion de qualité d'une image vidéo numérique est une caractéristique difficile à évaluer, d'autre part parce que les effets des algorithmes de compression varient suivant la nature même des images, et enfin parce que les bénéfices qualitatifs du nouveau standard par rapport à l'ancien sont variables en fonction des plages de débits concernés, l'amélioration étant plus particulièrement marquée pour les plus bas débits. En 2003, l'UER (Union Européenne de Radio-Télévision) indiquait, lors d'une comparaison entre MPEG-4 Part 10 et MPEG-2 que si le gain qualitatif pouvait être évalué entre 40 et 50% pour des applications de télévision standard (SDTV), il n'était plus que de 20 à 40% pour des formats HD mais pouvait être compris entre 50 et 60 % pour des images en quart d'écran au format CIF (Common Interchange Format).

Les résultats d'aujourd'hui sont plus que prometteurs pour les années à venir. Thomson, dans un tableau qui a été publié en 2004 dans le rapport du CGTI (Conseil Général des Technologies de l'information) destiné au Ministre de l'industrie, chiffre l'évolution des débits moyens de MPEG 4 et la marge de progression attendue pour les années à venir. Pour la télévision standard, par comparaison avec un débit de 3,8 Mb/s pour le standard MPEG-2 (et qui ne bougera sensiblement plus dans les années à venir, rappelons le) et à qualité d'image égale, le débit nécessaire en MPEG-4 AVC pourrait n'être que de 2,7 Mb/s à la fin de cette année, puis de 1,9 Mb/s (la moitié de ce qui est nécessaire en MPEG-2 !) au second semestre 2006 pour ensuite atteindre 1,6 Mb/s à la fin de l'année 2007. A la même époque, la haute définition serait possible pour des débits compris entre 5 et 8 Mb/s.

Débit nécessaire	Sept 2005	2S 2006	2S 2007
HD- MPEG 2	12-18 Mbps	12-18 Mbps	12-18 Mbps
HD- MPEG 4	8,4 - 12,6 Mbps	6,0 - 9,0 Mbps	5,0 - 7,6 Mbps
SD- MPEG 2	3,75 Mbps	3,75 Mbps	3,75 Mbps
SD- MPEG 4	2,625 Mbps	1,875 Mbps	1,575 Mbps
Amélioration de la performance des encodeurs			
SD- MPEG 4 / SD- MPEG 2	30%	50%	58%
HD- MPEG 4 / HD- MPEG 2	30%	50%	58%

Evolution des débits de MPEG-4. Source: Thomson cité dans CGTI

Les bénéfices techniques et économiques de cette réduction des débits sont importants.

- En premier lieu avec la possibilité de transmettre des programmes de qualité sur des canaux qui ne le permettaient pas jusqu'alors. MPEG 4 part 10 se pose ainsi comme la solution du futur pour la diffusion de la télévision standard sur des réseaux IP avec des

débits de l'ordre de 1,5 Mb/s. Un débit qui serait accessible dès la fin de l'année 2007 pour les nouvelles générations de normes ADSL, si on se réfère aux prévisions de Thomson.

- En second lieu, avec la multiplication du nombre de programmes qui peuvent être diffusés sur un canal donné.

Ces améliorations auront toutefois un coût. Pour la télévision standard (au format 4/3) une multiplication par quatre de la complexité du décodage par rapport à MPEG-2 et par 8 ou 10 de l'encodage. Une complexité qui nécessite des terminaux disposant d'une puissance de calcul suffisante. MPEG-4 est aujourd'hui normalisé, stable et en cours d'industrialisation : au dernier IBC, de nombreux encodeurs en temps réels et décodeurs (chips) ont été annoncés. Pour la TNT en France, les premiers décodeurs multi-standards (MPEG 2 et MPEG 4 SD) devraient être mis sur le marché grand public à partir de l'automne 2005. Tout n'est pas abouti, loin s'en faut : les chips d'encodage et de décodage pour les équipements mobiles relèvent encore du défi technologique.

MPEG-4 AVC et la haute définition

MPEG-4 AVC sera aussi la norme du futur en ce qui concerne la haute définition. Avec les derniers développements réalisés autour du codage H264, le débit « raisonnable » pour la haute définition se situe aujourd'hui autour de 9 à 12 Mb/s, mais il devrait rapidement baisser compte tenu des progrès technologiques qui vont être réalisés dans le domaine de la compression. Toujours selon les projections qui ont été effectuées par Thomson, la fourchette pourrait être comprise entre 6 et 9 Mb/s au cours du deuxième semestre 2006 (des débits comparables à ceux qui étaient nécessaires en 1996 pour un flux SD MPEG-2), et même sans doute descendre entre 5 et 8 Mb/s pour la fin de l'année 2007. Ces chiffres sont des valeurs moyennes. Ils sont à moduler en fonction du contenu et de la complexité des images : à qualité subjectivement équivalente, la retransmission d'images de sport nécessite un débit supérieur à celui qui est nécessaire pour la diffusion d'un film (par exemple 9 Mb/s pour le premier, 6 Mb/s pour le second). Avec de tels résultats, il deviendra possible de diffuser jusqu'à 5 ou 6 programmes en haute définition là même où on ne proposait il y a encore que quelques années qu'un seul et unique programme en analogique (un répéteur satellite permet un débit d'environ 40 Mb/s, pour le câble c'est un peu moins).

Une étude comparative de l'efficacité de MPEG-4 AVC par rapport à MPEG-2 a été réalisée par la Blu-Ray Disc Association. Ces tests ont été réalisés "en aveugle" avec un panel de spectateurs à qui l'on proposait des séquences vidéo encodées en MPEG-2 et en MPEG-4 AVC à différents débits. La majorité des spectateurs ne perçoivent pas de différence entre la vidéo originale (non compressée) et la séquence en MPEG-4 AVC lorsque celle-ci est encodée avec un débit de 16 Mb/s. Jusqu'à cette valeur, l'impression de qualité croît avec le débit mais il est inutile d'aller au-delà, le « plus » qualitatif n'étant plus perceptible. La perception de l'amélioration qualitative n'est pas toujours proportionnelle à l'augmentation de débit : en accroissant le débit de 8 à 12 Mb/s (soit une augmentation de 50%), la note donnée par les spectateur reste pratiquement constante.

A impression visuelle constante, il faut un débit trois fois plus élevé avec MPEG-2 qu'avec MPEG-4 (24 Mb/s contre 8 Mb/s avec MPEG-4 AVC).

MPEG-4 : une norme complexe

Une norme complexe : 21 parties aujourd'hui

Rançon de son ambition de couvrir tout le champ du multimédia et de l'audiovisuel, y compris désormais la vidéo professionnelle, la norme MPEG 4 adopte une approche modulaire et se caractérise par une complexité certaine que traduisent les divers livres (parts) qui la composent.

Le coeur est constitué de la partie Systems (part 1), Visual (part 2) et Audio (part 3) auxquels il faut ajouter DMIF (Delivery Multimedia Integration Framework, part 6) qui définit l'interface entre la couche application et celles liées au stockage et au réseau. Conformance (part 4) expose comment tester une implémentation de la norme et la part 5 propose un Reference Software. La part 7 propose un exemple d'encodeur optimisé. Si le transport n'est pas défini dans la norme, la part 8 définit cependant comment intégrer un flux MPEG 4 dans un réseau IP. La part 9 donne une description de Reference Hardware.

La part 10, concerne les derniers développements en matière de réduction de débit autour du codec H264. C'est devenu la facette principale de MPEG-4 qui éclipse aujourd'hui la part 2.

D'autres tomes ont été définis ou sont en préparation : un additif, issu de la part 1 et spécifiquement destiné à la description de scènes (part 11) ; l'ISO Media File Format (part 12) ; une extension de la gestion IPMP (Intellectual Property Management and Protection, version MPEG 4 du DRM, part 13) ; MP4 File Format (part 14, fondé sur part 12) ; AVC File Format (part 15, également fondé sur part 12) ; AFX (Animation Framework eXtensions) et MuW (Multi-users Worlds) pour la part 16....

Part 1	Systems
Part 2	Visual
Part 3	Audio
Part 4	Conformance testing
Part 5	Reference Software
Part 6	Delivery Multimedia Integration Framework
Part 7	Optimised software for MPEG-4 tools
Part 8	MPEG-4 on IP framework
Part 9	Reference Hardware Description
Part 10	Advanced Video Coding
Part 11	Scene Description and Application Engine
Part 12	ISO Base Media File Format
Part 13	IPMP Extensions
Part 14	MP4 File Format
Part 15	AVC File Format
Part 16	Animation Framework eXtension (AFX)
Part 17	Streaming text format
Part 18	Font compression and streaming
Part 20	Lightweight Application Scene Representation
Part 21	MPEG-J extensions for rendering

21 parties aujourd'hui pour définir la norme MPEG-4

Scalability : « échelonnabilité »

Ce mot revient à diverses reprises dans les textes de présentation de MPEG-4. Il signifie "adaptabilité", "échelonnement"... et désigne la possibilité de transmettre et de fournir dans un même signal tous les éléments correspondant à différents niveaux de qualité (codage hiérarchique) qui pourront être exploités différemment en fonction des conditions de transmission ou du type de terminal utilisé par l'utilisateur. La totalité des informations peut ainsi être ainsi transmise selon un flux principal et des flux complémentaires qui ne sont exploités qu'à la mesure de la performance du réseau ou du récepteur utilisé. On pense notamment à la réception d'un même contenu sur un téléphone portable et sur un récepteur de télévision fixe ou mobile. La qualité de restitution du document est directement proportionnelle à la complexité du terminal. Avec les décodeurs les plus performants, on assure le décodage de la totalité du flux ce qui permet une restitution optimum tandis qu'avec un terminal plus basique, on ne décode d'une partie du flux et on obtient la restitution d'un document de qualité moindre.

Différentes solutions sont envisageables :

- "Spatial scalability" (adaptabilité spatiale) : un décodeur peut n'utiliser qu'une fraction du flux de données ce qui permet l'affichage des images et des vidéos avec une résolution spatiale réduite (cela peut être par exemple l'utilisation de 1 pixel sur x, ce qui donne une image plus petite mais une sensation de qualité et une fluidité maintenue).
- "temporal scalability" (adaptabilité temporelle) : une séquence vidéo peut être reproduite avec une résolution temporelle réduite. Différentes méthodes sont possibles : réduction du nombre d'images par seconde, augmentation des intervalles d'estimation (prédiction, extrapolations...).
- "quality scalability" (adaptabilité qualitative) : la répartition des données en un certain nombre de couches (un niveau de base complété par une ou plusieurs couches d'extension (Enhancement Layer) pouvant permettre par exemple un affichage progressif de la qualité. L'image a une définition réduite au démarrage puis s'affine si les conditions le permettent. Les améliorations peuvent en outre porter sur la totalité de l'image ou simplement sur une portion (un objet).

En s'appuyant sur ces principes, le groupe MPEG a décidé il y a environ deux ans de démarrer les premiers travaux pour la définition d'un nouveau standard baptisé MPEG SVC (pour Scalable Video Coding) qui présenterait la particularité, à partir d'un seul flux (et donc d'une phase unique d'encodage), de satisfaire à tous les cas de figure possible dans des schémas de distribution comportant un grand nombre de configurations possibles.

Une structure hiérarchique fondée sur des profils et niveaux

Davantage qu'une norme unique, MPEG 4 constitue en fait une formidable boîte à outils regroupant une variété de solutions de compression correspondant à différentes situations d'utilisation. Reprenant les structures qui avaient été développées pour la norme MPEG 2, on a également défini pour MPEG-4 une série de profils et de niveaux (profiles et levels dans la littérature anglaise). Cette notion de profils a été définie, non seulement pour la vidéo et l'audio (audio profiles et visual profiles), mais également pour la partie système de la norme : profils graphiques, profils relatifs à la description de scènes...

Les profils correspondent à une approche qualitative : ils regroupent les outils de codage et les algorithmes les plus appropriés pour diverses situations d'utilisation : streaming, mobile, diffusion, production... Les profils Visual définissent ainsi les types d'images qui seront

utilisés (Images I, P, B...), le mode de codage employé pour la chrominance (4.2.0, 4.2.2...), la précision de la prédiction de mouvement...

Les décodeurs n'auront pas à implémenter tous les outils de la norme mais seulement ceux destinés à l'usage visé. Limitant les choix, restreignant volontairement la syntaxe de décodage susceptible d'être utilisée, les profils jouent un rôle essentiel du point de vue de l'interopérabilité. Ils garantissent le fait qu'un flux généré par un encodeur particulier pourra être décodé par tous les décodeurs acceptant ce profil.

Chaque profil comporte ensuite différents niveaux. Ils apportent une approche quantitative, décrivant différents paliers de performance pour un profil donné et correspondant à autant de degrés de complexité dans la mise en oeuvre. Les niveaux imposent des contraintes à certains paramètres clés du flux. Ils vont par exemple spécifier les limites supérieures pour la définition des images, le nombre d'image par seconde, les gammes de débits utilisables, le nombre d'objet...

Le couple Profiles et Levels constitue ainsi la signature caractérisant une conformité.

Face à la diversité et à la complexité des profils, les consortiums Mpeg IF (MPEG Industry Forum) et ISMA (Internet Streaming Media Alliance) ont pris des initiatives intéressantes en matière de certification et d'interopérabilité. MpegIF a présenté un programme de qualification garantissant la compatibilité des matériels et la conformité des performances et qui débouche sur l'attribution d'un logo MP4.

L'Internet Streaming Media Alliance propose de constituer un écosystème garantissant l'interopérabilité en s'intéressant à la spécification du système de bout en bout. Le mécanisme de certification qui se met en place et qui repose sur des auto-vérifications faites par les fabricants validées ensuite par l'ISMA selon une procédure standardisée, veut aider les utilisateurs à identifier les produits compatibles. Pour ce faire, l'ISMA a élaboré un programme de conformité reposant sur des produits et fichiers de référence et du matériel de test. La version 2.0 des spécifications de l'ISMA fondée sur les codecs avancés audio et vidéo sera publiée pour commentaires très prochainement. Le travail se poursuivra ensuite à la fois sur les nouvelles versions de codecs et sur la gestion des droits (DRM).

Un maquis touffu

La segmentation en profils permet aussi de faire évoluer plus facilement la norme mais au prix d'un accroissement certain de sa complexité. Même si les divers profils constituent souvent un simple enrichissement de versions de base, leur multiplication donne naissance à un maquis assez impénétrable et nuit à la lisibilité de l'ensemble.

La version 1 de la norme comportait neuf profils Visual. Cinq correspondaient à des images naturelles et 4 à des images de synthèse. Les premiers comprenaient le Simple Visual Profile destiné aux applications mobiles, le Simple Scalable Visual Profile ouvrant la voie à plusieurs niveaux de réception, le Core Visual, ajoutant au Simple des objets de taille arbitraire afin d'autoriser l'interactivité ; le Main, ajoutant entrelacement et sprite au Simple ; le N-Bit destiné aux applications de surveillance. Les seconds étaient constitués du Simple Facial Animation dont le nom définit la fonction ; du Scalable Texture destiné aux textures d'images fixes à niveaux de qualité variables destinés aux jeux ; le Basic Animated 2D Texture combinait les deux applications précédentes ; le Hybrid associait le dernier au Core.

La version 2 de la norme a ajouté trois profils à chacune des catégories. D'une part, l'Advanced Real-Time Simple améliorant la résistance aux erreurs et destiné aux applications de visioconférence et visioconférence ; Le Core Scalable, orienté Internet ; l'Advanced Coding

Efficiency pour la réception broadcast mobile. D'autre part, l'Advanced Scalable Texture optimisé pour l'accès aléatoire aux images fixes sur Internet ou les PDA ; l'Advanced Core, plus spécialement adapté au Rich Media ; le Simple Face and Body Animation, améliorant le Simple Facial.

Les versions suivantes ont ajouté l'Advanced Simple Profile, qui améliore les possibilités du Simple ; le Fine Granularity Scalable, le Simple Studio Profile et le Core Studio Profile.

En pratique, seul un nombre restreint de profils est actuellement implémenté. Le Simple Profile et l'Advanced Simple Profile (ASP) constituaient actuellement les deux profils vedettes en matière d'image naturelle animée jusqu'à l'apparition de H264. Ils restent encore d'actualité pour les applications mobiles.

Les profils de H264

L'avènement de H264 a apporté quatre nouveaux profils : Baseline (pour les applications avec une résolution réduite, comportant seulement des images I et P, avec une précision de compensation au 1/4 de pixel) ; Extended (spécialement dédié au streaming et aux mobiles ; il s'agit d'une extension de Baseline avec davantage de résistance aux erreurs de réseau) ; Main (qui emploie les images I, P et B avec un balayage entrelacé, et utilise CABAC...) et High pour les usages professionnels (production, contribution).

Profils AVC	Application cible
Baseline	Basse résolution, TV mobile, applications de visioconférence. DVB-H
Extended	Streaming, mobiles
Main	Vidéo entrelacée, broadcast, diffusion TV numérique standard
High Profile	HDTV

Domaines d'application des profils AVC

Jusqu'à l'année dernière, MPEG-4 était surtout ciblé sur la diffusion et H.264 s'imposait avec le Main Profile dans le monde du broadcast. Très logiquement, les comités techniques avaient cependant poursuivi le travail d'exploration vers de nouveaux profils en 4 :2 :2 ou plus avec un échantillonnage sur 10 ou 12 bits répondant à des besoins professionnels notamment en contribution. Ce travail prospectif s'est brusquement accéléré, semble-t-il sous la pression des partisans du DVD haute définition, pour aboutir à l'émergence surprise d'un nouveau profil orienté HD, le High Profile. Le Blu-Ray Disc avait en effet formulé des exigences pour des profils de très haute qualité avec des hauts débits (15-24 Mbits/s) et des formats d'image HD 1920x1080 et 1280x720. Quatre nouveaux profils ont donc été ajoutés à H.264 sous l'appellation générique de H.264 AVC FRExt pour Fidelity Range Extensions. Ils ont tous en commun d'être des extensions du Main Profile complété par une matrice de macro-blocs de quantification 8x8 de la prédiction de mouvement. Au final, semble-t-il, une réelle amélioration de l'ordre de 20% à ces hauts débits. Ces nouveaux profils correspondent à une logique d'empilement en peaux d'oignon, chaque nouveau profil intégrant le précédent. De plus, ils ont en commun de pouvoir supporter en option un Alpha Channel de haute qualité et une modélisation des caractéristiques du grain du film.

Nom du profil	Format max chroma	Echantillonnage max
High (HP)	4:2:0	8 bits
High 10 (Hi10P)	4:2:0	10 bits
High 4:2:2 (H422P)	4:2:2	10 bits
High 4:4:4 (H444P)	4:4:4	12 bits

Les nouveaux profils AVC (source document ISO/IEC JTC1)

Les profils audio

Différents profils ont été également définis pour l'audio.

Pour le son, MPEG 4 offrait quatre profils dans la version 1 : le Speech, fondé notamment sur le codec CELP adapté aux bas débits (de 3,6 à 24 Kbps) ; le Synthesis, destiné aux interfaces Text-To-Speech (TTS) ; le Scalable, amélioration du Speech pour des débits de 6 à 24 Kbps ; le Main, un “superset” des précédents. La version 2 a ajouté High Quality Audio qui intègre le CELP et l’AAC (Advanced Audio Coding, débits de 16 à 256 Kbps) ; le Low Delay Audio regroupant CELP, TTS et AAC ; le Natural Audio ne comprenant aucun outil destiné aux sons de synthèse ; le Mobile Audio Networking destiné aux applications de communication recourant à des outils non-MPEG. Pour terminer, il faut également mentionner le très efficace HE-AAC (High Efficiency Advanced Audio Coding).

Les objets de synthèse :

C'est l'une des spécificités originelle de MPEG-4 que d'intégrer les objets de synthèse. Pour la reproduction du corps humain, MPEG-4 définit un modèle standard en 3D et va assurer la description des paramètres pour la gestion et l'animation des visages et des corps. A la base, un modèle “neutre”, (préalablement téléchargé ou déjà inclus dans le terminal) et la possibilité d'une animation (expression, configuration des lèvres pour la parole) réalisée grâce aux différents paramètres transmis dans le flux numérique. Il est aussi possible de personnaliser ce modèle (en modifiant la forme et les textures du visage par exemple...) ou de faire appel à des modèles complets (les modèles ne sont pas normalisés, seuls les paramètres le sont). La même technique peut être appliquée pour la représentation du corps dans sa totalité : 175 paramètres (dont 50 pour les deux mains) répartis en 19 groupes sont alors nécessaires.

Même si cet aspect de MPEG-4 n'est plus (provisoirement) sous les feux de la rampe, des travaux ont néanmoins été poursuivis dans ce sens. Ainsi par exemple l'unité de projet ARTEMIS au sein de l'INT (Institut National des Télécommunications à Evry) qui travaille sur un dispositif destiné aux personnes déficientes auditives et qui permet, à travers l'animation d'un personnage virtuel, la traduction semi-automatique de texte dans le langage des signes.

Alain CHAPTAL
Philippe GASSER

Encadré

MPEG (Moving Picture Experts Group) est un groupe de travail issu des deux instances de normalisation que sont l'ISO et la IEC. Ce groupe a en charge l'établissement de standards internationaux pour la vidéo et l'audio numérique. C'est une œuvre de longue haleine puisque les premiers travaux ont commencé à la fin des années 80, il y a plus de 15 ans.

MPEG-1 avait pour cible les applications et le stockage de vidéos sur les disques CD de la première génération c'est à dire avec un débit ne dépassant pas 1,5 Mb/s¹. La qualité de l'image est souvent comparable à celle obtenue à la lecture d'une cassette VHS. L'image est au format CIF (288 x 352 pixels). Ce standard a été approuvé en novembre 1992.

MPEG-2 : Novembre 1994 a vu l'aboutissement de ce nouveau standard développé pour la diffusion de la télévision numérique sous toutes ses formes (satellite, réseaux câblés, hertzien, DVD...). Les débits pour la télévision standard (format 4/3) et le DVD peuvent atteindre 15 Mb/s. La télévision haute définition qui devait faire l'objet de MPEG-3 a été intégrée dans MPEG-2. MPEG-3 ne verra donc jamais le jour (Ne pas confondre MPEG-3 et MP3 !). Un profil spécifique pour la production a été également créé. Le débit maximum est de 50 Mb/s.

Ces deux premiers standards ont été un véritable succès et ont profondément modifié le paysage audiovisuel. MPEG-2 a été adopté dans le monde entier : des millions de décodeurs satellites ou pour les réseaux câblés, le DVD... utilisent cette norme et sont aujourd'hui en service.

MPEG-4 est le troisième standard développé par le groupe. Les travaux de normalisation ont débuté en 1993 (soit plus d'un an avant que MPEG-2 ne soit définitivement adoptée) mais la première version n'a été finalisée qu'en octobre 1998. Des développements complémentaires ont été ensuite réalisés, d'abord avec la version 2 en décembre 1999... et puis plus récemment avec les évolutions articulées autour de H264 et dont les premières applications commerciales commencent simplement à voir le jour. Les applications visées sont nombreuses : diffusion TV, postproduction, télévision numérique terrestre (TNT), télévision haute définition, multimédia, DVD, téléphonie, visiophonie, streaming, télésurveillance...

D'autres normes sont en chantier, mais elles ne concernent pas le domaine de la compression vidéo. MPEG-7 est orientée sur la gestion des contenus et la recherche des informations. Plus loin dans le futur, MPEG-21 a vocation à traiter de la distribution et de la consommation des contenus multimédias dans leur globalité.

¹débit obtenu avec un lecteur de CD ROM de base. Aujourd'hui, les lecteurs peuvent tourner beaucoup plus vite (X40) et donc fournir des débits beaucoup plus importants.