

## **RESUME**

- GAUTHIER Silvère -

### **Contexte**

MPEG, sigle de Moving Picture Experts Group, est le groupe de travail SC 29/WG 11 du comité technique mixte JTC 1 de l'ISO et de la CEI pour les technologies de l'information. Ce groupe d'experts est chargé du développement de normes internationales pour la compression, la décompression, le traitement et le codage de la vidéo, de l'audio et de leur combinaison, de façon à satisfaire une large gamme d'applications.

MPEG a développé les normes suivantes :

- MPEG-1 : première norme audio et vidéo utilisé plus tard pour les Vidéo CD. Ce format offre une résolution à l'écran de 352 x 240 pixels à 30 images par seconde ou de 352 x 288 à 25 images par seconde avec un débit d'environ 1,5 Mbit/s. Elle comprend le populaire format audio MPEG-1 partie 3 audio couche 3 (MP3).
  - MPEG-2 : norme applicable au codage de l'audio et la vidéo, ainsi que leur transport pour la télévision numérique : télévision numérique par satellite, télévision numérique par câble, télévision numérique terrestre, et (avec quelques restrictions) pour les vidéo-disques DVD ou SVCD. C'est notamment le format utilisé jusqu'à présent pour la TV sur ADSL. Les débits habituels sont de 2 à 6 Mbit/s pour la résolution standard (SD), et de 15 à 20 Mbit/s pour la haute résolution (HD)
  - MPEG-4 : norme applicable aux bas débits (jusqu'à 2 Mbit/s), exclus de la matrice des décodeurs de MPEG-2. Permet, entre autres, de coder des objets vidéo/audio, le contenu 3D et inclut le DRM. La partie 2 de MPEG-4 (Visual) est compatible avec la partie baseline de H.263 et a connu du succès grâce à la mise en application DivX ainsi que dans les téléphones mobiles. La partie 10 appelée MPEG-4 AVC permet des gains d'un facteur 2 à 3 par rapport à MPEG-2 et a déjà été retenue comme le successeur de celui-ci pour la TV haute définition, la TV sur ADSL et la TNT. L'extension de cette partie, appelée Scalable Video Coding (SVC) permet de proposer différents niveaux de qualité à partir d'un même flux codé.
  - MPEG-7 : Norme de description pour la recherche du contenu multimédia.
  - MPEG-21 : Norme proposant une architecture pour l'interopérabilité et l'utilisation simple de tous les contenus multimédia.
  - MPEG-A : tourné vers les applications multimédia. Est en cours de standardisation.
-

## **Bref résumé de la méthode/application présentée**

Le MPEG regroupe actuellement 5 standards regroupés en 2 familles. La première famille, la plus connue, a pour vocation de compresser des programmes audiovisuels et les rendre utilisables par le plus grand nombre. Ce sont le Mpeg-1, le Mpeg-2 et le Mpeg-4.

La seconde famille travaille sur un registre différent. Le mpeg-7 définit des outils capables de décrire de façon standardisée un contenu multimédia afin de rendre rapide et efficace des recherches et des classifications de documents. Le Mpeg-21 vise quant à lui la commercialisation automatisée et contrôlée de tout type de document multimédia.

### **MPEG-1**

#### **MJpeg**

Le JPEG a donné l'exemple sur des images fixes. Suite à la sortie de ce standard, différents codecs MJpeg ont fait leurs apparitions. L'idée était de dire qu'une vidéo n'était qu'une suite d'images fixes et que donc chacune de ces images pouvait être compressée en JPEG.

Afin de gagner un maximum de place le Mpeg-1 ne travaille pas sur une image vidéo pleine cadre à 50 frames entrelacées par seconde. Le Mpeg-1 travaille sur 1/4 de l'écran, ne code qu'une frame sur 2 (l'autre étant répétée) et donc pas d'entrelacement. Ce format s'appelle le SIF et supprime 75% des informations.

#### **Compression vidéo spatiale**

Pour compresser, il faut commencer par supprimer les données redondantes, ensuite intervient une phase destructrice où l'on joue sur les défauts de la vision humaine pour que cette destruction soit la plus discrète possible. Pour travailler sur une image, le Mpeg-1 la divise en blocs de 8x8 pixels. Ensuite on applique une transformation appelée DCT (Transformée Cosinus Discrète) pour avoir une représentation fréquentielle de l'image.

Des études psychophysologiques sur la vision humaine montrent que l'œil est moins sensible aux détails fins d'une image (les hautes fréquences) pour lesquelles une approximation sera difficile à détecter pour l'œil. Suite à ce constat, on détermine quels coefficients il vaut mieux arrondir dans la phase de quantification (phase destructrice).

On obtient donc une matrice 8x8 que l'on va stocker en zig zag parce que l'énergie de l'image est concentrée en haut à gauche de la matrice donc on obtient les valeurs les plus hautes en haut à gauche et les plus basses en bas à droite. Du coup les derniers chiffres de cette suite sont nuls ou quasi nuls.

Il suffit d'arrondir les derniers chiffres à 0 de les remplacer par un End Of Block puisqu'un block fait toujours la même taille.

Pour terminer on utilise un codage entropique (sans perte) : le code de Huffman ou VLC. Ce codage consiste à coder les symboles qui ont une probabilité d'apparition forte avec des mots de longueur plus faible que les symboles qui ont une probabilité d'apparition faible.

#### **Compression temporelle**

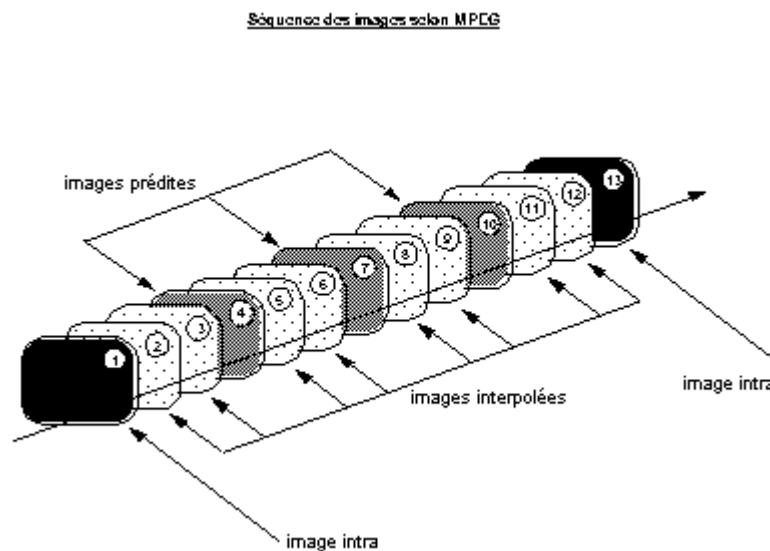
Le but de la compression temporelle est de tirer partie de la redondance temporelle d'une vidéo. Or la plupart du temps cette redondance est forte et il y a donc beaucoup de place à gagner

sans trop perdre en qualité. Le but est donc de ne coder que ce qui change dans le temps.

Pour cela on utilise 3 différents types d'image :

- Image I (Intra) ou Key frame : Image codée individuellement (comme une JPEG).
- Image P (prédite) : Image prédite à partir d'une image I ou P qui se situe avant ou après elle. Elle est en moyenne 3 fois moins volumineuse qu'une image I.
- Image B (bidirectionnelle) : Image interpolée en utilisant les informations d'images I ou P passées ou futures. Elle est en moyenne 6 fois moins volumineuse qu'une image I.

On voit donc l'interdépendance de ces images que l'on regroupe dans un GOP (Group Of Picture). Un GOP est en général composé de 12 à 15 images en MPEG1. On remarquera que le décodeur devra décoder les images dans le désordre (les P puis les B).



*exemple : IBBPBBPBBPBB*

Un GOP peut être ouvert ou fermé. Dans un GOP fermé, la dernière image B n'a pas besoin de l'image I du groupe suivant pour être décodée et le flux de bit peut être coupé à la fin de ce GOP. Plus un GOP est long, plus le débit est faible pour une qualité donnée.

Compression du son : Mpeg1 layer 2 : mp2

Compression du son : Mpeg1 layer 3 : mp3

## Applications

La norme Mpeg-1 définit avec précision la syntaxe du flux de données compressé et les opérations effectuées par le décodeur, mais elle ne standardise pas la technique mise en oeuvre par un codeur. C'est grâce à cela que le Mpeg-1 est une norme permettant de relire une vidéo avec n'importe quel codec Mpeg-1 tout en permettant l'évolution des codeurs et donc une amélioration constante de la qualité de ceux-ci.

## **MPEG-2**

Iso certifié en 1994, le Mpeg-2 est le standard actuel de diffusion de vidéo. TV par câble, par satellite, le DVD... il a permis le passage de l'analogique au numérique au grand public. Le Mpeg-2 est une sorte de généralisation du Mpeg-1 auquel on aurait enlevé la plupart des limitations.

### **Les profils et les niveaux du Mpeg-2**

Le Mpeg-2 définit des profils de décodeurs. Un décodeur Simple profil ne sait pas décoder les B frames par exemple. Il utilise donc moins de ressource (moins cher ou plus rapide) .

Le profil le plus utilisé est le Main profil. C'est celui des DVD par exemple. Un autre profil très utilisé est le 4:2:2 qui comme son nom l'indique accepte du 4:2:2 en entrée et non plus du 4:2:0. Il est utilisé en général en studio, travaille sur des GOP beaucoup plus court (1 ou 2 images).

Le Mpeg-2 permet aussi la haute définition. Le Mpeg-3 devait concerner la HDTV mais a été abandonné puisque le Mpeg-2 le faisait.

Compression du son : La norme AAC est apparue avec le Mpeg-2 et a été complétée dans la norme Mpeg-4.

## **MPEG-4**

### **Le codage orienté objet**

Le Mpeg-4 a pour but de stocker et de diffuser des éléments tels que de la vidéo, de l'audio, de la 2D ou de la 3D. Le but est d'avoir un ensemble composé d'objets pouvant être un décor 2D de forme variable avec une animation 3D, une voix, une musique de fond, etc.

On obtient une liste d'objets codés individuellement et organisés de façon hiérarchique. Un script décrit leurs interactions. Les objets peuvent donc être transmis indépendamment au décodeur ou ils seront ré-assemblés. L'utilisateur lui peut inter-agir avec la scène reçue en supprimant ou en déplaçant certains objets par exemple, voir même cliquer sur un objet pour obtenir des informations ou lui appliquer des actions spécifiques.

Il est possible également de stocker les objets récurrents en local afin de ne pas avoir à les retransmettre. Dans les objets dit synthétiques (calculés), on retrouve les techniques du jeu vidéo, avec le maillage par triangles ou l'animation d'un visage...

Autre avantage du Mpeg-4 face au Mpeg-2 : le principe d'échelonnabilité. Ce principe permet à partir d'un seul encodage la diffusion à des débits différents. L'idée est de décomposer le flux de données en différentes couches qui se complètent. La couche basse est reçue par tous les décodeurs et plusieurs couches supplémentaires ajoutent chacune des informations pouvant améliorer la résolution de l'image, sa qualité, etc.

Le but est ici de n'avoir qu'une vidéo qui pourra aussi bien être diffusée en bas débit vers des téléphones portables (qui n'auront que la couche basse) et au complet via satellite vers des téléviseurs haute définition.

## Les profils du mpeg4

Les profils du Mpeg-4 sont beaucoup plus nombreux et complexes que ceux du Mpeg-2. En effet chaque type d'objet se voit attribuer des outils de compression optimisés pour lui.

Un profil qui ne fera que confirmer que le Mpeg-4 n'est pas réservé au bas débit, bien au contraire, c'est le Studio profil. La qualité sera bien meilleure que celle du Mpeg-2 avec une partie réservée au montage et l'autre à la distribution par exemple de films pour les cinéma.

D'autres profils sont très spécifiques. Par exemple la gestion de l'animation d'un visage de synthèse permettant d'envoyer l'objet image avec les méthodes de déformation du visage, des lèvres... pour n'avoir ensuite que du texte à diffuser pour voir ce visage le lire.

### Le Mpeg-4 simple profile

Le simple profil ne traite que les vidéos rectangulaires. Il ressemble donc beaucoup au Mpeg-2. Ce sont les codecs que nous connaissons actuellement (Divx, Xvid, etc.). Mais évidemment il y a eu une progression depuis le Mpeg-2 : la précision d'un mouvement est de 1/2 pixel voire 1/4 pour le Qpel, gestion globale des mouvements panoramiques, le GOP a grossi, etc.

Ceci permet donc de gagner encore et toujours des kb/s de débit pour avoir la même qualité.

### Le H264 ou Mpeg-4 part 10

La norme H264 est une reprise du simple profil mais encore et toujours plus efficace. Le but est d'améliorer de 50% l'efficacité de compression pour les débits inférieurs à 1Mb/s et de 15 à 30% au dessus. Donc avoir en 500kb/s ce que l'on avait avant en 1Mb/s !

Autre but : améliorer la robustesse du format en cas de diffusion avec risque de défaillance réseau (paquets perdus ou retardés). L'idée est de généraliser ce profil pour des réseau type UMTS.

Pour la robustesse, tout est dans l'organisation des données.

Application de deux principes :

- Ne jamais diffuser ce qui ne se voit pas ou ne s'entend pas
- Supprimer toute redondance

Cette fois ci c'est la redondance temporelle qui est particulièrement visée:

- L'estimation de mouvement s'effectue sur des blocs de pixels plus petits et donne des résultats plus précis sur des petits objets.
- L'estimation de mouvement peut se faire à partir de plusieurs images de références
- La notion de mouvements répétitifs est prise en compte (ex : une roue qui tourne...)
- La DCT a été améliorée
- Le codage entropique a été amélioré surtout dans les bas débits

Mais tous ceci a un prix : multiplication par 3 à 5 de la complexité du codeur et du décodeur.