

La compression des images animées

Introduction	2
1. Principes de bases	3
1.1. La vidéo	3
1.2. La compression vidéo	4
1.3. Les données vidéo	6
2. La compression MPEG	7
2.1. Le codage spatial	7
2.2. Le codage prédictif	7
2.3. La compensation de mouvement	8
2.4. Le codage bidirectionnel	10
2.5. La partie système	12
3. Normes et produits	13
3.1. MPEG-1	13
3.2. MPEG-2	13
3.3. MPEG-4	14
3.4. MPEG-7	15
3.5. MPEG-21	15
3.6. Formats propriétaires	16
Conclusion	17
Références	18

Introduction

L'apparition du CD audio au début des années 80 est à l'origine de ce que l'on appelle communément aujourd'hui le "multimédia". Sur une petite galette de 12 cm de diamètre, on pouvait stocker d'un coup près de 650 Mo de données alors que les disques durs de l'époque atteignaient péniblement des capacités de 500 Mo. Très vite, on a pris conscience que le CD pouvait servir non seulement à stocker du son, mais aussi toutes sortes de données informatiques et notamment des images et de la vidéo. Pour cette dernière, il a fallu tout de même attendre quelques années pour que la puissance des micro-ordinateurs permette de la gérer avec suffisamment d'aisance.

Pour la vidéo, le problème de taille est qu'il s'agit d'une application très gourmande en débit. En effet, transmettre une vidéo au format XVGA (1024 x 768) nécessite un débit de l'ordre de 60 Mo/s ! D'où la volonté d'établir des normes de compression de données vidéo. En 1988 fut fondé le comité **MPEG** (officiellement pour *Moving Picture Experts Group*, mais on trouve aussi dans la littérature la traduction *Motion Picture Experts Group*) dans le but de créer une norme permettant de stocker de la vidéo compressée sur CD. Cette norme, connue sous le nom de MPEG-1, voit le jour en 1992. D'autres normes ont suivi (MPEG-2, MPEG-4, MPEG-7 et maintenant MPEG-21), avec à chaque fois des améliorations et/ou spécifications.

Dans cet exposé, nous nous attacherons à décrire le mode de fonctionnement de ces normes MPEG. Nous considérons comme acquis les principes fondamentaux de compression des données, et plus particulièrement compression d'images fixes et de sons. Nous ne développerons pas la partie audio des normes MPEG pour nous concentrer uniquement sur les parties vidéo et, dans une moindre mesure, système. Enfin, nous laisserons volontairement de côté les formats de compression vidéo propriétaires.

1. Principes de base

1.1. La vidéo

Avant toute chose, il est utile de bien définir de quoi nous allons parler ici : la vidéo, c'est-à-dire une succession d'images animées. Le principe fondamental de la vidéo est que l'oeil humain a la faculté de retenir pendant un certain temps (de l'ordre du dixième de seconde) toute image imprimée sur la rétine. Il suffit donc de faire défiler un nombre suffisant d'images par seconde, pour que l'oeil ne se rende pas compte qu'il s'agit d'images distinctes.

Il existe deux grandes familles de systèmes vidéo : les systèmes vidéo analogiques et les systèmes vidéo numériques. Commençons par les **systèmes vidéo analogiques**, et revenons à la bonne vieille télévision noir et blanc. La caméra balaye l'image bidimensionnelle qu'elle a devant elle par un faisceau d'électrons qui se déplace très rapidement de gauche à droite et plus lentement de haut en bas, et produit un voltage fonction du temps. Elle enregistre ainsi l'intensité lumineuse, et à la fin du balayage on a ce qu'on appelle une **trame**, et le faisceau revient à l'origine pour recommencer. Le récepteur va recevoir cette intensité fonction du temps, et pour reconstruire l'image, va répéter le processus de balayage.

Les paramètres précis de ce balayage varient d'un pays à l'autre, mais deux grandes familles existent : en Europe (système PAL/SECAM, pour *Phase Alternating Line / SEquentiel Couleur Avec Mémoire*) le système utilisé a 625 lignes (seulement 576 sont affichées), un rapport vertical/horizontal de 4/3 et 25 trames par seconde ; en Amérique et au Japon (système NTSC, pour *National Television Standards Committee*), on a seulement 525 lignes (483 affichées) mais 30 trames par seconde.

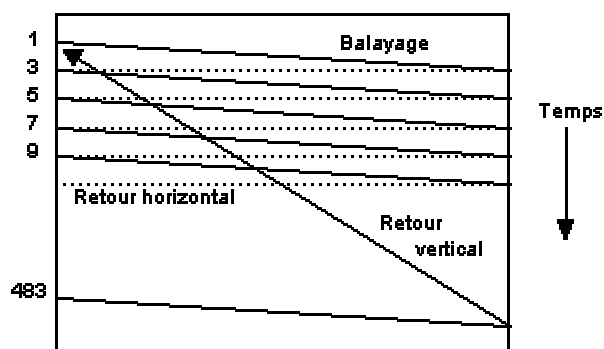


Figure 1 - Principe de balayage utilisé pour la vidéo et la télévision NTSC

Mais l'on s'est aperçu que même avec 25 trames par seconde, certaines personnes pouvaient percevoir des papillotements. Pour résoudre ce problème, au lieu d'augmenter le débit des trames, ce qui exigerait plus de bande passante, on a préféré afficher d'abord les lignes impaires puis les lignes paires. Chacune de ces deux demi-trames est appelée **champ**. Les expérimentations ont montré que là où des gens notaient des papillotement à 25 trames par seconde, ils n'en notaient plus à 50 champs par seconde. Cette technique est appelée **l'entrelacement** (et on parle alors d'images entrelacées). On dit de la télévision ou de la vidéo non entrelacée qu'elle est **progressive**.

La télévision couleur utilise ce même principe de balayage, à ceci près qu'au lieu de balayer l'image avec un seul faisceau, on en utilise trois, un par couleur primaire : rouge, vert et bleu (RVB, ou RGB en anglais pour *Red-Green-Blue*). Les signaux RVB sont ensuite combinés linéairement en un signal de **luminance** et deux signaux de **chrominance**. La télévision haute définition (TVHD) utilise le même principe, mais double grosso-modo le nombre de lignes, pour obtenir une bien meilleure qualité. En outre, elle utilise un format 16/9 au lieu de 4/3, et ceci pour mieux s'adapter au format des films de cinéma.

Passons maintenant aux **systèmes vidéo numériques**. La vidéo numérique est en fait tout simplement une suite de trames formées d'une matrice rectangulaire de pixels. Pour obtenir des images en couleur, il faut utiliser au moins 8 bits par pixels, ce qui correspond à 256 couleurs. En fait, avec 8 bits par pixels, on obtient de la vidéo numérique noir et blanc de haute qualité. Pour la vidéo numérique couleur, on utilise 8 bits pour chaque couleur RVB, soit donc 24 bits par pixel, ce qui correspond à environ 16,8 millions de couleurs. Le principe de balayage utilisé est sensiblement le même que pour la vidéo analogique.

L'entrelacement n'est généralement pas utilisé en vidéo numérique. En effet, les moniteurs rafraîchissent l'écran 75 fois par seconde ou plus, ce qui oblige à afficher trois fois de suite la même trame pour une vidéo à 25 trames/s, et ce qui empêche les papillotements. Enfin, un petit calcul simple pour se persuader de l'utilité de la compression des données vidéo : imaginons que l'on veuille afficher une vidéo numérique à 25 trames/s sur un écran XGA (1024 x 768) en 16,8 millions de couleurs (24 bits par pixels). Cela nécessite un débit minimum de 60 Mo/s...

1.2. La compression vidéo

Toutes les normes de compression vidéo répondent à quelques principes de base que nous allons énumérer ici. Tout d'abord, le signal vidéo lorsqu'il est analogique est converti en paquets de données numériques qui seront plus faciles à transporter sur un réseau. Un des atouts majeurs du format numérique est notamment la non dégradation du signal.

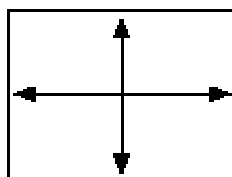
Le principe fondamental de la compression vidéo est de réduire autant que possible les redondances d'informations dans les données, sans bien sûr que cela ne soit visible de manière flagrante pour l'oeil humain. Toute la difficulté est là, dans le dosage entre un taux de compression qui s'améliore en même temps que la qualité d'image devient médiocre.

Deux grandes méthodes de compression existent : la compression sans perte et la compression avec perte.

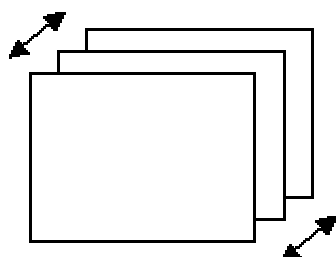
- Dans le cas de la **compression sans perte**, les données décodées à l'arrivée par le récepteur sont strictement identiques aux données codées au départ par l'émetteur. Ce type de compression, permettant au mieux un taux de compression de 2:1, est évidemment insuffisant pour la compression vidéo.
- Dans le cas de la **compression avec perte**, les données sont différentes à la sortie du décodeur par rapport à l'entrée du codeur. C'est ce type de compression qui est utilisé en vidéo, car on peut accepter des pertes d'information qui ne sont pas toujours visibles à l'oeil et qui se traduisent par de nets gains de compression. En compression avec pertes, on peut atteindre des taux de compression allant jusqu'à 300:1.

Dans une séquence vidéo, il existe deux sortes de redondances : la redondance spatiale et la redondance temporelle.

- La **redondance spatiale** est celle dans chaque image prise indépendamment des autres. On peut diminuer cette redondance en codant chaque image séparément en JPEG (*Joint Photographic Experts Group*). Cette approche est parfois utilisée lorsqu'on a besoin de pouvoir accéder de façon aléatoire à chaque image individuellement, comme par exemple lors d'un montage vidéo. On parle alors de MJPEG (*Motion JPEG*).
- Mais on peut aller plus loin en constatant que deux images qui se suivent dans une séquence vidéo sont quasiment identiques : c'est la **redondance temporelle**. Le but est alors de ne stocker que ce qui est modifié lors du passage d'une image à une autre. Les images ainsi compressées sont de deux types : les images I (images intracodées) et les images P (images prédictives). Les **images I** sont des images complètes codées en JPEG, on les appelle aussi images-clés. Les **images P**, par contre, ne contiennent que les pixels modifiés par rapport à l'image précédente, qui peut être elle-même une image I ou une image P. Nous reviendrons plus tard sur la manière précise de coder les images P. Des techniques particulières, comme la **compensation de mouvement**, permettent d'optimiser la génération et la compression des images P.



La compression spatiale a pour but de diminuer la redondance de données dans une seule image



La compression temporelle a pour but de diminuer la redondance de données entre plusieurs images

Figure 2 - Compression spatiale et compression temporelle

Enfin, il ne faut pas oublier qu'une séquence vidéo ne se compose pas seulement d'images, mais aussi de son. Il y a donc deux flux à gérer : le flux vidéo et le flux sonore. Les normes de compression vidéo comme MPEG ont donc trois parties : une partie vidéo, une partie audio, et une partie système qui gère l'intégration des deux premières. Les codeurs/décodeurs vidéo et audio travaillant indépendamment l'un de l'autre, la partie système est là pour résoudre le problème de multiplexage et démultiplexage.

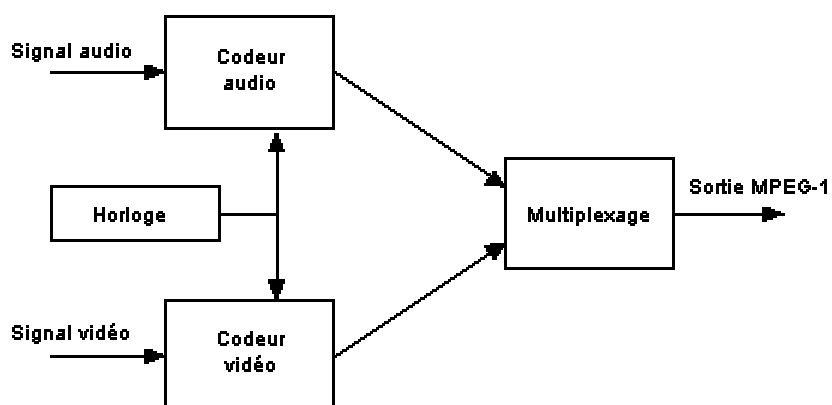


Figure 3 - Synchronisation des flux audio et vidéo dans MPEG-1 à l'émission

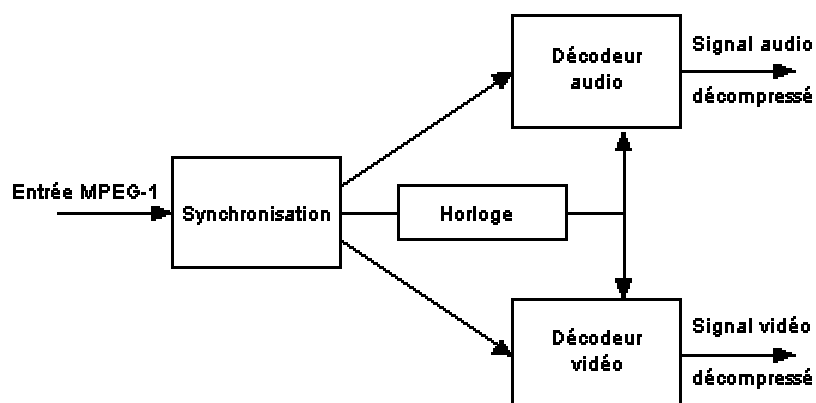


Figure 4 - Synchronisation des flux audio et vidéo dans MPEG-1 à la réception

1.3. Les données vidéo

Dans le flux vidéo, les données sont hiérarchisées selon une manière bien précise. Tout d'abord, la séquence vidéo commence par un code de début de séquence, contient un ou plusieurs groupes d'images, et se termine par un code de fin de séquence. Chaque groupe d'images débute également par un en-tête, et comporte une ou plusieurs images.

Nous arrivons ensuite à l'image qui est l'unité élémentaire pour le codage d'une séquence vidéo. Il existe plusieurs façons de représenter une image dont les deux principales sont la représentation par matrices RVB et la représentation par matrices de luminance et chrominances (notée YUV, où Y est la luminance, U est la chrominance blanc-rouge et V la chrominance blanc-bleu). Ces deux notations sont équivalentes, et on utilisera ici la deuxième, plus économique en place mémoire.

Sans trop entrer dans les détails, une image est donc constituée de trois matrices où chaque élément de la matrice représente un pixel. A noter que les matrices U et V sont de dimension deux fois plus petites que la matrice Y (l'essentiel de l'information de l'image est stockée dans la matrice Y, car l'oeil est plus sensible à la luminance). Plusieurs formats d'images existent et on les note selon les caractéristiques de leurs matrices YUV. Ainsi, le format 4:2:2 indique que sur chaque pixel est échantillonné en luminance (Y), tandis qu'un pixel sur deux est échantillonné pour la différence blanc-rouge (U) et blanc-bleu (V). Le format 4:2:2 est le plus répandu en vidéo professionnelle.

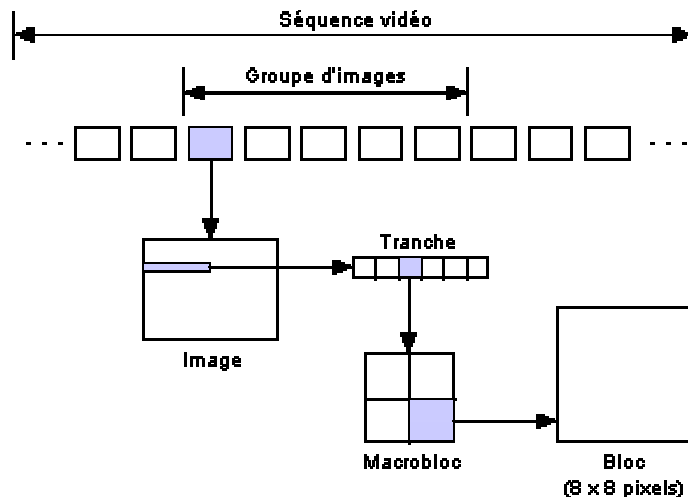


Figure 5 - Hiérarchie des données dans le flux vidéo

Chaque image est découpée en tranche. Une tranche est constituée de un ou plusieurs macroblocs adjacents ordonnés de gauche à droite. Ce sont des éléments importants pour la gestion des erreurs. Si une tranche contient une erreur, le décodeur passe alors immédiatement à la tranche suivante. Bien entendu, plus il y a de tranches dans une image, meilleur est le traitement des erreurs mais plus l'image prend de place également.

Enfin, un macrobloc est une matrice de dimension 2 constituée de blocs, lesquels sont des pavés de 8 x 8 pixels. Un macrobloc couvre donc 16 x 16 pixels dans l'espace de luminance et 8 x 8 pixels dans l'espace de chrominance.

2. La compression MPEG

Nous avons évoqué précédemment le principe de base de la compression vidéo, à savoir la diminution de la redondance spatiale d'une part et de la redondance temporelle d'autre part. Nous allons étudier ici en détail ce principe dans MPEG.

2.1. Le codage spatial

Le flux de sortie de MPEG comporte trois types d'images : les images I (intracodées), les images P (prédictives) et les images B (bidirectionnelle). Étudions tout d'abord les **images I**.

Comme nous l'avons déjà dit, il s'agit d'images complètes codées en JPEG. Sans entrer dans les détails de la compression JPEG, l'algorithme de compression d'une image fixe s'effectue en quatre étapes : transformation en cosinus discrète (DCT, pour *Discrete Cosinus Transform*), quantification, codage en longueur de plage (RLC, pour *Run Length Coding*) et codage de Huffman. Vous pourrez trouver sur le web ou dans la littérature beaucoup d'informations sur la compression d'images fixes.

Typiquement, une image I est intercalée dans le flux vidéo toutes les 10 à 15 images. Avec un débit de 25 à 30 images par seconde, cela veut dire qu'il y a 2 ou 3 images I par seconde dans le flux. Trois raisons principales font que les images I sont indispensables. Tout d'abord, MPEG-1 peut être utilisé pour faire de la diffusion multi-destinataires, chaque utilisateur pouvant se connecter à tout moment. Si toutes les images dépendaient de la précédente, en remontant ainsi jusqu'à la première, et que l'on ait manqué la première, on ne pourrait plus décoder le flux. D'autre part, si une image était erronée, on serait bloqué de la même manière. Enfin, les images I permettent la lecture avant ou retour rapide, sans que le décodeur ait besoin de calculer chaque image survolée.

2.2. Le codage temporel

Les **images P** (images prédictives) sont codées par rapport à l'image précédente. Elles ne codent en fait que la différence bloc par bloc avec l'image précédente. On recherche donc dans l'image précédente un macrobloc identique ou semblable pour optimiser le codage. La différence entre les deux images est ensuite codée spatialement, comme les images I.

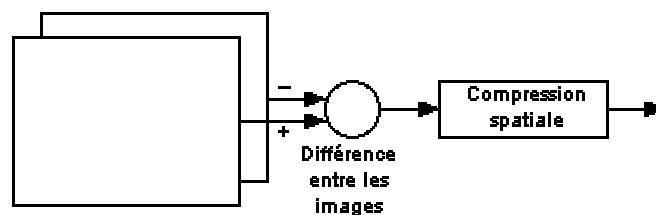


Figure 6 - La compression temporelle : on code la différence entre deux images.
Le résultat est compressé spatialement

La figure 7 illustre un exemple dans lequel les images P sont très utiles. Nous avons trois images consécutives qui ont le même arrière-plan mais différent par la position d'un animal. Les macroblocs qui représentent l'arrière-plan restent les mêmes d'une image à l'autre, mais ceux qui contiennent l'animal devront être décalés d'une certaine valeur qui est à calculer.

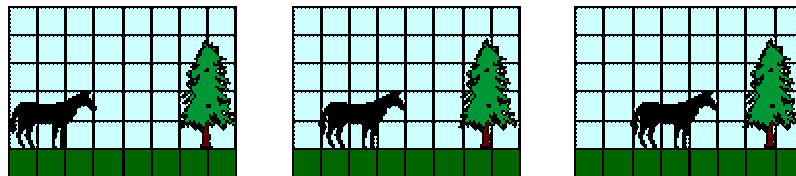


Figure 7 - Trois images consécutives dans une séquence vidéo

Cette technique est appelée la **compensation de mouvement**. Quant un macrobloc est compressé par cette technique, le fichier ainsi compressé contient les informations suivantes :

- un vecteur de déplacement entre le macrobloc de référence et le macrobloc qui va être codé. Ce vecteur rend donc compte du déplacement du macrobloc dans chaque direction par rapport à sa position précédente.
- la différence entre le contenu du macrobloc de référence et le macrobloc qui va être codé. Cette différence est appelée le terme d'erreurs, et est codée en JPEG.

La figure 8 montre le profil type d'une application MPEG basique. Une image I est codée périodiquement, tandis qu'entre chacune d'entre elles, on code des images P qui sont basées sur les images précédentes. Deux problèmes majeurs d'ordre technique se posent immédiatement : si une erreur se produit lors du codage d'une image P, toutes les images P suivantes contiendront également cette erreur ; et si une image I se perd, le décodage des images P suivantes est impossible jusqu'à ce qu'une nouvelle image I arrive.



Figure 8 - Un signal basique MPEG

2.3. La compensation de mouvement

Revenons un peu sur le principe de compensation de mouvement. Ce principe est, on l'a vu, à la base de la génération des images P. La figure 9 montre le fonctionnement d'un système de compression par compensation de mouvement. La partie gauche de la figure représente le codeur, et la partie droite le décodeur. On remarque que le codeur contient en lui-même un décodeur, et ceci pour calculer le terme d'erreurs. En effet, après avoir calculé les vecteurs de déplacement, le codeur les utilise comme le ferait un décodeur pour calculer l'image prédictive uniquement à partir de l'image précédente et de ces vecteurs. Le résultat soustrait à l'image actuelle donne alors le terme d'erreurs. Ce sont donc les vecteurs de déplacement ainsi que le terme d'erreurs qui sont transmis au décodeur. Le décodeur prend donc l'image précédente, lui applique les vecteurs de déplacement puis ajoute le terme d'erreurs, pour avoir finalement l'image actuelle.

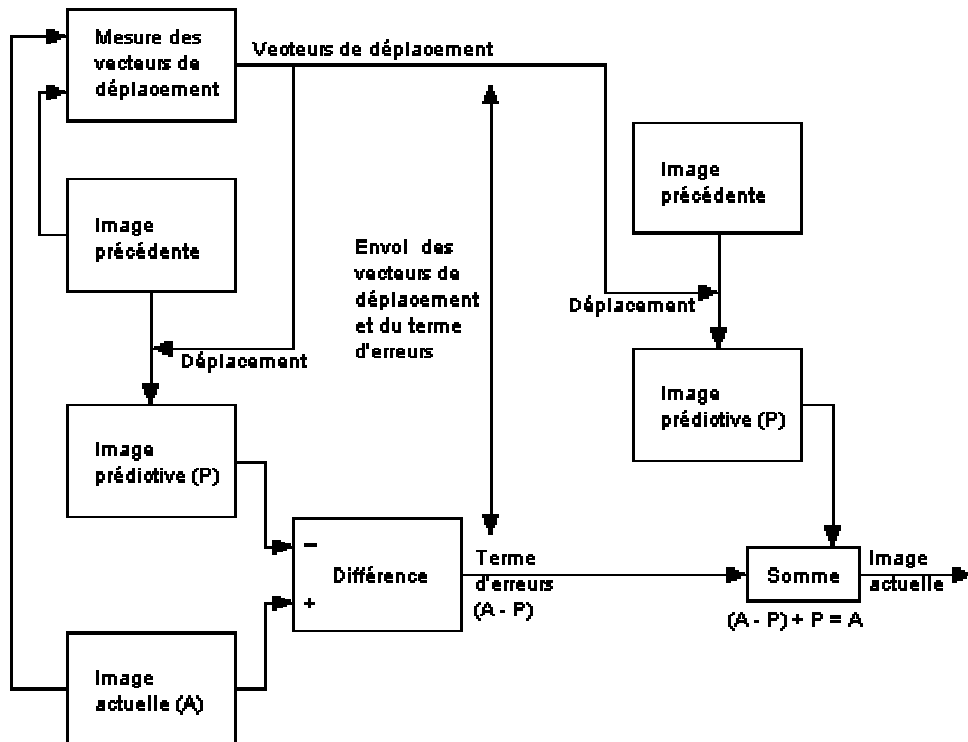


Figure 9 - Un système de compression par compensation de mouvement

Mais comment calcule-t-on précisément les vecteurs de déplacements ? La norme MPEG ne spécifie pas comment faire la recherche, ni jusqu'où la faire, ni même le seuil de ressemblance que l'on doit atteindre. Chaque implémentation est libre de fixer ses paramètres et d'utiliser des algorithmes plus ou moins performants. Trois grandes méthodes existent : la similitude de blocs (*block matching*), la similitude de gradients (*gradient matching*) et la corrélation de phase (*phase correlation*).

La méthode par similitude de blocs est la plus courante et la plus simple à comprendre. Dans une image donnée, un bloc de pixels est choisi et stocké en mémoire. Si le bloc contient un objet en mouvement, un bloc similaire doit exister dans une image suivante, mais pas à la même place. Comme le montre la figure 10, on bouge simplement le bloc de référence autour de sa position initiale dans la seconde image pour trouver les pixels similaires. Quant une similitude est jugée acceptable, le vecteur de déplacement correspond au déplacement entre le bloc de référence et le bloc similaire.

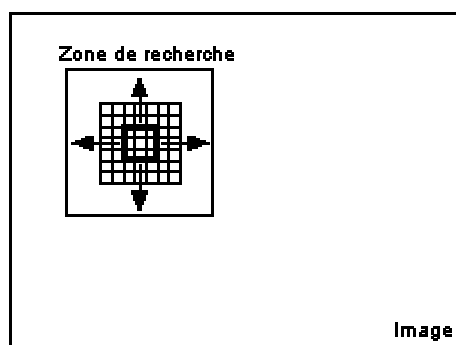


Figure 10 - Similitude de blocs : on bouge le bloc de référence à la recherche d'un bloc similaire

Aussi simple que cette méthode a l'air, elle n'en est pas moins a priori très gourmande en calculs, à cause du grand nombre de possibilités à tester avant de tomber sur le bon bloc. Cependant, en partant du principe qu'un objet en mouvement ne change pas de direction à chaque image, on peut réduire ces calculs en testant d'abord les blocs obtenus par les vecteurs de déplacements précédents, et en continuant éventuellement la recherche à partir de là.

Les méthodes par similitude de gradients et par corrélation de phase sont autrement plus complexes et ne seront pas développées ici.

2.4. Le codage bidirectionnel

La technique de compensation de mouvement avec les images P est très puissante, mais elle ne résout pas tous les problèmes de la compression vidéo. En effet, dans les films, il y a beaucoup de situations où cette technique échoue lamentablement : considérons par exemple une scène où un personnage ouvre une porte, il n'y a aucun moyen de prédire les détails de la pièce qui se trouve derrière la porte avec l'image précédente où la porte était fermée !

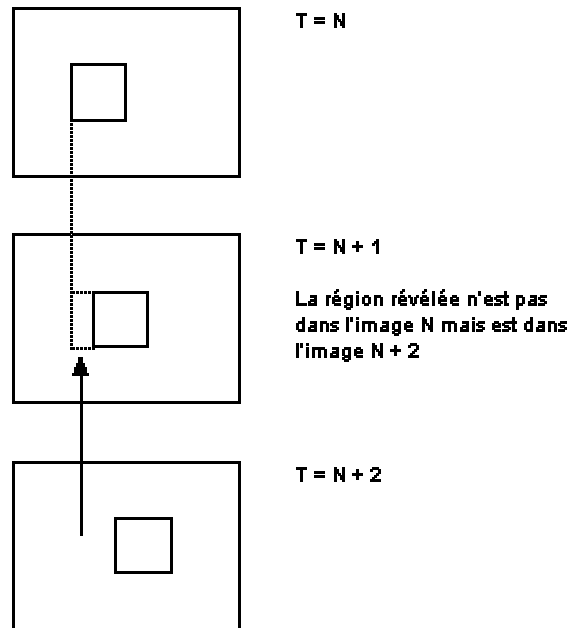
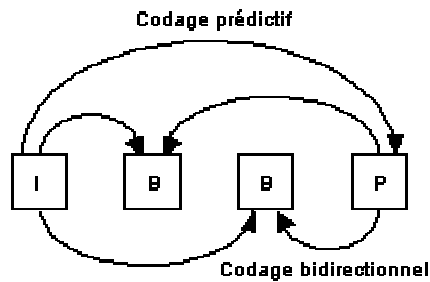


Figure 11 - Codage bidirectionnel : il peut être plus efficace de chercher les données dans l'image suivante que dans l'image précédente

D'où l'utilité des **images B** (images bidirectionnelles) : elles ressemblent aux images P, mis à part le fait qu'elles permettent au bloc de référence de se trouver soit dans l'image précédente soit dans l'image suivante. Cela peut être utile de se référer à l'image suivante, comme l'illustre la figure 11.

Les images B peuvent se référer soit à une image I soit à une image P, mais jamais à une autre image B. Elles ne propagent donc pas, a priori, leurs propres erreurs. La figure 12 montre le fonctionnement du codage bidirectionnel.



I = Codage spatial. Image clé
P = Codage prédictif.
Le codeur envoie la différence entre I et P
B = Codage bidirectionnel.
L'image peut être codée à partir d'une
image I ou P précédente, ou à partir d'une
image I ou P suivante.

Figure 12 - Codage bidirectionnel : on peut insérer des images B entre des images I ou P

Le codage bidirectionnel est très puissant. Par exemple, une séquence IBBP consomme 60 % de bits en moins qu'une séquence IIII, pour une qualité équivalente. Cependant, l'encodage et le décodage d'une telle séquence sont plus coûteux en temps de traitement.

L'algorithme MPEG n'impose pas la fréquence et la place des images-clés (images I). C'est au codage que l'on choisit la fréquence des images-clés. Par exemple, pour une application où l'accès aléatoire est important, une image-clé peut être positionnée toutes les demi-secondes. Le nombre d'images prédites (images P et B) est aussi au choix de l'encodeur. Il dépend en fait souvent de la puissance du décodeur (mémoire), ainsi que des caractéristiques du film (rapidité des plans).

A cause des images B, MPEG est obligé de changer l'ordre des images dans le flux vidéo. En effet, le décodeur serait dans l'incapacité de décoder une image B sans avoir encore reçu l'image de référence si celle-ci se situe plus loin dans la séquence vidéo. La figure 13 illustre ce principe. Ainsi, chaque image de référence est déjà dans le décodeur pour permettre que le décodage bidirectionnel commence. On peut aussi noter que l'image I du prochain groupe d'images est envoyée avant les dernières images B du groupe d'images courant.

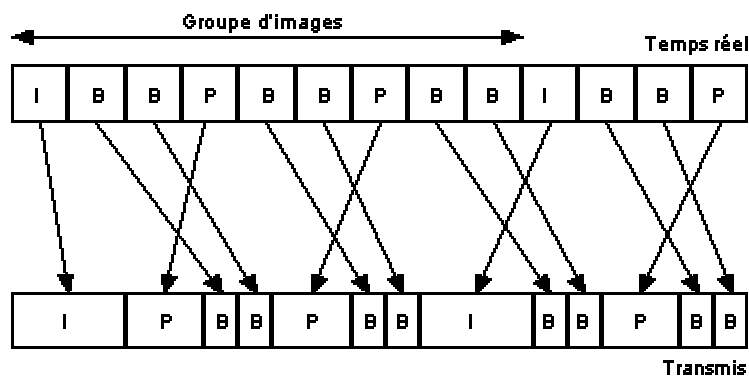


Figure 13 - Composition du flux vidéo

La figure 13 montre également que la quantité de données requise par chaque image est différente. Les images I, qui sont uniquement compressées spatialement, requièrent beaucoup de données. Les images P, créées par déplacement par rapport aux images I, nécessitent moins de données, tandis que les images B sont celles qui sont le moins gourmandes en bits.

2.5. La partie système

Il reste un problème que nous n'avons pas encore étudié. En effet, même si cet exposé se concentre principalement sur la partie vidéo de la compression MPEG, n'oublions pas que les films comportent la plupart du temps du son. Nous n'allons pas étudier la compression audio, sur laquelle il y aurait beaucoup à dire (MPEG est notamment à l'origine du désormais célèbre format MP3). Vous trouverez sur le web ou dans la littérature beaucoup de documentations sur la compression audio. Par contre, il nous faut parler de la partie système de MPEG, qui gère la synchronisation des deux flux vidéo et audio.

La norme MPEG définit donc un mécanisme qui inclut deux paramètres : l'horloge de référence du système (SCR, pour *System Clock Reference*) et le marquage temporel (PTS, pour *Presentation Time Stamps*). La valeur de ces paramètres et leur utilisation est quelque peu différente d'une norme MPEG à l'autre, mais le principe général reste le même. Par exemple, dans MPEG-1, on utilise une horloge système cadencée à 90 kHz qui donne l'heure courante aux deux codeurs vidéo et audio. Ces valeurs sont sur 33 bits, ce qui permet d'avoir des films qui durent 24 heures sans rebouclage de l'horloge. Ces marqueurs temporels sont inclus dans les paquets de données codées et le destinataire les utilise pour synchroniser les flux vidéo et audio à la décompression (voir figures 3 et 4).

3. Normes et produits

3.1. MPEG-1

La norme MPEG-1 est enregistrée à l'ISO (*International Organisation for Standardisation*) sous le code ISO/IEC 11172. Elle a été finalisée en 1992. Elle contient cinq parties : une partie système, une partie vidéo, une partie audio, une partie tests de conformité et une partie simulation logicielle.

Le but des chercheurs qui ont travaillé sur cette première norme était de stocker et de reproduire de la vidéo avec le son associé sur des supports de stockage, en qualité magnétoscope (320 x 240), avec un débit maximum égal à 1,5 Mbits/s. Concrètement, MPEG-1 est devenu par la suite la norme de stockage de vidéos sur CD-ROM au format CD-I ou CD-vidéo.

Le principe de fonctionnement de MPEG-1 est exactement celui décrit dans la partie 2 de cet exposé, avec les restrictions de formes suivantes : affichage maximum de 30 trames/s, images en mode progressif uniquement, format de couleur 4:2:0 au mieux.

Lorsque la norme est sortie, la compression des séquences vidéo en MPEG-1 nécessitait des quantités de mémoire (16 Mo de RAM minimum) que les ordinateurs personnels de l'époque n'avaient pas encore (4 Mo de RAM en moyenne). Aussi la compression vidéo MPEG-1 nécessitait de coûteux investissements (3 000 à 4 500 euros pour une carte de compression, et plus de 20 000 euros pour une station de compression complète !). Aujourd'hui, n'importe quel micro-ordinateur d'entrée de gamme peut compresser sans aucun problème des séquences vidéo en MPEG-1, via des logiciels souvent gratuits (*freeware*).

Enfin, nous pouvons dire un petit mot sur la partie audio du MPEG-1, décomposée en MPEG-1 Audio Layer I, II et III. Ce dernier format, plus connu sous le nom de MP3, est à l'origine d'une véritable révolution de l'industrie discographique, après en avoir fait trembler les bases à la fin des années 90.

3.2. MPEG-2

La norme MPEG-2 est enregistrée à l'ISO sous le code ISO/IEC 13818, et a été finalisée en 1996. Elle comporte neuf parties : les cinq parties semblables au MPEG-1, plus quatre autres parties composées d'extensions diverses.

MPEG-2 a été développée pour la compression de la vidéo de qualité télévision à un débit de 4 à 6 Mbits/s (ce qui correspond à un canal de télévision normal). Un peu plus tard, lorsque MPEG-3 fut abandonné, MPEG-2 intégra la compression de la télévision à haute définition (TVHD). Aujourd'hui, MPEG-2 est aussi le format utilisé pour stocker les films sur DVD.

Si sur le fond, le principe général de fonctionnement de MPEG-2 est très sensiblement identique à celui de MPEG-1, il existe de nombreuses différences de formes entre les deux normes. Tout d'abord, MPEG-2 permet la compression d'images entrelacées là où MPEG-1 ne traite que les images en mode progressif, et ceci bien évidemment pour servir à la télévision numérique. La prédiction de mouvement ne se base donc plus que sur des trames, mais aussi sur des champs. En MPEG-2, on peut faire afficher jusqu'à 120 champs/s, soit 60 trames/s.

Ensuite, au lieu de n'avoir qu'un seul niveau de résolution, MPEG-2 en permet quatre : basse résolution (352 x 288), résolution normale (720 x 576), haute résolution 1440 (1440 x 1152) et haute résolution (1920 x 1152). MPEG-2 définit aussi plusieurs profils, chaque profil visant un type d'application en particulier. Le profil normal, allié au niveau de résolution normal, est le profil le plus utilisé. Le profil simple ne comporte pas d'images B, pour simplifier codage et décodage. Les autres profils concernent principalement la télévision haute définition (TVHD). Ces profils diffèrent en termes de résolution de chrominance (4:2:0 ou 4:2:2) et de débit (de 4 à 100 Mbits/s).

Enfin, le multiplexage des données vidéo et audio est plus général dans MPEG-2 que dans MPEG-1. On dispose de plusieurs flux élémentaires : vidéo et audio bien sûr, mais aussi des flux de données comme des sous-titres en plusieurs langues. Bien d'autres différences minimes existent entre les deux normes, et il serait trop long de toutes les énumérer ici.

Au point de vue produits, la compression MPEG-2 est à la portée de toutes les bourses : un micro-ordinateur d'entrée de gamme suffira, et vous pourrez trouver des logiciels gratuits (*freeware*) ou partagés (*shareware*) bon marché (aux alentours de 30 euros) qui réalisent des compressions MPEG-2 à des bons niveaux de qualité. Pour des applications professionnelles et de la très haute qualité, comme la création de DVD, comptez plutôt de 100 à 600 euros pour un logiciel professionnel. On peut aussi signaler que bon nombre de laboratoires universitaires proposent des encodeurs et décodeurs MPEG-2 gratuits.

3.3. MPEG-4

La norme MPEG-4 est enregistrée à l'ISO sous le code ISO/IEC 14496, et a été finalisée dans sa version 2 en 1999. La finalisation définitive est prévue pour février 2002. Elle comporte sept parties, toujours sur le même schéma des cinq parties semblables au MPEG-1 et MPEG-2, plus deux autres parties composées d'extensions diverses. D'autres extensions sont encore attendues.

Les objectifs de cette norme sont assez nombreux et touchent de vastes domaines. En fait, MPEG-4 a pour ambition de fusionner trois mondes : l'informatique, les télécommunications et la télévision. Cette norme s'appuie sur les succès futurs de la télévision numérique, des applications graphiques interactives et du multimédia (web), son but étant d'assurer une standardisation technologique à tous les niveaux : production, distribution et diffusion. Ses domaines d'application sont immenses : communication temps réel (visiophone), multimédia mobile, téléconférence, post-production (cinéma et télévision), stockage (DVD) et recherche d'informations basée sur le contenu, etc.

Du coup, le principe de fonctionnement de MPEG-4 commence à devenir bien plus complexe que ceux de MPEG-1 et MPEG-2. Le MPEG-4 appréhende désormais la vidéo selon une méthode orientée objets : une scène devient alors une composition d'objets médias hiérarchisés, chaque objet étant décomposé en différents paramètres. La norme n'en devient que plus interactive (accès aux objets, manipulation, organisation). Enfin, le dernier point fort de MPEG-4 est sa flexibilité et son évolutivité.

La norme MPEG-4 propose donc une approche radicalement différente pour le codage des vidéos, en décomposant chaque scène en plusieurs objets médias hiérarchisés. Ainsi, dans l'arborescence de cette hiérarchie, on trouve des images fixes (arrière-plan), des objets vidéo (objets en mouvement sans arrière-plan) et des objets audio (la voix associée à l'objet en mouvement). La figure 14 illustre le schéma de structure d'une scène MPEG-4. Cette approche objet permet alors de très nombreuses opérations sur une séquence MPEG-4 : ajout, suppression ou déplacement d'un objet, transformation géométrique, changement de point de vue, etc. En bref, l'utilisateur interagit vraiment avec les objets de la séquence vidéo.

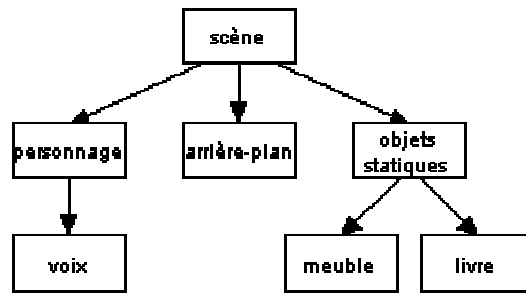


Figure 14 - Structure d'une scène MPEG-4

Cette approche objet nécessite la définition d'objets audiovisuels (AVO, pour *Audio-Video Object*), qui peuvent être soit des composants vidéo, soit des composants audio, soit les deux. Ainsi, une scène audiovisuelle doit être comprise comme la composition d'objets audiovisuels selon un script décrivant leurs relations spatiale et temporelle. Intéressons-nous plus particulièrement aux composants vidéo (VOP, pour *Video Object Plan*). On retrouve alors les mécanismes de compression communs aux normes MPEG, qui sont toutefois généralisés ici aux VOP. Le codage des VOP se fait donc également grâce à une DCT qui peut être adaptée à la forme de l'objet vidéo, suivie toujours de quantification, codage RLE et codage de Huffman. On trouve également des I-VOP (*Intra-VOP*), P-VOP (*Predicted-VOP*) et B-VOP (*Bidirectional-VOP*) qui fonctionnent selon un procédé similaire aux images I, P et B de MPEG-1 et MPEG-2, ainsi que le même principe de compensation de mouvement, le tout toutefois avec des algorithmes bien plus puissants.

MPEG-4 est encore en développement et ne s'est pas encore vraiment imposée au regard de ses ambitions. Concrètement, MPEG-4 apparaît uniquement pour l'instant comme le standard de la visioconférence et du multimédia mobile, dans son profil bas (10 images/s, débit inférieur à 64 Kbits/s). MPEG-4 commence également à se faire un nom auprès du grand public, grâce à son association au codec DivX;-) qui permet de stocker des films sur CD-ROM en qualité proche du DVD.

3.4. MPEG-7

La norme MPEG-7 est enregistrée à l'ISO sous le code ISO/IEC 15938, et est en élaboration depuis 1997. La finalisation d'une première version de la norme MPEG-7 est prévue pour septembre 2002.

MPEG-7 a pour principal objectif de simplifier et d'optimiser la recherche de fichiers multimédia. Pour cela, MPEG-7 va spécifier une description standard de différents types d'informations multimédia. Cette description devra être associée au contenu lui-même pour permettre la recherche rapide et efficace des informations qui intéressent l'utilisateur. Cela peut s'appliquer aussi bien à de la vidéo qu'à du son, des images, des graphiques, des animations 3D, ou encore à la façon dont des objets vidéo sont combinés dans une séquence.

On peut simplifier en disant que MPEG-7 est une "sur-couche" de MPEG-4 dans le sens où MPEG-4 fournit la manière de coder les séquences audiovisuelles tandis que MPEG-7 ajoute une couche d'informations décrivant ces séquences. Cette description, basée sur le langage XML, peut avoir plusieurs niveaux. Dans un premier temps, on trouve les informations relatives à la forme, la couleur, la texture, la taille, la mélodie, l'agencement d'une séquence audiovisuelle. Ensuite, on trouve les informations concernant l'auteur, la date de création et le genre de la séquence, ainsi que des informations plus précises à propos de l'agencement de la séquence et des objets la composant. Enfin, une dernière couche donne les informations relatives à la taille du fichier, sa durée, ses méthodes de codage et la langue utilisée.

Cette norme, une fois finalisée, touchera des domaines d'application aussi vastes que les bases de données multimédia, la télévision et le cinéma, le commerce électronique, le journalisme, les services culturels et éducatifs, etc.

3.5. MPEG-21

La norme MPEG-21 est enregistrée à l'ISO sous le code ISO/IEC 18034, et est en élaboration depuis 2000. Une première finalisation est prévue, au mieux, en 2002.

La dernière née des normes MPEG ajoute notamment à MPEG-7 une couche "sécurité" en proposant tout un éventail de solutions permettant des échanges de matériaux audiovisuels en toute sécurité (déclaration, identification, protection). MPEG-21 redéfinit également la structure interne des séquences audiovisuelles en optimisant les interactions possibles entre les objets audiovisuels, et entre les utilisateurs et ces mêmes objets.

3.6. Formats propriétaires

MPEG n'a pas le monopole de la compression vidéo, loin de là. De nombreux formats propriétaires ont été développés, souvent par de grandes firmes.

On peut notamment citer QuickTime, développé par Apple en 1991 pour son nouveau système d'exploitation "Système 7" pour Macintosh. Il s'agit d'un environnement complet de développement orienté vidéo. On peut donc y réaliser le montage vidéo, en plus de la compression des séquences vidéo. Cette compression s'effectue par le procédé Indeo.

Le procédé Indeo a été développé par Intel, et a été adopté notamment par Microsoft et Apple. En toute logique, Intel a développé un algorithme de compression basé uniquement sur la performance du processeur utilisé. D'autres algorithmes ont été développés (Cinepak, ou les Video Codec de Microsoft) et sont couramment utilisés sur le web pour coder de petits fichiers.

Enfin, l'un des algorithmes privés les plus connus en ce moment est le DivX;-), qui a été développé à partir du codec MPEG-4 de Microsoft par deux pirates anonymes. Ce codec permet d'optimiser la compression MPEG-4 et de stocker un film entier sur CD-ROM en qualité proche du DVD. Comme le MP3 fait trembler l'industrie du disque depuis quelques années, le DivX;-) commence à inquiéter sérieusement les majors de l'industrie du cinéma.

Conclusion

Depuis l'apparition du multimédia dans l'informatique, les méthodes de compression vidéo n'ont cessé de s'améliorer, afin d'optimiser le stockage et la diffusion à travers les réseaux de séquences audiovisuelles. On l'a vu, le comité MPEG travaille en ce sens, et le fruit de ses recherches est aujourd'hui à la base de toutes les technologies multimédia existantes.

On peut désormais obtenir, dans les meilleurs des cas, des taux de compression vidéo proches de 300:1, avec une qualité très raisonnable. De nouvelles méthodes de compression d'images fixes, en utilisant les fractales ou les ondelettes, laissent espérer de bien meilleurs résultats encore dans le domaine de la vidéo (remplacement du JPEG par fractale ou ondelette pour les images-clés).

Enfin, rien ne vaut un bon exemple pour terminer cet exposé et illustrer la puissance de MPEG. Vous pouvez télécharger en cliquant [ici](#) un clip publicitaire bien connu pour des pastilles rafraîchissantes. Ce clip dure exactement 30 secondes et possède 750 images (soit 25 images/s). La résolution est de 352 x 288. Compressé en MPEG-1, ce fichier vidéo a une taille de 2,1 Mo. Non compressé, sa taille est de 222,6 Mo. Soit un taux de compression supérieur à 100:1, pour une qualité visuelle plutôt bonne. CQFD...

Références

Livres

- MPEG-2, John Watkinson, Focal Press, 1999
- Réseaux, Andrew Tanenbaum, Dunod, 1996

Sites web

- www.mpeg.org
- [MPEG home page](#)
- [FAQ MPEG](#)
- [Exposé norme MPEG - EMN Réseaux](#)
- [Video compression](#)
- [Exposé norme MPEG-4](#)