

La compression vidéo

Principes de l'audiovisuel

I. Principe.....	3
II. Le format MJPEG	3
III. LE MPEG1.....	4
1) A l'origine... ..	4
2) Principe du MPEG1.....	4
IV. Le MPEG2	5
1) Pourquoi le MPEG2 ?.....	5
2) L'exemple de la diffusion vidéo.....	6
3) Principe de réduction du débit.....	7
A) Debit constant (CBR).....	7
B) Débit variable (VBR).....	7
4) Un codage spatial et temporel.....	8
5) Prédiction inter images à compensation de mouvement.....	9
6) Les GOPs	11
A) Les images I(I-Frames):	11
B) Les images P(P-Frames):	11
C) Les images B(B-Frames):.....	12
D) Structure de GOP.....	12
7) Profils et niveaux	13
8) Un codeur complet... ..	14
9) Un format populaire	15
V. Le MPEG4 part 10 (AVC)	16
1) Prédiction temporelle ou Inter-prédiction	16
2) Prédiction Spatiale	17
3) Adaptation des transformées discrètes	17
4) Plusieurs codeurs entropiques.....	18
5) Deblocking	18
6) Deux modes d'entrelacement.....	18
7) L'approche Objet.....	19
8) Des résolutions multiples	20
9) Performances générales et cadre d'utilisation.....	20
VI. Les supports vidéos	21

Principes de l'audiovisuel

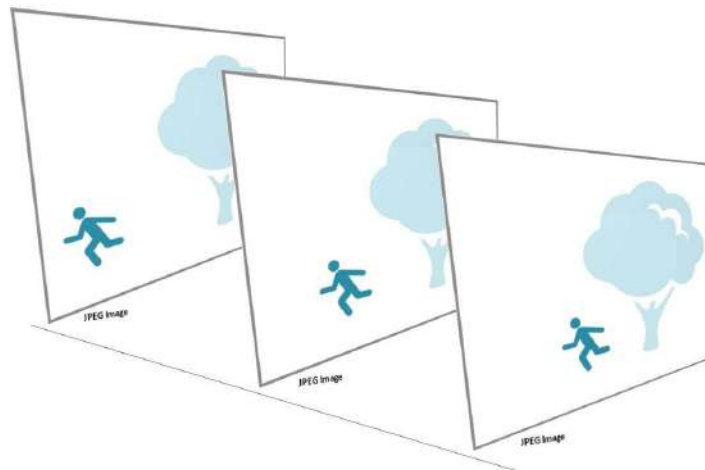
I. Principe

La compression vidéo a pour but de réduire le poids et ou le débit de séquence vidéos animées. Elle repose sur des techniques variées. Dans un premier temps nous allons considérer une vidéo comme une simple suite d'images fixes, puis en tenant compte de la redondance temporelle nous pourrions mettre en œuvre de nouveaux outils pour réduire la quantité d'information.

II. Le format MJPEG

C'est une adaptation du JPEG à l'image vidéo. Chaque image est traitée et compressée indépendamment des autres. On parle alors de compression intra-image.

Les images compressées se suivent alors à 25 ou 30 images par secondes pour rendre le mouvement de la vidéo.



Il ne s'agit pas à proprement parler d'une norme mais plutôt d'une famille de compressions basées sur la norme JPEG.

Plusieurs constructeurs de solutions matériels ont en effet choisi d'adopter ce principe pour coder les vidéos, sans qu'un standard ne soit véritablement défini. Ainsi les tables de quantification peuvent être différentes, tout comme la manière de parcourir la matrice quantifiée.

Ce manque d'uniformisation entraîne des problèmes lors de l'échange de vidéos compressés en MJPEG. Les risques de mauvaise interprétation sont nombreux lors du décodage d'un fichier MJPEG entre différents systèmes.

Principes de l'audiovisuel

Permettant d'obtenir une bonne qualité visuelle, les taux de compression sont assez faibles, impliquant des fichiers encore assez volumineux.

Le format est jugé sans perte jusqu'à 3:1, la perte est quasiment invisible jusqu'à 10:1. Nous verrons par la suite qu'il est possible de faire beaucoup mieux...

Néanmoins ce type de compression est assez largement utilisé lorsqu'il s'agit d'enregistrer des données sur bande. Les formats Digital Betacam, HDCAM et D5-HD se basent sur le principe de la compression MJPEG. Ces formats propriétaires bénéficient en effet des avantages du MJPEG vis à vis de la vidéo non compressée, sans que le décodage ne soit problématique.

Le MJPEG n'est plus développé aujourd'hui mais continue à être très utilisé... C'est la base du codec DV et de ses différentes déclinaisons.

III. LE MPEG1

1) A l'origine...

Le MPEG1 est un format de compression utilisé principalement pour le multimédia. Il permet d'obtenir une qualité comparable à la VHS pour un débit très faible de 1,5 Mb/s (audio+video). Ce débit est comparable à celui d'un signal audio Stéréo non compressé. Il permettait de faire tenir un film sur un CD, c'est le vidéo CD... ancêtre du DVD.

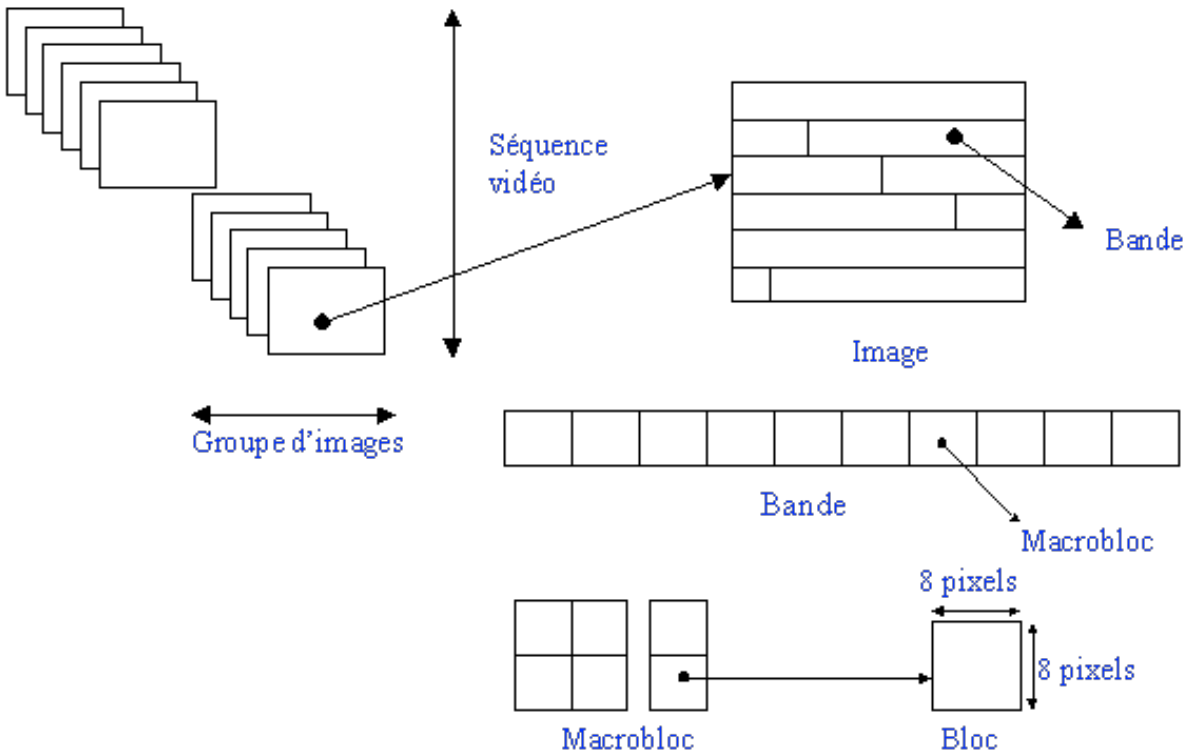
Pour arriver à un débit aussi faible, la taille de l'image est réduite à un quart de la résolution standard soit 360x288 pixels. Le MPEG1 fonctionne en 25 trames progressives, il supprime une trame sur deux et réaffiche deux trames similaires par la suite. Ces restrictions sont définies par la norme SIF (Standard Image Format).

Les images ainsi réduites sont alors codées en tirant parti de la redondance spatiale (intra image) et de la redondance temporelle (interimages).

2) Principe du MPEG1

- L'élément de base est un Bloc de 8x8 pixels.
- Les Blocs sont associés en Macroblock pour tenir compte du sous échantillonnage 4:2:0. Un Macroblock correspond à 6 Blocs (4 Blocs Y + 2 Blocs UV).
- Les Macroblocks forment des tranches (Slice) qui constituent les images
- Les éléments sont traités de la même manière que dans l'algorithme JPEG à savoir DCT et quantification.
- Les images sont assemblées en GOP (Group Of Picture) on tire alors parti de la redondance temporelle en utilisant des vecteurs de déplacement des macroblocks d'une image à l'autre à l'intérieur d'un GOP.

Principes de l'audiovisuel



Le son dans le MPEG est compressé, selon une technique appelée SUB BAND MASKING, en fait, on exploite ici un phénomène physiologique de l'oreille et la réaction des dispositifs de haut parleurs, ou les sons les plus puissants masquent ceux plus faibles.

On élimine ainsi les informations liées aux sons masqués, sans réelle perte de qualité auditive. C'est tout simplement la base du format MP3 (MPEG 1 layer 3) permettant de compresser dans un rapport de 10:1 le poids des fichiers originaux.

Ce format de compression autorise des débits très faibles pour une qualité médiocre. Il est très utile lorsque la qualité n'est pas le facteur le plus important. (échange de fichiers de travail, diffusion par internet, Vidéo CD...).

IV. Le MPEG2

1) Pourquoi le MPEG2 ?

Les défauts du MPEG1 sont nombreux :

- Débit limité à 1,5 Mb/s
- Taille d'image réduite
- Echantillonnage
- Mauvaise gestion des images entrelacées
- ...

Principes de l'audiovisuel

L'explosion des supports numériques dans les années 90 nécessitait de se pencher sur la question. Les experts à l'origine des formats JPEG et MPEG1 ont alors publié une évolution majeure, le MPEG2.

MPEG-2 est une extension du standard MPEG-1. Il est principalement un format "broadcast" à des débits supérieurs. Il propose des outils algorithmiques pour encoder efficacement la vidéo entrelacée, supporte une grande échelle de débits et permet l'encodage audio surround à multiples canaux.

Le format vidéo de postproduction numérique utilise 270 Mbit/s de débit pour coder les images. Sachant qu'un canal satellite accepte autour de 45 Mbit/s (valeur la plus courante), on voit rapidement qu'il va falloir faire suivre un sérieux régime à nos images.

Le standard MPEG-2 est capable d'encoder un signal télévision standard à un débit allant de 3 à 15 Mbits/s (audio + vidéo + données auxiliaires) et un signal télévision haute définition avec un débit de 15 à 30 Mbits/s.

Les décodeurs MPEG-2 sont également capables de décoder les signaux MPEG-1.

2) L'exemple de la diffusion vidéo

Les stations de télévision européennes diffusent à une fréquence image de 25Hz. Chaque image est formée de deux trames "entrelacées", donnant une fréquence trame de 50Hz. La première trame de chaque image contient seulement les lignes impaires de l'image (la première du haut sera la #1). La seconde trame contient quant à elle les lignes paires (environ 20 ms après le premier champ). Le signal télévision nord-américain est lui aussi "entrelacé", mais avec une fréquence image d'un peu moins de 30 Hz (29,97 Hz).

On pourra utiliser des échantillonnages 4:2:2 ou 4:2:0 sur 8 bits. La région active d'une image vidéo numérique standard est de 720 pixels X 576 lignes, à 25 images par secondes. En utilisant 8 bits pour chaque pixels Y, U ou V, calculons les débits sans compression pour ces deux cas de figure :

- En 4:2:2 / 8bits -> $720 \times 576 \times 25 \times 8 + 360 \times 576 \times 25 \times (8+8) = 166 \text{ Mbits/s}$
- En 4:2:0 / 8 bits -> $720 \times 576 \times 25 \times 8 + 360 \times 288 \times 25 \times (8+8) = 124 \text{ Mbits/s}$

MPEG-2 est capable de compresser le débit d'un signal vidéo standard 4:2:0 jusqu'à un taux de 3 Mbits/s soit un taux de compression de l'ordre de 45 environ. Plus le débit est bas, plus la qualité du signal encodé devient compromise.

Pour un signal de diffusion numérique terrestre standard, comme pour le DVD, un débit d'environ 6 Mbits/s est généralement accepté comme un bon compromis entre la qualité d'image et les impératifs de bande passante du transmetteur ou du support d'enregistrement (Câbles, Satellites, etc...).

Principes de l'audiovisuel

3) Principe de réduction du débit

Le système de réduction du débit fonctionne en enlevant l'information redondante du signal avant la transmission, grâce au codeur et en la réinsérant grâce au décodeur. Deux types de redondances ont été ciblés pour y arriver:

La redondance temporelle où un élément commun à plusieurs images consécutives n'est transmis qu'avec la première image. Pour les autres images, on ne transporte que sa position dans l'image. C'est le cas d'un plan fixe, où l'on voit une voiture qui va de gauche à droite. On transmet une fois le décor et la voiture et, pour les images suivantes, on indique seulement la position de la voiture.

L'autre redondance exploitée est la redondance spatiale. Si, dans la même image, il y a trois voitures identiques, elle n'est codée qu'une seule fois. En y ajoutant la position de chacune dans l'image, on peut reconstituer la scène. La valeur de chaque pixel n'est donc pas indépendante, mais reliée à ses voisins, autant à l'intérieur de la même image que par rapport aux images voisines.

Jusqu'à un certain point, la valeur d'un pixel est prédictible en tenant compte de ses voisins.

A) Débit constant (CBR)

Dans ce cas la compression est faite de telle manière que le débit en sortie est constant tout au long du programme. La transmission et la réception se font alors de manière synchronisée.

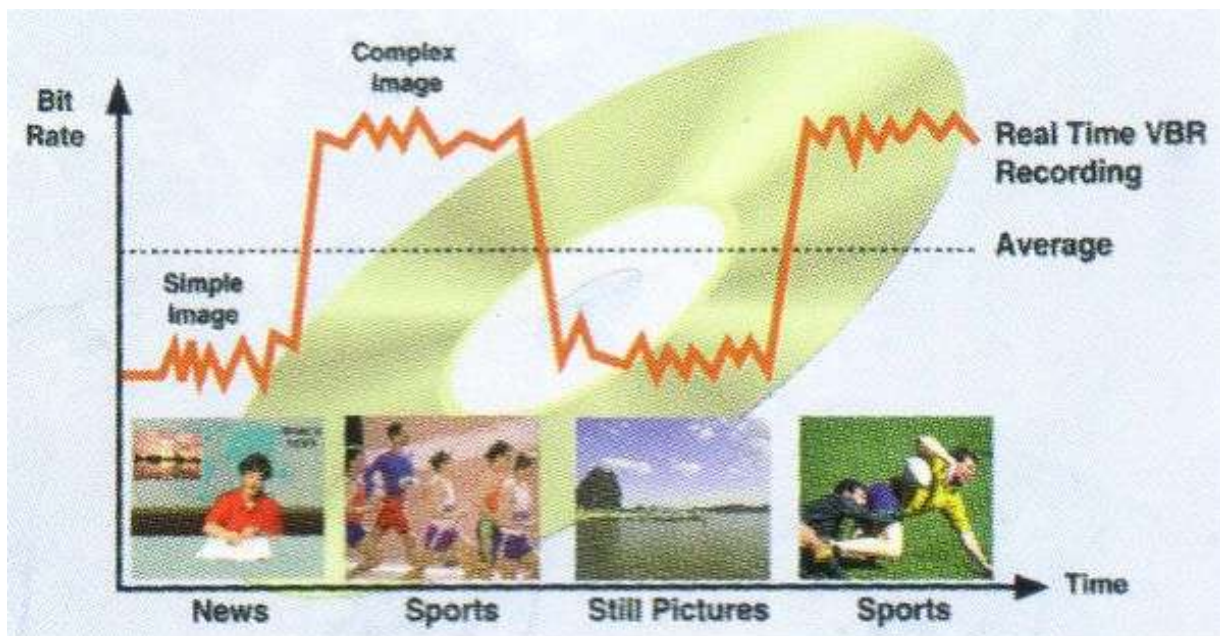
C'est la solution la plus simple, mais pas nécessairement la plus adaptée au contenu de la vidéo. En effet certains passages pourraient se contenter d'un débit plus faible sans que le rendu ne soit altéré (image fixes), alors que d'autres passages nécessiterait un débit plus élevé pour les décrire convenablement. C'est le cas des scènes dont la redondance spatiale ou temporelle est très faible (pluie, explosions...)

B) Débit variable (VBR)

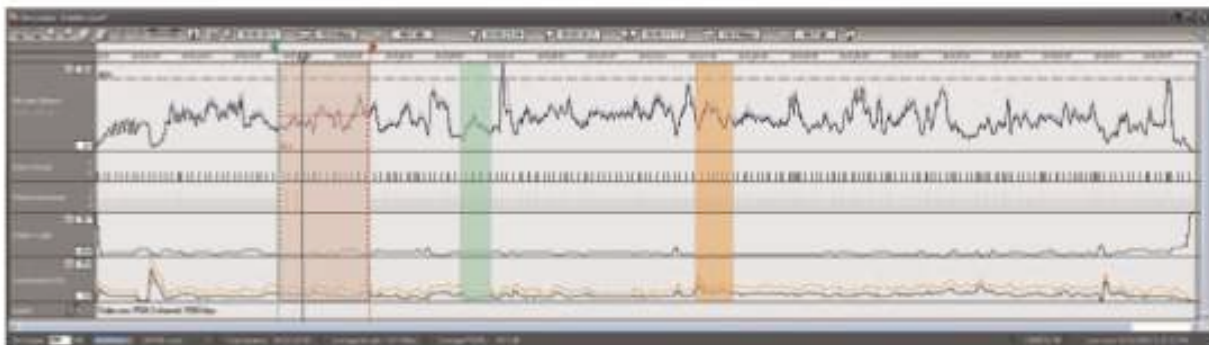
Le débit variable permet d'optimiser la répartition des informations en tenant compte du contenu à encoder. La vidéo est alors analysée puis le débit est automatiquement adapté en fonction des besoins.

On fixe alors un débit moyen et deux limites inférieures et supérieures pour définir les valeurs minimales et maximales du débit.

Principes de l'audiovisuel



Évolution du débit en fonction de l'analyse du contenu de la vidéo.



Des outils d'analyse permettent de contrôler la qualité générale de l'encodage en mesurant le rapport signal/bruit lié à l'encodage.

4) Un codage spatial et temporel

Comme nous l'avons dit, la compression vidéo utilise les avantages des deux redondances (spatiale et temporelle).

La compression spatiale utilise la similarité entre des pixels adjacents sur une surface unie et tient compte des fréquences spatiales dominantes existant dans les zones en amont. Les techniques utilisées sont les mêmes que celles mise en œuvre dans le cas de la compression JPEG. On les combinera avec des techniques liées à l'exploitation de la redondance temporelle.

Principes de l'audiovisuel

En MPEG-2, la redondance temporelle est d'abord réduite en utilisant les similitudes entre deux images successives. La plus grande partie possible de l'image courante est créée (ou prédite) en utilisant l'information de l'image déjà émise. Quand on utilise cette technique, il suffit de transmettre une image de différence qui élimine les différences entre l'image actuelle et l'image de prédiction.

L'image de différence est ensuite soumise à une compression spatiale. Pour des raisons pratiques, il est plus facile d'expliquer la compression spatiale avant d'aborder la compression temporelle.

5) Prédiction inter images à compensation de mouvement

Cette technique exploite la redondance temporelle en essayant de prédire l'image qui doit être codé à partir d'une image de référence.

La prédiction ne peut être basée sur une image "source" car la prédiction doit pouvoir être répétée dans le décodeur, là où les images "sources" n'existent pas. L'image décodée n'est pas identique à l'image source car le procédé de réduction du débit introduit des petites distorsions dans l'image décodée (codage avec perte...)

La prédiction la plus simple du bloc à encoder est celle qui utilise le bloc le plus près de l'image de référence. Ceci permet de réaliser une bonne prédiction pour les régions stationnaires de l'image, mais réussissent moins bien dans les régions où il y a mouvement. Quand un objet se déplace sur l'écran TV, il apparaît à un endroit différent, mais il ne change pas beaucoup d'aspect.

On introduit donc une méthode plus sophistiquée, appelée motion-compensated inter-frame prediction, qui consiste à copier tout mouvement translationnel qui existe entre le bloc qui est encodé et l'image de référence et à utiliser le bloc ainsi créé comme prédiction.

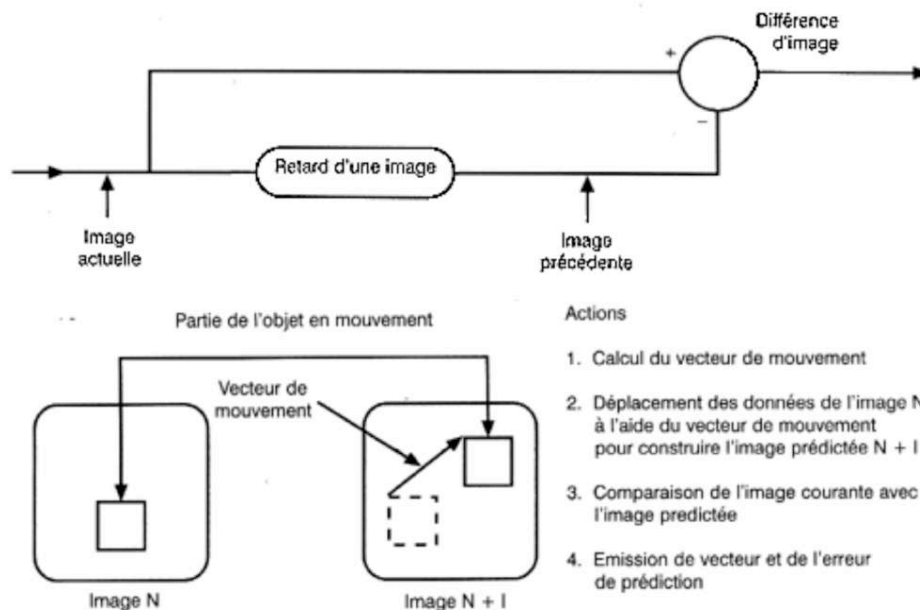
La différence d'image peut être réduite en mesurant le déplacement au codeur. Ce déplacement est transmis au décodeur sous la forme d'un vecteur. Le décodeur utilise ce vecteur pour décaler une partie de l'image précédente vers l'emplacement approprié dans la nouvelle image.

En pratique, chaque vecteur concerne le déplacement d'un macrobloc. La taille d'un macrobloc est déterminée par le codage DCT et la structure de sous-échantillonnage couleur.

Principes de l'audiovisuel



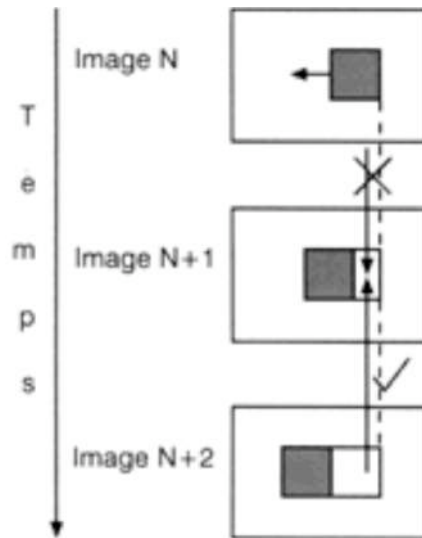
Sur cette image, des vecteurs de déplacement sont associés à chaque macroblocks indiquant les déplacements par rapport à l'image de référence.



Les blocs de prédiction sont créés de plusieurs manières différentes. Par exemple, un bloc pourra être prédit à partir d'une image précédente (forward predicted), d'une image suivante (backward predicted) ou bi-directionnellement (bidirectionally predicted) en faisant une moyenne des prédictions précédentes et suivantes. La méthode utilisée peut changer d'un bloc à l'autre.

Principes de l'audiovisuel

Le codage bidirectionnel réduit considérablement la quantité de données de différence nécessaire à l'amélioration du degré possible de prédiction. Pour chaque bloc à coder, l'encodeur choisit la meilleure méthode pour maximiser la qualité de l'image selon les contraintes du débit. La méthode choisie est ensuite transmise au décodeur pour reconstituer l'image correctement.



6) Les GOPs

En MPEG-2, trois différents types d'images sont nécessaires pour effectuer le codage différentiel et le codage bidirectionnel avec un minimum d'erreurs de propagation:

A) Les images I(I-Frames):

Sont intra-codées et ne nécessitent pas d'informations supplémentaires pour être décodées. Elles nécessitent beaucoup de données comparativement aux autres types d'images et c'est pourquoi elles ne sont transmises que lorsque cela est nécessaire. Elles consistent essentiellement en coefficients de transformées et n'ont pas de vecteur de mouvement. Elles autorisent la commutation de voies et bloquent la propagation des erreurs.

B) Les images P(P-Frames):

Sont celles qui sont déduites d'une image antérieure qui peut être de type I ou P. Les données d'une image P sont constituées de vecteurs décrivant où chaque macrobloc doit être pris dans l'image précédente et des coefficients non transformés décrivant la correction ou les données de différence à ajouter à ce macrobloc. Les images P comportent pratiquement la moitié des données d'une image I.

Principes de l'audiovisuel

C) Les images B(B-Frames):

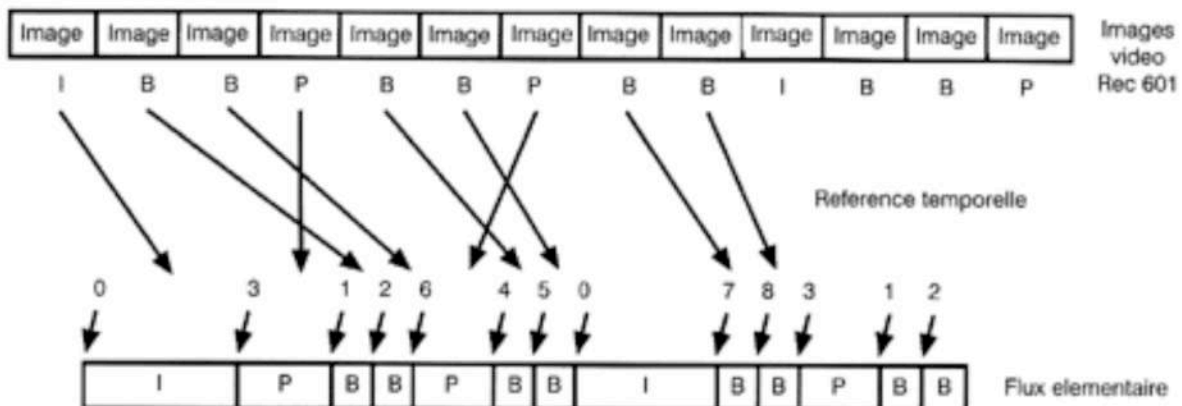
Sont prédites bi-directionnellement à partir d'images antérieures ou postérieures et de type I ou P. Les données des images de type B consistent en vecteurs décrivant l'endroit où les données doivent être prises dans les images antérieures ou postérieures. Elles contiennent également les coefficients de transformées fournissant la correction. La prédiction bidirectionnelle est si efficace que les données de correction sont minimales et que l'image de type B utilise pratiquement le quart des données par rapport à une image de type I.

D) Structure de GOP

À partir de ces images I, B et P, on forme des Groupe d'Images(GOP, Group Of Pictures). Le GOP commence par une image I, suivie de quelques images P espacées et entre lesquelles se placent les images restantes qui sont de type B. La fin du GOP se situe à la dernière image précédant immédiatement une nouvelle image I.



La longueur d'un GOP est variable, mais la valeur la plus courante se situe entre 12 et 15. En fait, si les données d'une image B doivent être utilisées pour construire une image ultérieure, ces données doivent rester disponibles dans le décodeur. Par conséquent, le codage bidirectionnel implique que les données soient extraites de la séquence et provisoirement sauvegardées.

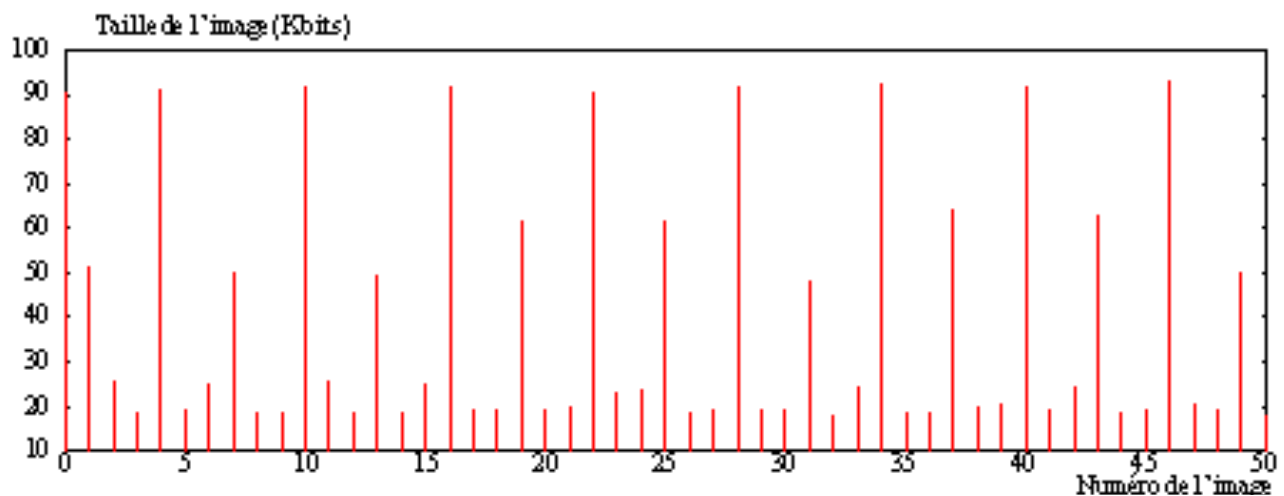


Principes de l'audiovisuel

La figure ci-dessus montre également que les données de l'image P sont émises avant celles de l'image B. Notez également que les dernières images B du GOP ne peuvent être transmises qu'après la première image I du GOP suivant, dans la mesure où elles ont besoin de son contenu pour être décodées bi-directionnellement.

Afin de replacer convenablement les images dans leur ordre, une référence temporelle est incluse dans chaque image. Comme des en-têtes sont régulièrement insérés dans le flux de données, un fichier MPEG-2 peut être affiché dans l'ordre chronologique sur un ordinateur, par exemple.

L'extraction des données d'une séquence nécessite un supplément de mémoire dans le codeur et le décodeur mais aussi génère du retard. Le nombre d'images bidirectionnelles insérées entre des images d'autres types doit être réduit pour diminuer le coût des équipements et limiter le retard si celui-ci doit répondre à des contingences. Un compromis doit être fait entre le facteur de compression et le retard de codage.



Ce schéma représente l'impact du type de codage sur le poids de chaque image. Plus les GOP sont long plus la compression est efficace.

Un problème se pose lorsque l'on souhaite manipuler des GOP pour du montage, il est impossible d'intervenir à l'intérieur d'un GOP directement...à moins de le décompresser.

7) Profils et niveaux

Le format MPEG2 supporte une grande variété de réglages possibles pour permettre de s'adapter au mieux à l'utilisation recherchée (qualité élevée en HD, diffusion en SD, etc...). On définit ainsi des profils et des niveaux de référence.

Principes de l'audiovisuel

Le profil indique les techniques particulières mises en œuvre (algorithme de compression, sous échantillonnage de la chrominance...)

MPEG-2 Profiles

Abbr.	Name	Picture Coding Types	Chroma Format	Aspect Ratios	Scalable modes	Intra DC Precision
SP	Simple profile	I, P	4:2:0	square pixels, 4:3, or 16:9	none	8, 9, 10
MP	Main profile	I, P, B	4:2:0	square pixels, 4:3, or 16:9	none	8, 9, 10
SNR	SNR Scalable profile	I, P, B	4:2:0	square pixels, 4:3, or 16:9	SNR (signal-to-noise ratio) scalable	8, 9, 10
Spatial	Spatially Scalable profile	I, P, B	4:2:0	square pixels, 4:3, or 16:9	SNR- or spatial-scalable	8, 9, 10
HP	High profile	I, P, B	4:2:2 or 4:2:0	square pixels, 4:3, or 16:9	SNR- or spatial-scalable	8, 9, 10, 11
422	4:2:2 profile	I, P, B	4:2:2 or 4:2:0	square pixels, 4:3, or 16:9	none	8, 9, 10, 11
MVP	Multi-view profile	I, P, B	4:2:0	square pixels, 4:3, or 16:9	Temporal	8, 9, 10

Le niveau précise la taille de l'image et le débit maximum notamment, ce qui se traduit par une notion de qualité finale de la vidéo compressée.

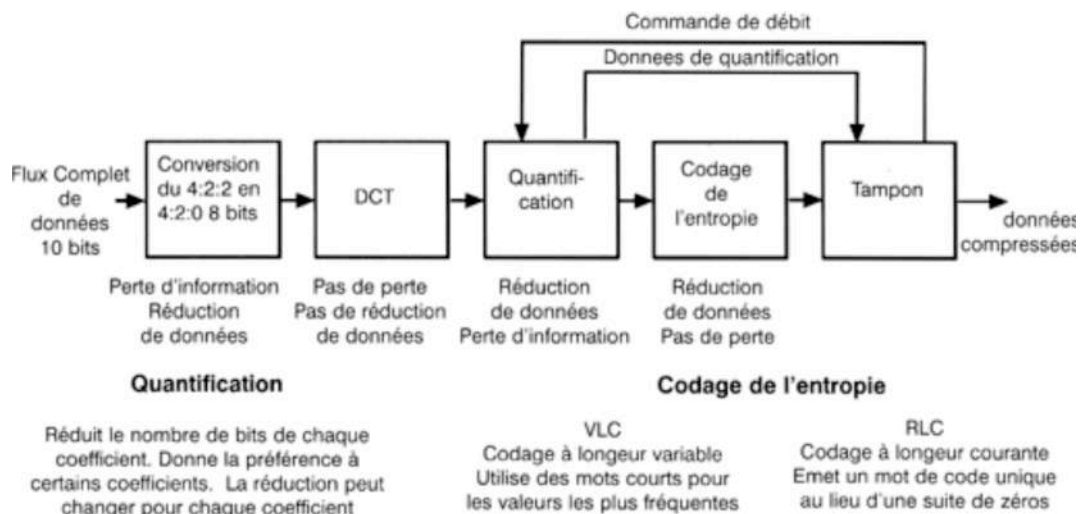
MPEG-2 Levels

Abbr.	Name	Frame rates (Hz)	Max horizontal resolution	Max vertical resolution	Max luminance samples per second (approximately height x width x framerate)	Max bit rate in Main profile (Mbit/s)
LL	Low Level	23.976, 24, 25, 29.97, 30	352	288	3,041,280	4
ML	Main Level	23.976, 24, 25, 29.97, 30	720	576	10,368,000, except in High profile, where constraint is 14,475,600 for 4:2:0 and 11,059,200 for 4:2:2	15
H-14	High 1440	23.976, 24, 25, 29.97, 30, 50, 59.94, 60	1440	1152	47,001,600, except that in High profile with 4:2:0, constraint is 62,668,800	60
HL	High Level	23.976, 24, 25, 29.97, 30, 50, 59.94, 60	1920	1152	62,668,800, except that in High profile with 4:2:0, constraint is 83,558,400	80

On parle par exemple d'une vidéo compressée en mpeg2 en MP@ML pour indiquer que l'on a utilisé le profil **MP** et le niveau **ML**.

8) Un codeur complet...

La figure suivante regroupe l'ensemble des concepts de codage précédemment évoqués.



Principes de l'audiovisuel

Le signal d'entrée est supposé être à la norme 4:2:2 série en 8 ou 10 bits(SDI). Le MPEG n'utilise cependant qu'une résolution de 8 bits; aussi, un étage de traitement sera nécessaire pour arrondir les valeurs si le signal entrant est à 10 bits.

Les profils MPEG effectuent un échantillonnage du type 4:2:0; un étage d'interpolation verticale/filtre passe-bas sera alors nécessaire.

L'arrondi et le sous-échantillonnage couleur introduisent une légère mais irréversible perte d'informations, mais aussi une réduction du débit. Le format d'entrée de la scrutation d'écran doit être tel qu'il puisse être converti en blocs de 8x8 pixels.

L'étage DCT transforme l'information d'image dans le domaine fréquentiel. La DCT n'effectue pas de compression par elle même. Après la DCT, les coefficients sont tronqués et pondérés, ce qui correspond à une première compression. Les coefficients sont ensuite scrutés en zigzag pour accroître la probabilité de commencer par les coefficients les plus significatifs.

Après le dernier coefficient non-nul, un code de fin de bloc(EOB, End of Block) est généré. Les données afférentes aux coefficients sont ensuite compressées à l'aide de codages RLC et VLC.

Dans les systèmes à débit variable, la quantification est fixe mais, dans un système à débit constant, une mémoire tampon est utilisée pour absorber les variations intervenant au cours du codage.

À la suite de la transformée inverse, le bloc de 8x8 pixels est recréé. Pour obtenir un signal de sortie d'écran balayé, les blocs sont stockés dans une RAM qui est lue ligne par ligne.

Pour avoir une sortie 4:2:2 à partir de données 4:2:0, il faudra utiliser une interpolation verticale pour recréer les données manquantes.

9) Un format populaire

Le Mpeg2 est très répandu tout au long de la chaîne de traitement :
BETACAM SX, IMX, XDCAM HDet EX, HDCAM-SR,BLU-RAY, DVD, VCD, Diffusion TNT SD (DVB T), Télévision par ADSL... Tous ces formats utilisent le codec Mpeg2 en tirant profit des nombreux profils disponibles. Cependant, ce format bien que compatible avec la haute définition se trouve moins efficace pour des applications très haute qualité.

On arrive aux limites de ce que le Mpeg2 est capable de faire...

Principes de l'audiovisuel

V. Le MPEG4 part 10 (AVC)

C'est assez logiquement qu'une nouvelle norme a fait son apparition pour améliorer les principes du MPEG2. On est passé directement à la 4^o version, les objectifs visés par la 3^o version n'étant plus adaptés aux évolutions techniques.

La norme MPEG4 regroupe plusieurs parties définissant chacune un contexte précis d'utilisation. Par exemple, la partie 3 est une norme de compression pour le codage perceptuel et les signaux audio ; elle spécifie notamment le format audio AAC.

La partie 10 est une norme avancée de compression vidéo appelée aussi H.264 ou AVC (Advanced Video Codec).

Cette norme a pour but de réduire les débits obtenus par des normes plus anciennes tels que le Mpeg2. Les technologies ont en effet progressé depuis l'époque de la création du mpeg2 et les besoins sont bien plus grand (avènement de la HD, diffusion numérique...).

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

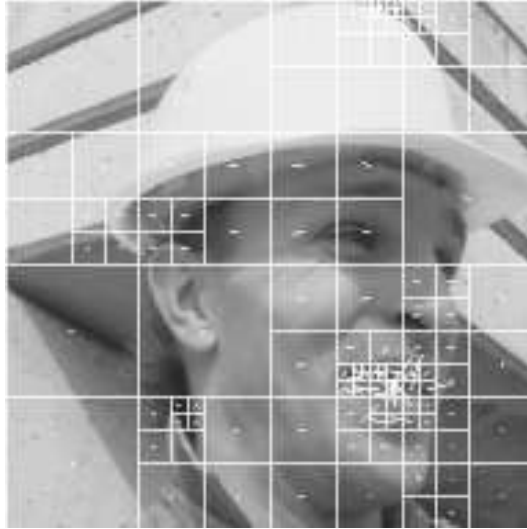
1) Prédiction temporelle ou Inter-prédiction

Une compensation de mouvement pouvant être effectuée par rapport à plusieurs images de référence déjà codées. Le choix de l'image de référence intervient au niveau macrobloc et sous-macrobloc. Ceci permet d'utiliser dans certains cas jusqu'à 32 images de référence (contrairement aux précédentes normes, qui étaient limitées à une ou dans le cas d'images B conventionnelles, à deux) et jusqu'à 4 références différentes pour un même macrobloc.

Cette fonctionnalité particulière permet habituellement des améliorations modestes au niveau du débit et de la qualité dans la plupart des scènes. Mais dans certains types de scènes, comme par exemple les scènes contenant des flashes rapides et répétitifs ou des scènes réapparaissant fréquemment, il permet une réduction du débit réellement significative.

Une compensation de mouvement pouvant utiliser 7 tailles de blocs différentes (16×16, 16×8, 8×16, 8×8, 8×4, 4×8, 4×4) permet une segmentation très précise de zones se déplaçant. La taille variable des blocs permet une gestion plus fine des détails.

Principes de l'audiovisuel



L'image n'est plus découpée en blocs de tailles identiques. On voit que l'image découpée ainsi possède des grandes zones de vecteurs identiques. Sur ces zones, on a un gain important en compression. D'autre part, certaines zones ont été découpées beaucoup plus finement qu'avec des blocs de taille constante. Ici on a un gain en qualité visuelle. Les petites zones mobiles sont beaucoup mieux codées.

La précision va jusqu'au quart de pixel pour la compensation de mouvement, permettant une description très précise du déplacement des zones en mouvement. Pour la chrominance, la précision de la compensation de mouvement se fait même au huitième de pixel.

La compensation de mouvement pondérée par des poids et des décalages permet au codeur de construire des prédictions s'adaptant au changement de luminance et de chrominance de la scène courante. Ceci apporte en particulier un gain pour les scènes comportant des transitions dues à des flashes ou des fondus entre scènes effectués au montage.

2) Prédiction Spatiale

Une prédiction spatiale sur le bord des blocs voisins pour un codage « intra » vient s'ajouter à la prédiction sur les coefficients continus présente dans MPEG-2.

3) Adaptation des transformées discrètes

Une transformée entière est effectuée sur des blocs de taille 4×4 pixels (proche de la DCT classique).

Une autre transformée, celle de Hadamard, est effectuée sur les coefficients moyens de la transformée spatiale primaire (pour la chrominance et éventuellement la luminance dans certains cas) pour obtenir encore plus de compression là où l'image est adoucie.

Principes de l'audiovisuel

4) Plusieurs codeurs entropiques

Des outils de codages entropiques plus puissants peuvent être utilisés compte tenu des évolutions techniques. On pourra utiliser en fonction des besoins :

- Un codage arithmétique, CABAC, qui est une technique sophistiquée de codage entropique qui produit d'excellents résultats en termes de compression mais possède une grande complexité.
- Un codage adaptatif de type Huffman à longueur variable, le CAVLC, qui est une alternative moins complexe que CABAC pour le codage des tables de coefficients de transformation. Bien que moins complexe que CABAC, CAVLC est plus élaboré et plus efficace que les méthodes habituellement utilisées jusqu'à présent pour coder les coefficients.
- Une technique simple et hautement structurée de codage à longueur variable ou pour de nombreux éléments de syntaxe non codés par CABAC ou CAVLC, considéré comme du code exponentiel-Golomb (Exp-Golomb).

5) Deblocking

Un filtrage anti-blocs ou deblocking filter, est effectué dans la boucle de codage et opéré sur les blocs 4×4, permettant de réduire les artefacts caractéristiques du codage avec transformation en bloc (frontière de bloc visible).

6) Deux modes d'entrelacement

Codage en Picture-adaptive frame-field (PAFF ou PicAFF): l'encodeur choisit d'adapter le codage de la trame courante soit de façon progressive (une seule trame), soit par entrelacement (deux champs ou fields).

Codage en Macroblock-adaptive frame-field (MBAFF): même principe que pour le PAFF sauf que l'application ne s'effectue pas au niveau de la trame mais au niveau des macroblocs 16x16. Le sens du codage est légèrement différent puisqu'il traite deux lignes de macroblocs en même temps au lieu d'une seule habituellement.

Principes de l'audiovisuel

7) L'approche Objet

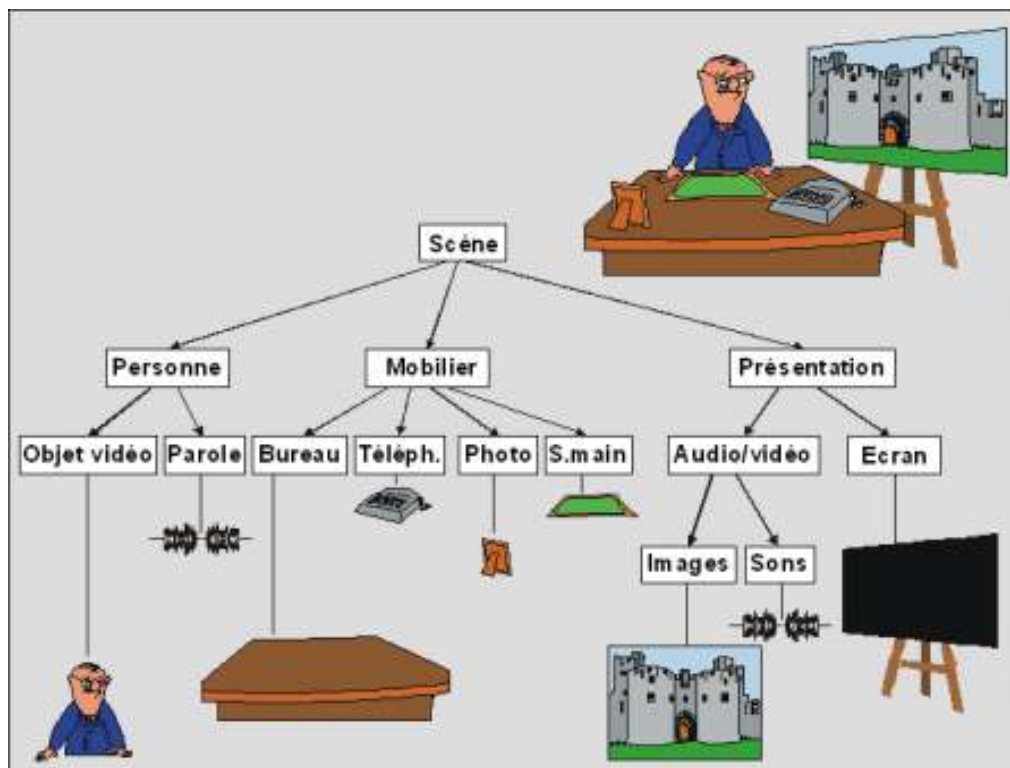
Toutes les techniques décrites précédemment sont des évolutions de celles exploitées par le MPEG2.

Le MPEG4 propose en plus une nouvelle fonctionnalité dans laquelle une scène est décrite comme la combinaison de plusieurs objets dynamiques.

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.



Principes de l'audiovisuel

Pour l'heure cette approche n'est que peut utilisée mais permettra à l'avenir d'offrir une interactivité bien plus grande pour le spectateur de la vidéo en lui donnant la possibilité d'agir concrètement sur le contenu des images (point de vue, affichage/masquage des objets...).

8) Des résolutions multiples

La norme MPEG4 intègre également la possibilité d'accéder à plusieurs résolutions à l'intérieur d'un même fichier. Ce principe permet une meilleure adaptation à la ligne de diffusion en proposant plusieurs niveaux de qualité à partir d'un seul et même fichier source.

9) Performances générales et cadre d'utilisation

Ces techniques, ainsi que plusieurs autres, aident MPEG4 à dépasser significativement les standards précédents, dans une grande variété de circonstances et dans une grande variété d'environnements d'application. MPEG4 peut fonctionner souvent nettement mieux que la vidéo MPEG-2 en obtenant la même qualité avec un débit binaire diminué de moitié, voire plus.

Un même format de fichier permet ainsi de s'adapter à des contenus très légers (diffusion sur internet) jusqu'à des formats beaucoup plus exigeants qualitativement (Blu-ray, HDTV...).

MPEG-4 est une révolution, tant du point de vue de sa structure physique, que de l'idéologie qui en découle. C'est la base de tout ce qui touchera la communication et le multimédia dans les années à venir, et quand on voit l'explosion de toutes ces technologies, on imagine l'importance d'une telle norme.

MPEG-4 n'en est qu'à ses débuts, mais ses applications sont vastes et beaucoup de choses ont déjà été réalisées. Malgré tout, le travail à effectuer est encore long avant d'obtenir des résultats convainquant pour tous les compartiments de la norme. En particulier, il reste deux problèmes majeurs à résoudre pour le codage des vidéos naturelles, à savoir, l'estimation de mouvement et la segmentation de la vidéo en objets.

On retrouvera le codec MPEG-4 à la base des formats HDCAM-SR, AVCHD, TNT HD, télévision HD par ADSL, AVC intra de Panasonic...

Principes de l'audiovisuel

VI. Les supports vidéos

Il existe une grande variété de supports physiques et virtuels permettant d'enregistrer de la vidéo. Ceux-ci font largement appel aux différentes techniques décrites précédemment.

Ce tableau regroupe leurs caractéristiques principales.

Spécifications Techniques								
Format	Constructeur	Taille média	Echantillonnage	Ratio	Débit	codec	media	Transport
HDCAM SR	SONY		4 : 2 : 2 ou 4 : 4 : 4	2,7 : 1 ou 4 : 1	440 - 880 Mbps			HDSDI
DVCpro HD	Panasonic	6.35mm	4:2:2/ 8 bits	6,7:1	100 Mbps	4DV	metal particule	SDTI
HDcam	SONY	½-pouce	17:6:6/ 8-10 bits	4,4:1	144 Mbps		metal particule	HDSDI/SDTI
HDV	SONY		4 : 2 : 0		25 Mbps	IBP GOP 6		IEEE1394
IMX	Sony	½-pouce	4:2:2 / 8 bits	3.3:1	50Mbps/s	MPEG-2		SDTI-CP
D1	Sony/BTS	19mm	4:2:2 / 8 bits	1:1	216 Mbits/s	ITU-R 601	metal	SDI
D2	Sony	19mm	composite digital	1:1	142 Mbits/s		metal	composite digital I/O
D3	Panasonic	½-pouce	composite digital	1:1	142 Mbits/s		metal	composite digital I/O
D5 / D5 HD	Panasonic	½-pouce	4:2:2 / 10 bits	4:1 @ 8 bit 5:1 @ 10 bit 1:1 @ 576/50i	288 Mbits/s		metal	SDI
D6 (voodoo)		3/4pouce	22:11:11 / 10-8 bits	1:1	1 Gbits/s		metal particule	
D7(DVCPR O50)	Panasonic	6.35mm	4:2:2 / 8 bits	3.3:1	50Mbps/s	Double DV	metal particule	SDI
DCT	Ampex	19mm	4:2:2 / 10 bits	2:1	137 Mbits/s	DCT	metal	SDI
Digital Betacam	Sony/Thomson	½-pouce	4:2:2 / 10 bits	2.2:1	128 Mbits/s	DCT	metal	SDI
BetacamSX	Sony	½-pouce	4:2:2@ML	10:1 / Gop : IB	18Mbps/s	MPEG-2		QSDI/SDI
DVCAM	Sony	6.35mm	4:1:1@ML	5:1 / I frames	25Mbps/s	DV	metal	SDI/QSDI/ 1394
DVCPRO	Panasonic	6.35mm	4:1:1@ML	5:1 I frames	25Mbps/s	DV	metal particle	SDI/1394
DVCPRO50	Panasonic	6.35mm	4:2:2@ML	3.3:1 I frames	50Mbps/s	Double -DV	metal particle	SDI
Digital- S(D9)	JVC	½-pouce	4:2:2	3.3:1	50Mbps/s	Double -DV		SDI
miniDV et D8	consortium	6.35mm	4:1:1	5:1 I frames	25Mbps/s	DV	metal	S-Video/ 1394
D-VHS	consortium	½-pouce		variable	14.1Mbps/s	MPEG-2		S-Video/ 1394
D-VHS HD	JVC	½-pouce			19-28Mbps/s	720p/1 080i		S-Video/ 1394
DVD	consortium	12cm	4:2:0 P@ML	variable / longs GOP	6Mbps/s environ	MPEG-2	polycarbonate	S-Video/ 1394
VideoCD 2.0	consortium	12cm	4 : 1 : 1	variable / N:12 M:3	1.2Mbps/s	MPEG-1	polycarbonate	composite
Laserdisc	consortium	12-pouces	Composite			NTSC/ Pal	PMMA	composite

Principes de l'audiovisuel