

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
ECOLE NATIONALE POLYTECHNIQUE



DÉPARTEMENT D'ÉLECTRONIQUE

Mémoire de fin d'études

En vue de l'obtention du diplôme de Master en Electronique

Thème :

Étude des différentes normes de compression vidéo

Réalisé par :

Abd Ennour BOUZENAD

Soutenu devant le jury composé de:

Pr. M.MEHENNI	Président
Pr. R.SADOUN	Examineur
Pr. R.AKSAS	Examineur
Mr. L.ABDELOUEL	promoteur
Mr. M.TAGHI	co-promoteur

Promotion : **juin 2013**

ملخص

يندرج هذا المشروع في إطار معالجة الفيديو. قمنا بشرح المبادئ الأساسية لضغط الفيديو و ركزنا خصوصا على المقاييس المختلفة للنظامين ISO/IEC و UIT-T .

الكلمات المفتاحية : ضغط الفيديو، ISO/IEC ، MPEG ، UIT .

Résumé

Ce projet s'inscrit dans le domaine du traitement de la vidéo. Nous avons expliqué les principes fondamentaux de la compression vidéo et nous nous sommes particulièrement intéressé aux différentes normes de compression vidéo du système ISO/IEC et du UIT-T.

Mots clés : Compression vidéo, ISO/IEC, MPEG, UIT.

Abstract

This project subscribes to the field of video processing. We have explained the basic principles of video compression and we have particularly focused on the different norms of video compression, of ISO/IEC and those of UIT-T.

Keywords: Video compression, ISO/IEC, MPEG, UIT.



Projet de Fin d'Études réalisé au Laboratoire des

Dispositifs de Communications et de Conversions Photovoltaïques

École Nationale Polytechnique

10 avenue Hacène BADI – El Harrach

BP182–16200 Alger

Algérie

Tél : (+213) 21 52 53 01/03

Fax : (+213) 21 52 29 73

Web : <http://www.enp.edu.dz/>

Remerciements

Au terme de ce projet de fin d'études réalisé à l'Ecole Nationale Polytechnique d'Alger, nous tenons à exprimer notre profonde gratitude à notre encadrant **Mr L.ABDELOUEL** qui, grâce à sa disponibilité, aide et rigoureux conseils, nous avons pu mener à bien ce projet.

Nos remerciements s'adressent également à **Mr M.TAGHI** pour son aide et disponibilité, et aux membres du jury qui ont accepté de juger ce travail.

Nous remercions également nos familles qui nous ont soutenus tout au long de nos cursus, nos amis qui nous ont encouragés et à tous ceux qui nous ont aidés, de près ou de loin, à accomplir ce projet.

Abd Ennour BOUZENAD

Contenu

ملخص	2
Résumé	2
Abstract	2
Table des figures.....	6
Introduction Générale	7
Problématique	7
Chapitre 1 : Principe de la compression vidéo	8
Introduction.....	8
1.1 Introduction aux vidéos numériques.....	8
1.1.1 Généralités sur les images numériques.....	8
1.1.2 Compression vidéo :	11
1.1.3 Evaluation de la qualité de l'image	13
1.2 Architecture générique de compression vidéo	14
Chapitre 2 : Normes de compression vidéo	16
2.1 Les normes de l'ISO/IEC.....	16
2.1.1 Norme MPEG-1.....	16
2.1.2 Norme MPEG-2.....	17
2.1.3 Norme MPEG-4.....	17
2.1.4 Norme MPEG-7.....	18
2.2 Les normes de l'UIT-T	19
2.2.1 Norme H.261	19
2.2.2 Norme H.263	20
2.2.3 Norme H.263+	20
2.2.4 H.264 / MPEG-4 AVC	21
2.2.5 La norme H.265	21
Conclusion	23

Table des figures

1.1 : Représentation numérique d'une image	7
1.2 : Représentation des pixels	9
1.3 : Redondance spatiale et temporelle	11
1.4 : Exemple de la redondance temporelle.	12
1.5 : Schéma global d'un encodeur vidéo	13
1.6 : Modules d'un système de compression vidéo.	14
2.1 : évolution des normes de compression vidéo.	15
2.2 : situation du macro-bloc dans un GOP.	18
2.3 : comparaison entre l'espace disque requis.	24
2.4 : performance des normes populaires de compression vidéo en termes de qualité PSNR en fonction du débit.	24

Introduction Générale

Introduction

Avec l'introduction de la télévision numérique haute définition (TVHD), et le développement des communications via Internet, la vidéo numérique est actuellement en plein essor. Les anciennes normes PAL, SECAM et NTSC laissent peu à peu leur place à des normes de codage numérique.

La compression des données joue ici un rôle primordial, que ce soit dans le stockage ou dans la diffusion des vidéos, puisqu'elle doit permettre de stocker plus de données sur moins d'espace, de diffuser plus de chaînes numériques sur une même bande passante disponible, etc.

Problématique

La vidéo numérique a connu une énorme émergence ces deux dernières décennies, afin de maintenir la capacité de transmission et pouvoir stocker ces vidéos de plus en plus gourmandes en mémoire et en bande passante plusieurs normes de compression en vue le jour.

L'objectif de ce master est l'étude des normes d'encodage vidéo et tout particulièrement la norme MPEG4/H.264 qui est la plus populaire ces dernières années.

Chapitre 1 : Principe de la compression vidéo

Introduction

Le but de ce chapitre est d'introduire les notions générales nécessaires à cette étude. Nous définissons dans un premier temps les images numériques et la compression vidéo en général. Le transport d'un flux vidéo numérique nécessite de compresser celui-ci afin de l'adapter au débit du canal de transmission ou à la capacité du support. Des généralités sur la vidéo numérique et les principales techniques de codage sont présentées.

1.1 Introduction aux vidéos numériques

1.1.1 Généralités sur les images numériques

Une image numérique est définie par une matrice de points appelés pixels (figure 1.1). La représentation d'une image numérique est introduite ici afin de bien comprendre les techniques de traitement vidéo.



Fig. 1.1 : Représentation numérique d'une image

Une vidéo numérique non compressée se caractérise par son système de représentation des pixels, sa taille et son rafraîchissement. Plusieurs formats existent selon l'application visée.

Représentation des pixels Les pixels peuvent être représentés de plusieurs manières. Dans une image à niveaux de gris, un pixel possède la plupart du temps 256 niveaux de gris (codés sur 8 bits). Le niveau 0 code le noir et le niveau 255 le blanc. Une image couleur est représentée avec trois composantes. Il existe principalement deux systèmes :

- R : V : B : un pixel de couleur est la somme des trois composantes Rouge Verte et Bleue. Dans le cadre de la vidéo, ce format est principalement utilisé pour l'affichage, il peut toutefois être utilisé en post-production de cinéma.

- Y : Cb : Cr : Un changement de repère est effectué par rapport au RVB. Un pixel est représenté par sa composante de luminance (Y) et deux composantes de chrominance (Cb et Cr). Ce format est aussi appelé YUV.

Le passage aux composantes YCrCb à partir des composantes RVB s'effectue à l'aide des formules suivantes :

$$\begin{pmatrix} Y \\ Cr \\ Cb \end{pmatrix} = \begin{pmatrix} 0.257 & 0.504 & 0.098 \\ 0.439 & 0.368 & 0.071 \\ -0.148 & 0.291 & 0.439 \end{pmatrix} \times \begin{pmatrix} R \\ V \\ B \end{pmatrix} + \begin{pmatrix} 16 \\ 128 \\ 128 \end{pmatrix}$$

Pour reconstruire l'image originale

$$\begin{pmatrix} R \\ V \\ B \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1.140 \\ 1 & -0.395 & -0.580 \\ 1 & 2.032 & 0 \end{pmatrix} \times \begin{pmatrix} Y \\ Cr \\ Cb \end{pmatrix} + \begin{pmatrix} 464.2541 \\ -335.3364 \\ 670.3702 \end{pmatrix}$$

Le système YCrCb est historiquement utilisé dans les systèmes audiovisuels car il permet de diffuser une image couleur tout en gardant une compatibilité avec le système noir et blanc. De plus, l'œil humain est plus sensible à la luminance qu'à la couleur, ce qui permet de réduire la définition des composantes de chrominance et donc d'obtenir facilement une première réduction des données [1]. On distingue trois formats (figure 1.2) :

- 4:4:4 : chaque composante est codée de la même manière, il n'y a pas de sous-échantillonnage. Ce format est utilisé lorsque toute la qualité de la vidéo doit être gardée, c'est à dire pour le stockage ou la post-production en studio, le cinéma numérique.

- 4:2:2 : on ne garde qu'une ligne sur deux pour les composantes de couleur. Seulement la moitié de l'information de couleur est gardée. C'est un format de qualité studio.

- 4:2:0 : on ne garde qu'une ligne sur deux et une colonne sur deux pour les composantes de couleur. Un quart de l'information de couleur est conservé. C'est le format le plus utilisé pour la vidéo grand public.

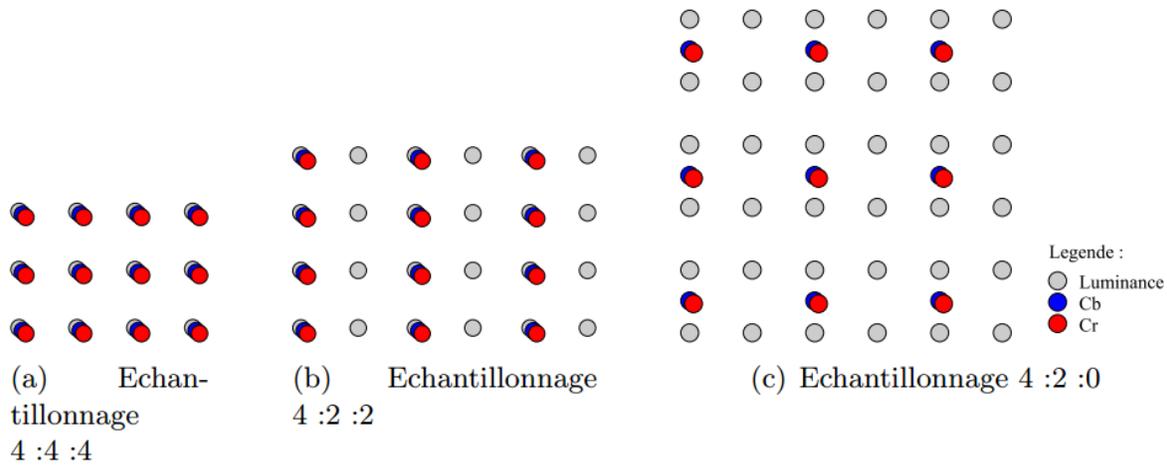


Fig. 1.2 : Représentation des pixels

Chaque composante du signal vidéo peut être codée sur 8 à 10 bits. 8 bits sont suffisants pour représenter une composante puisque l'œil ne distingue pas deux niveaux voisins. Cependant, en post-production 10 bits peuvent être utilisés pour une meilleure précision et une meilleure gestion des différents niveaux de post-production. D'une manière générale, la distribution des niveaux est linéaire, mais il se peut toutefois qu'une échelle logarithmique soit utilisée pour offrir une plus grande dynamique, notamment pour le cinéma numérique qui utilise des caméras pouvant avoir un contraste élevé.

D'une manière générale, les composantes du signal vidéo sont le plus souvent codées sur 8 bits avec un format 4:2:0. On a donc en moyenne 1,5 octet par pixel.

Le balayage Le balayage peut être progressif (« progressive scan » ou p) ou entrelacé (« interlaced scan » ou i). Pour une séquence progressive, toute l'image est affichée d'un seul coup, alors que pour une séquence entrelacée les lignes paires et impaires sont acquises et affichées alternativement.

La taille La taille ou définition d'une image est le nombre de lignes et le nombre de colonnes. Pour des définitions standard dans les applications audiovisuelles, seulement le nombre de lignes est spécifié : (ex : 720p : 1280 colonnes x 720 lignes en balayage progressif).

Le tableau 1.1 récapitule les paramètres des formats standards. Le débit nécessaire à la transmission des données non compressées est calculé pour un format en 4 :2 :0 avec 8 bits par composante (1.5 bits/pixel).

Définition	Balayage	Fréquence	Débit
QCIF : 176x144	progressif	10 à 30 Hz	380 Ko à 1.1 Mo/s
CIF : 352x288	progressif	10 à 30 Hz	1.5 à 4.6 Mo/s
SD (NTSC) : 640x480	entrelacé	60 Hz	13.8 Mo/s
SD (PAL, SECAM) 768x576	entrelacé	50 Hz	16.6 Mo/s
SD (DVD) (D1) 720x576	entrelacé progressif	50 à 60 Hz 25 à 30 Hz	15.6 à 18.6 Mo/s
HD (720p) : 1280x720	progressif	24 à 60 Hz	33.2 à 83 Mo/s
HD (1080i) : 1920x1080	entrelacé	50 à 60 Hz	78 à 93 Mo/s
HD (1080p) : 1920x1080	progressif	24 à 30 Hz	75 à 93 Mo/s

TAB. 1.1 - Formats standard

Comparées aux débits actuels des communications numériques (de quelques centaines de Kbits/s pour le téléphone portable (3G) à quelques dizaines de Mbits/s pour une chaîne satellite), les images non compressées représentent d'énormes quantités de données. Les débits sont ainsi beaucoup trop élevés par rapport aux bandes passantes disponibles pour leur diffusion. Il est donc nécessaire de compresser les données vidéo pour permettre leur diffusion avec des débits réalistes.

Exemple : pour une définition SD (PAL, SECAM) 768x576 une séquence vidéo de durée

$$\begin{aligned}
 1 \text{ minute (sans compression)} &= (24\text{bits}) \times (768 \times 576 \text{ pixels}) \times (25 \text{ image/sec}) \times (60 \text{ secondes}) \\
 &= 15925248000 \text{ bits} \\
 &= 1.85 \text{ Go}
 \end{aligned}$$

Tandis que :

$$1 \text{ minute (avec compression)} = 7.5 \text{ Mbits pour un format MP4}$$

$$\text{Soit un taux de compression de } \frac{1.85 \text{ Go}}{7.5 \text{ Mo}} = 252.5$$

1.1.2 Compression vidéo :

La compression (ou codage) d'une séquence vidéo a pour but de réduire son débit et par conséquent les coûts de stockage ou rendre possible sa diffusion sans surcharger le réseau, deux types de redondances sont éliminés pour aboutir aune bonne compression (figure 1.3) :

– La redondance spatiale : ce type de codage exploite la corrélation spatiale des données au sein d’une même image (indépendamment des autres images de la séquence). Généralement, dans une image, la différence de couleur et d’intensité lumineuse entre des pixels voisins est assez faible. Donc, plutôt que de mémoriser la valeur propre de chacun des pixels d’une image, il est plus efficace de coder que les différences de valeur entre les pixels successifs.

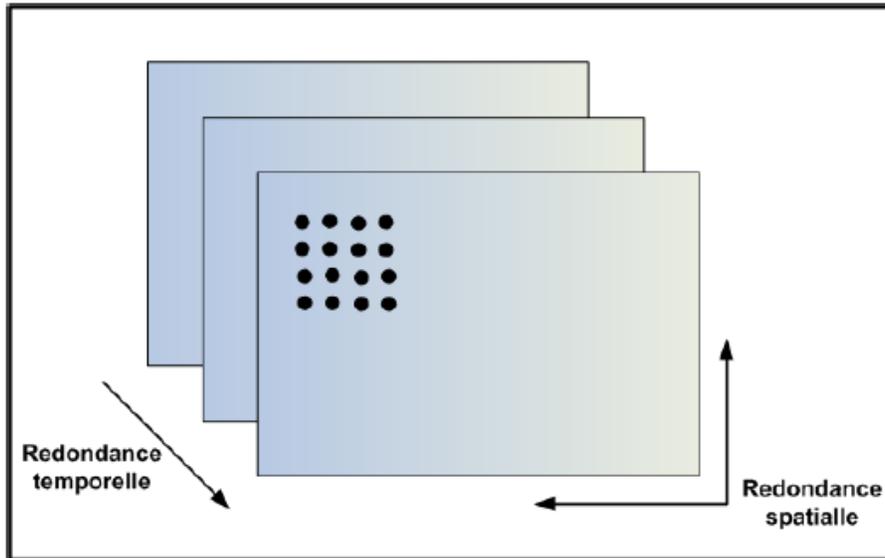


Figure 1.3 : Redondance spatiale et temporelle

– La redondance temporelle : ce type de codage exploite la corrélation temporelle entre des images successives. Elle est utilisée dans le cas où la différence entre deux images successives dans une séquence vidéo est relativement faible (illustrée dans la figure 1.4). Dans ce cas de figure, afin de minimiser le nombre d’informations à mémoriser afin de transmettre la seconde image, on cherche à identifier les déplacements des pixels entre ces deux images successives. Ainsi les données (valeurs de pixels) de la seconde image sont codées par rapport à l’image précédente. Seules les différences entre ces deux images sont alors transmises.

Ce type de codage est appelé codage Inter-frame.



Fig. 1.4 : Exemple de la redondance temporelle.

1.1.3 Evaluation de la qualité de l'image

La compression vidéo réduit de façon plus ou moins la qualité des trames vidéo compressées. L'évaluation de leurs qualités est très complexe. Il existe des métriques de qualité objectives, basées sur un calcul de distorsions comme le PSNR (Peak Signal to Noise Ratio) très utilisé dans le traitement du signal en général :

$$PSNR = 10 \times \text{Log}_{10} \left(\frac{(Amplitude\ Max)^2}{\frac{1}{N} \times \sum_0^{N-1} (Originale - Reconstituée)^2} \right)$$

Le PSNR mesure une erreur quadratique moyenne non pondérée sur toute l'image. Une alternative est proposée par la SSIM (Structural SIMilarity) [2] qui pondère les défauts en fonction de leur voisinage. En effet l'œil humain perçoit moins les défauts dans les zones texturées que dans les zones homogènes. D'une manière générale, ces métriques sont très limitées pour évaluer la façon dont l'œil humain perçoit les défauts dans une image et la façon dont le cerveau les interprète.

1.2 Architecture générique de compression vidéo

Tous les systèmes de compression vidéo actuellement utilisés fonctionnent selon le même principe et possèdent les mêmes étapes décisives. La compression vidéo consiste à réduire leurs tailles en supprimant la redondance ainsi que l'information imperceptible par l'œil humain. Elle exploite les redondances temporelles et spatiales pour extraire le minimum de données nécessaires pour la reproduction des images originales.

Une vue globale d'un système générique de compression vidéo comprend un encodeur et un décodeur comme l'illustre la (figure 1.5).

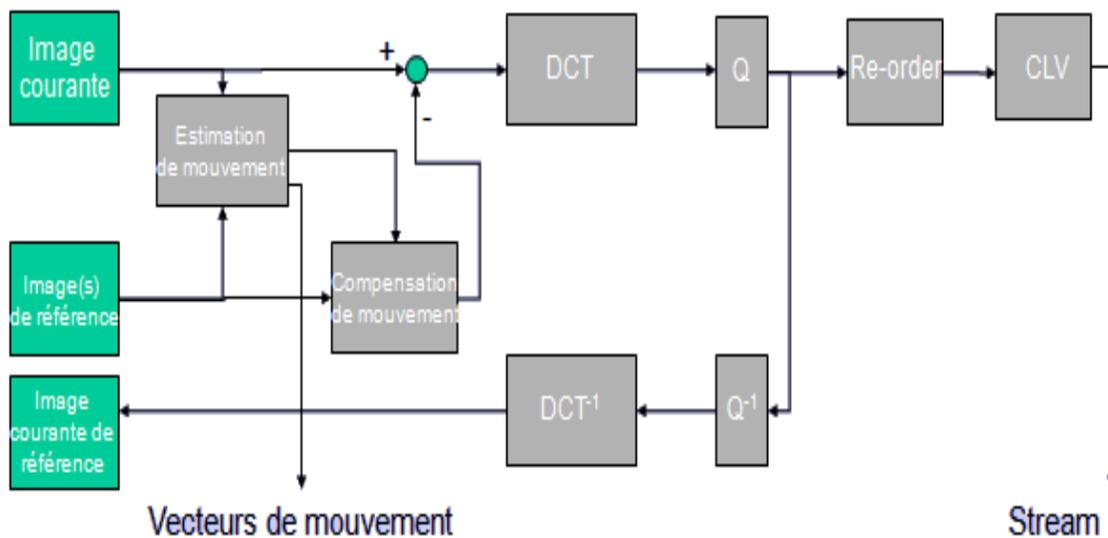


Fig. 1.5 : Schéma global d'un encodeur vidéo

L'encodeur consiste en trois modules : la compression spatiale, la compression temporelle (ou prédiction) et le codage entropique. A l'entrée de l'encodeur, nous fournissons une séquence vidéo paramétrée.

La compression spatiale consiste à éliminer la redondance entre pixels d'une même image et ce en exploitant la corrélation que chacun d'eux présente avec son voisinage. Une transformation

convertit une trame en coefficients, qui sont ensuite quantifiés pour obtenir une série de coefficients significatifs représentant la trame de manière plus compacte.

Vient ensuite le module compression temporelle (ou de prédiction) qui contrairement au module de compression spatiale ce module exploite la redondance entre les trames voisines et construit une prédiction des trames suivantes à partir de trames dites de référence.

Ces compressions convertissent une trame en coefficients. Ces coefficients subissent à leur tour une compression statistique par un codage entropique.

Le décodeur quant à lui reconstruit la vidéo à partir de la séquence de sortie de l'encodeur. Les coefficients et les vecteurs de mouvement sont décodés par un décodeur entropique et puis, le décodeur reconstruit une version de la trame résiduelle. Ensuite, à partir d'une trame précédemment reconstruite, des vecteurs de mouvement et de la trame résiduelle, le décodeur obtient la trame de départ.

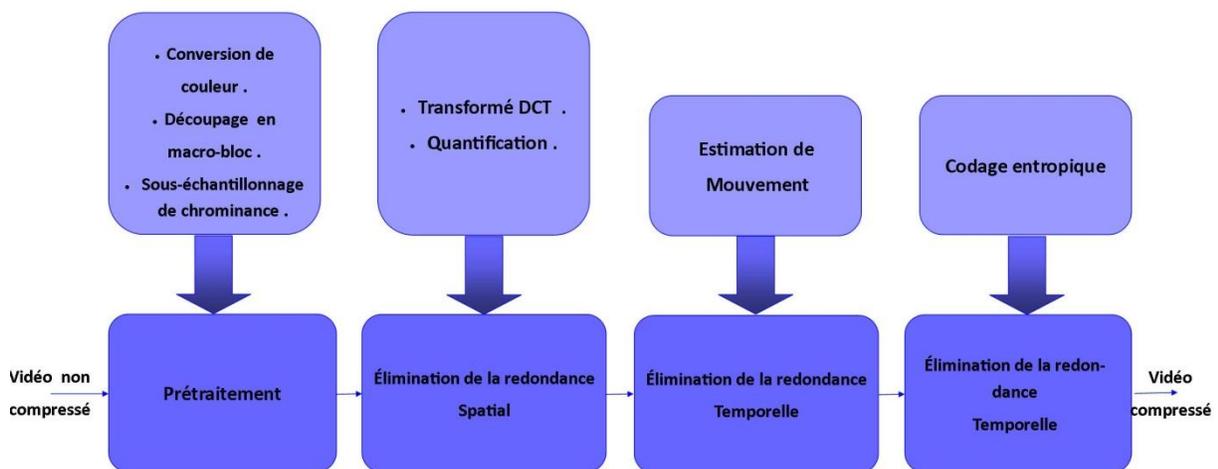


Fig. 1.6 : Modules d'un système de compression vidéo.

Chapitre 2 : Normes de compression vidéo

Durant la dernière décennie, les domaines d'application et les besoins en termes de compression et d'interopérabilité des contenus vidéo ont évolué. La Figure 1.8 synthétise l'évolution des différentes recommandations ou normes de compression vidéo.

Les organismes ITU-T [4] et ISO/MPEG [5] se partagent la normalisation des standards de compression vidéo. Le premier s'intéresse principalement aux applications nécessitant de faibles débits, applications de visiophonie et de visioconférence.

Le deuxième s'intéresse aux applications nécessitant des débits plus élevés : applications d'archivage et de diffusion [6].

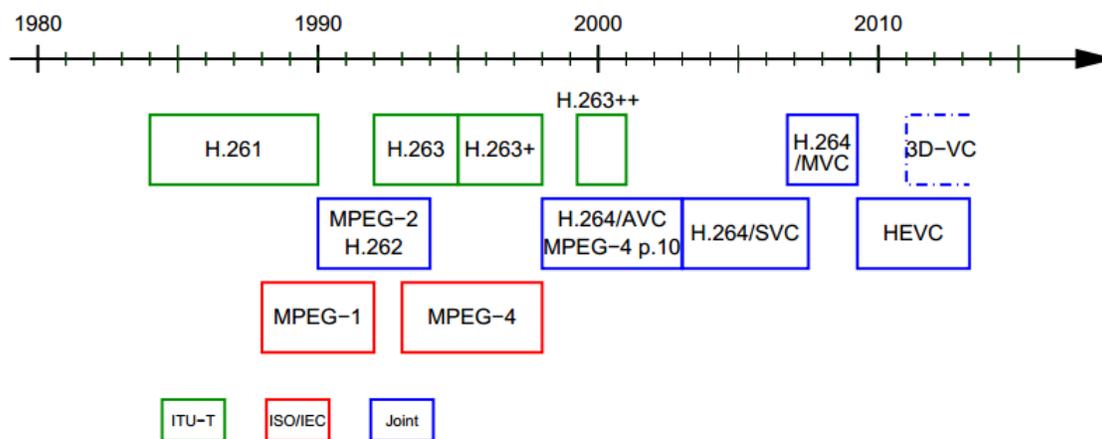


Fig. 2.1 : évolution des normes de compression video.

2.1 Les normes de l'ISO/IEC

2.1.1 Norme MPEG-1

Le premier standard international traitant de la compression vidéo a été publié en novembre 1992 sous le nom de MPEG-1. L'objectif de cette norme était le codage d'images en mouvement et du son associé en vue d'un stockage numérique. Les débits supportés par MPEG-1 peuvent atteindre 1,5 Mb/s (dont 1,2 Mb/s pour l'image animée, 256 kb/s pour l'audio stéréo et quelques kb/s pour les données annexes) [7, 8].

Bien que ce standard soit optimisé pour compresser un flux vidéo de type SIF sous une contrainte de débit à 1,5 Mb/s (352 X 240@30fps progressif ou 352 X 240@25fps en YUV4 : 2 : 0), il est toutefois possible d'adapter le débit et/ou la résolution.

Une séquence vidéo au format MPEG-1 est découpée en groupes d'une quinzaine d'images (GOP, Group Of Pictures). Chaque image d'un GOP peut être encodée de façon Intra (I), Inter (P) ou Bidirectionnelle (B). Dans ce standard, chaque image est découpée en sous-blocs (nommés macro-blocs) afin de réduire la complexité calculatoire des étapes de recherche et de compensation de mouvement. Les macro-blocs sont de taille fixe et se composent de 16 X 16 pixels. Lors des traitements réalisés durant les étapes d'encodage et de décodage (hors estimation de mouvements), ces macro-blocs sont toutefois décomposés en sous blocs de taille 8 X 8 pixels.

2.1.2 Norme MPEG-2

Le standard MPEG-2 [9] a été proposé en novembre 1994. Initialement élaboré pour la transmission de vidéo de qualité TV avec un débit compris entre 4 et 9 Mb/s, ce standard a absorbé la compression des vidéos TV Haute Définition (TVHD) alors dévolues à MPEG-3.

Le standard supporte la compression progressive ou entrelacée de vidéos avec des débits variant de 1,5 Mb/s à 100 Mb/s. Ces spécifications couvrent les besoins de la TVHD, des médias de stockage interactifs (ISM) et de TV par câble.

MPEG-2 génère un flux de données numériques permettant une extraction d'un flux de débit moindre assurant une reconstruction à résolution inférieure. D'une manière générale, il assure une compatibilité descendante et ascendante entre les codeurs et les décodeurs de normes et de résolutions différentes.

Les couches structurant les flux MPEG-2 sont similaires à celles définies par MPEG-1. Les améliorations se situent principalement aux niveaux de la compensation de mouvement, de la gestion des coefficients fréquentiels et de l'adaptabilité du flux vidéo (scalability).

2.1.3 Norme MPEG-4

Le standard MPEG-4 normalisé en 1998 se base sur le concept d'objets média (media objects) pour représenter des contenus audio-visuels. Ces objets peuvent être combinés et synchronisés. Ils interagissent avec l'utilisateur selon des limites fixées par leurs auteurs. Une hiérarchie lie les objets. Le niveau de base est constitué "d'objets médias primitifs" regroupant les images fixes, les vidéos, les sons, le texte, les graphiques, les visages parlants synthétiques et les sons synthétiques. MPEG-4 permet la composition de scènes complexes en standardisant la façon de placer, transformer, grouper, modifier et interagir avec les objets. La partie visuelle de la norme MPEG-4 décrit les méthodes pour la compression des images et des vidéos, des textures. Le codage flexible (scalable) des flux vidéo est intégré dans la norme MPEG-4.

Il est possible de ne décoder qu'une partie du flux vidéo. Cette caractéristique permet d'adapter le flux vidéo aux limitations de débit du canal de transmission ou aux capacités du décodeur.

Le codage des objets audio fournit des outils pour la représentation des sons aussi bien naturels (parole, musique) que synthétiques basés sur une description de structure. Le débit alloué varie entre 6 kb/s et 24 kb/s et permet également une transmission extensible. Enfin, afin d'améliorer la

qualité de service (par exemple pour les applications mobiles), le standard prévoit de techniques de resynchronisation du flux vidéo. Cela permet de corriger ou de limiter les effets des erreurs de transmission. Bien que les domaines d'application des standards MPEG-4 et H.263+ soient différents, ils sont parfois recouvrants. Par exemple, les deux standards tentent de coder efficacement les vidéos pour des débits compris entre 24 kb/s et 64 kb/s. Ces recouvrements proviennent du fait que de nombreux participants à l'élaboration du standard MPEG-4 Vidéo participaient également à l'UIT-T [10].

2.1.4 Norme MPEG-7

Développé comme ses prédécesseurs par le Moving Picture Experts Group, MPEG-7 [11] est une simple évolution ou amélioration de normes ultérieures. Par ailleurs, MPEG-7 propose d'adresser un large spectre d'applications. En effet son domaine d'application couvre : l'archivage radio, la télévision et le cinéma, les systèmes d'informations géographiques ou touristiques, la télédétection, le domaine biomédicale, etc.

Au sein de la norme MPEG-7, les applications sont divisées en deux catégories : les applications de type pull et celles de type push.

La catégorie pull regroupe les applications qui ciblent l'extraction et la recherche de données audiovisuelles. Dans ce contexte, il est explicitement fait mention des applications suivantes :

- Stockage et requête de l'information dans des bases de données vidéo.
- Fourniture d'images et de vidéos pour la production professionnelle.
- Applications commerciales musicales.
- Bibliothèques d'effets sonores.
- Bases de données de discours historiques.
- Recherche de films selon les événements.
- Enregistrement et recherche de marques enregistrées.

La catégorie push suit plutôt le paradigme du filtrage et de la sélection de données pour des applications spécifiques au monde du broadcasting ou du webcasting.

Enfin, MPEG-7 se propose également d'aborder un certain nombre d'applications ayant un profil hautement spécialisé voire professionnel. Dans ce cadre, nous retrouvons des applications telles que l'imagerie satellitaire (remote sensing application), la télésurveillance, le contrôle par la vision (visually-based control). . .

2.2 Les normes de l'UIT-T

Les principaux standards normalisés par l'ITU-T sont :

2.2.1 Norme H.261

La norme H.261 standardisée en 1988 a été développée pour permettre la transmission de vidéos à débits variable (px64 kb/s où p varie de 1 à 30). Elle cible particulièrement des applications de type vidéoconférence et téléphonie [12]. Le format d'image utilisé est le QCIF (144 X 176 pixels), optionnellement le CIF (288 X 352 pixels) et le Sub-QCIF (96 X 128 pixels). La fréquence de rafraîchissement des images de base est de 29.97 Hz. Chaque Macro-bloc composant une image peut être codé en mode Intra ou en Inter afin d'exploiter ou non les redondances temporelles entre les images

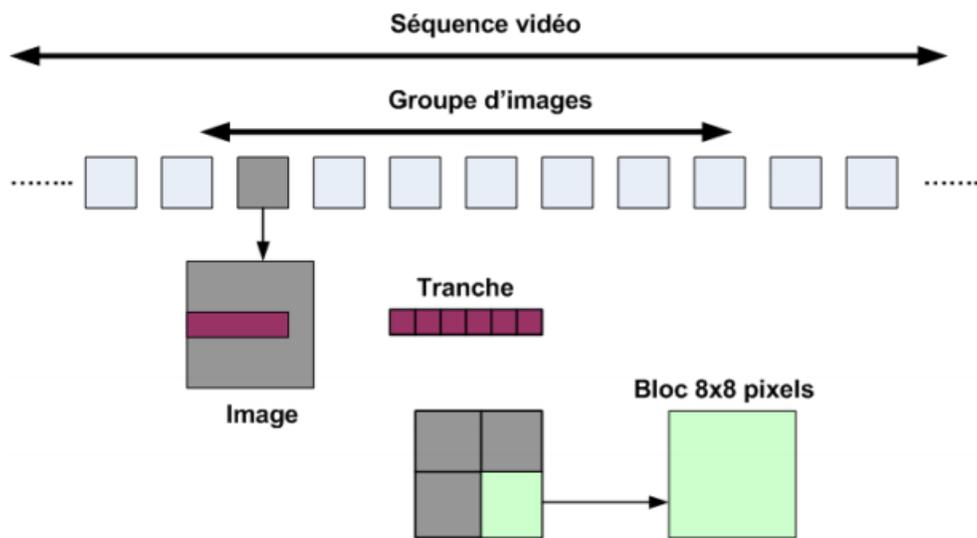


Fig. 2.2 – situation du macro-bloc dans un GOP.

La transformation des 4 blocs de 8X8 pixels d'un macro-bloc 16 X 16 de la représentation temporelle à la représentation fréquentielle est réalisée au moyen d'une TCD (Transformée en Cosinus Discret). Les 64 coefficients fréquentiels obtenus après traitement sont alors quantifiés. L'étape de quantification est la source principale de perte de la qualité visuelle lors de la compression d'un flux vidéo. Cette étape est contrainte par l'utilisateur en fonction du compromis souhaité (taux de compression, qualité visuelle). En effet, en fonction du compromis, le pas de quantification des coefficients fréquentiels de la DCT varie.

En aval de cette étape, les données issues du flux vidéo sont traitées par un codage à longueur variable. En mode Inter, l'estimation de mouvements transmet des vecteurs de mouvement entiers. Les vecteurs de mouvement ne peuvent pas excéder +/-15 pixels. Un seul vecteur de

mouvement est utilisé pour la luminance et la chrominance. Lors de la transmission des données, le flux vidéo est structuré en 4 niveaux comme cela est présenté dans la figure 2.2.

2.2.2 Norme H.263

Il s'agit d'une norme du codage vidéo dédiée aux applications de vidéocommunication à très bas débit. La première version de cette norme a été adoptée en 1995. La norme H263 [13] vise les applications de visiophonie et visioconférence sur réseaux RTC (Réseau Téléphonique Commuté) et RNIS (Réseau numérique à Intégration de Services). Cette norme repose sur les principes de base mis en place dans la norme H.261.

Le codage du flux vidéo est réalisé grâce à un partitionnement de chaque image en Macro-blocs composés de 16 X16 pixels pour la luminance et de 8X8 pixels pour la chrominance (Cb et Cr). Chaque Macro-bloc peut être codé en mode Intra ou en mode Inter. Les redondances spatiales sont exploitées par un codage TCD tandis que les redondances temporelles le sont par une compensation de mouvement. Cette compensation de mouvement utilise des vecteurs non bornés avec une précision au demi-pixel et permet une interpolation bidirectionnelle.

Le décodage d'un flux H.263 est basé sur les décodeurs H.261 améliorés pour supporter les diverses innovations. Quatre améliorations principales permettent d'améliorer les performances des codeurs H.263 par rapport à H.261 :

1. Non restriction des vecteurs de mouvement ce qui permet de référencer des blocs hors des images.
2. Utilisation d'un codage arithmétique à la place d'un codage de Huffman.
3. Utilisation de blocs 8X8 recouvrant pour la compensation de mouvement à la place d'un simple vecteur par Macro-bloc 16X16.
4. Possibilité de coder une image P et une image B ensemble sous forme d'une image PB.

Les formats supportés par cette norme sont le QCIF (144X176 pixels) et le SubQCIF (96 X 128 pixels) et optionnellement le CIF (288x352 pixels), le 4CIF (576x704 pixels) et le 16CIF (1152X1408 pixels).

2.2.3 Norme H.263+

La norme H.263+ est une extension de la norme H.263 possédant de nouvelles fonctionnalités telles que l'extensibilité SNR (Signal to Noise Ratio) et l'extensibilité spatiale et temporelle. Cette norme utilise un codage Intra avancé des Macro-blocs pour améliorer les performances de compression en utilisant notamment une prédiction spatiale des coefficients TCD et un filtrage a posteriori afin de réduire les effets de bloc. Le regroupement des Macro-blocs en tranches permet une réduction de l'erreur résiduelle, une optimisation de la transmission par paquets et une réduction des délais. Un mode de fonctionnement avec une résolution réduite permet de maintenir une fréquence d'images élevée lors de grands mouvements. Cet objectif est atteint en codant une mise à jour basse résolution tout en gardant une haute résolution sur les zones

stationnaires. Cette norme a l'avantage d'utiliser un décodage de l'erreur résiduelle en segments indépendants. Ainsi, elle permet de limiter la propagation d'erreurs dans le temps lorsque les données reçues sont corrompues.

De plus, l'utilisation d'un codage VLC (Variable Length Code) alternatif permet de réduire le nombre de bits nécessaires à l'encodage des blocs prédits composés de grands coefficients. De même, la modification de la technique de quantification améliore le contrôle du débit [14]

2.2.4 H.264 / MPEG-4 AVC

L'émergence du codage H.264/AVC (Advanced Video Coding) a permis d'atteindre une meilleure efficacité des étapes de compression à qualité visuelle équivalente. Cette augmentation des performances de compression est en partie due à une plus grande souplesse dans le partitionnement de l'image en blocs élémentaires vis à vis des normes antérieures. En effet, à partir des avancées proposées dans cette norme, il est possible de regrouper des blocs élémentaires en macro-blocs de taille et de forme variables (bandes, rectangles, carrés de résolution et de dimension variables).

De plus, la taille des blocs élémentaires sur lesquels sont estimés les mouvements entre les images successives a été réduite de 8x8 pixels à 4x4 pixels. Il résulte de ces deux évolutions une meilleure adéquation entre l'image compressée et les données d'origine. De plus, en ce qui concerne le codage des redondances temporelles, la norme H.264/AVC permet de rechercher un bloc similaire se trouvant à une distance temporelle beaucoup plus grande (jusqu'à une distance de 16 images). Actuellement, cette norme reste couramment utilisée par les producteurs et les diffuseurs de contenus vidéo. Toutefois ces principales améliorations possèdent un contre coût important sur la complexité des processus d'encodage et de décodage des flux vidéo.

La figure 2.3 exprime la relation entre les exigences d'archivage et la technologie de compression utilisée. Nous pouvons y observer que pour la même durée d'enregistrement, la compression H.264 requiert deux fois moins d'espace en disque que la compression en MPEG-2.

2.2.5 La norme H.265

La norme H.265 aussi appelée HEVC (High Efficiency Video Coding) ou H.265 (Next-Generation Video Coding) est en cours d'élaboration. Cette nouvelle norme devrait remplacer l'actuelle norme H.264/MPEG-4 AVC développé par le group JVT (Joint Video Team). Cette norme supportera les résolutions 1080p et Ultra HD. Une amélioration de 50% du taux de compression avec une qualité visuelle accrue de 25% est annoncée. Cette amélioration des performances est obtenue au détriment de la complexité calculatoire des étapes de compression et de décompression.

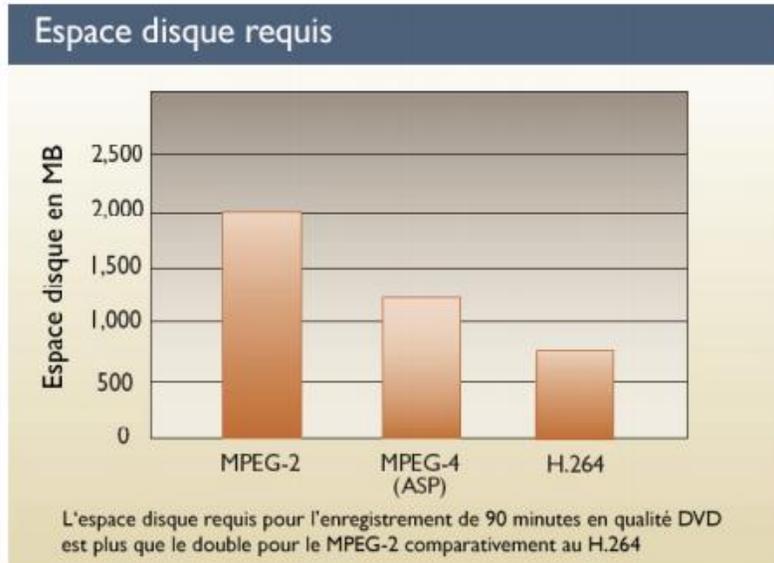


Fig. 2.3 – comparaison entre l'espace disque requis [15]

Les besoin de diffusion pour les réseaux a faible débit (internet, téléphonie, satellite, etc) ont fait émerger les standard de compression de la série H.26X.

Le H.264 (MPEG 4 niveau 10) représente la mouture la plus achevée de cet ensemble puisqu'il offre un PSNR se démarquant nettement des autres standards de compression.

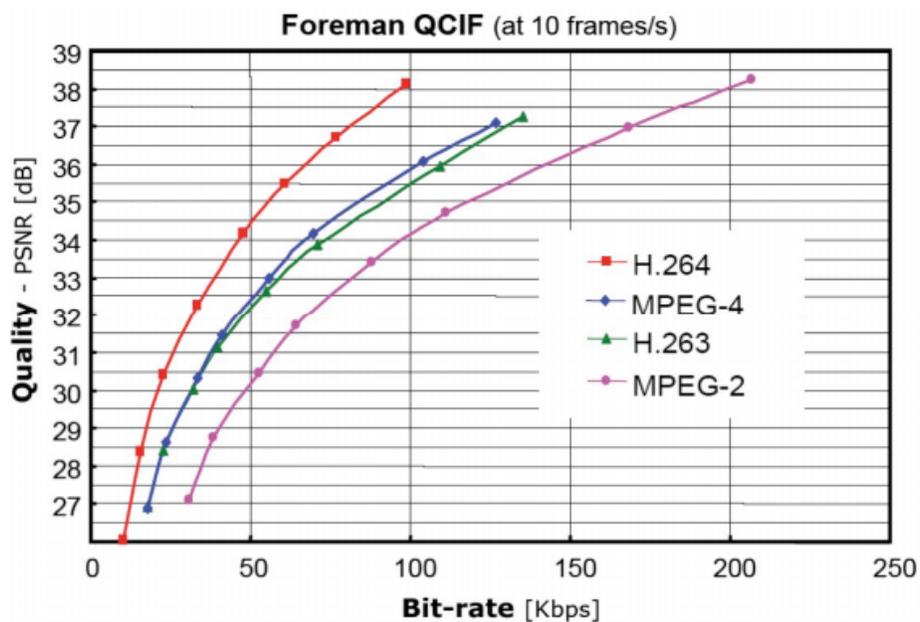


Fig. 2.4 – performance des normes populaires de compression vidéo en termes de qualité PSNR en fonction du débit.

Conclusion

Dans ce mémoire, après une présentation synthétique du principe de compression vidéo, nous avons présenté un historique des différentes normes puis nous avons détaillé le fonctionnement des techniques de compression et de décompression associés à la norme H.264/AVC. Ce standard fournit des performances nettement supérieures en comparaison aux normes antérieures.

Références :

- [1]: C. A. Poynton, A Technical Introduction to Digital Video. New York, NY: Wiley, 1996.
- [2]: Zhou Wang, Alan Conrad Bovik, Hamid Rahim Sheikh, and Eero P. Simoncelli. Image quality assessment: From error visibility to structural similarity. IEEE Transactions on Image Processing, 13(4) :600{612, April 2004.
- [3] : Implantation optimisée d'estimateurs de mouvement pour la compression vidéo sur plateformes hétérogènes multicomposants. Thèse de doctorat de Fabrice URBAN, Décembre 2007.
- [4] : "Itu." <http://www.itu.int>.
- [5] : "Mpeg." <http://www.Mpeg.org>.
- [6] : S. Roux, Adéquation algorithme architecture pour le traitement multimédia embarqué. PhD thesis, France Télécom RD DIH/OCF et TIMA, Institut national polytechnique de Grenoble, France, 2002.
- [7] : "Standard mpeg-1 : Iso/iec 11172-2. coding of moving pictures and associated audio for digital storage media at up to about 1,5 mbits/s," tech. rep., 1991.
- [8] : G. ROBERT, Représentation et codage De séquences vidéo Par hybridation de fractales et délimitations finis . PhD thesis, Université Joseph Fourier-Grenoble 1 Sciences Géographie, 2000.
- [9] : "Standard mpeg-2 : Iso/iec 13818-2. information technology-generic coding of moving pictures and associated audio information," tech. rep., 1994.
- [10] : "Standard mpeg-4 : Iso/iec 14496-2. information technology-generic coding of audio-visual objects," tech. rep., 1998.
- [11] : V. L. J.-M. M. Ahmed Mostefaoui, Françoise Prêteux, "Informatique et système d'information,"
- [12] : "H.261 : video codec for audiovisual services at p x 64 kbits/s. recommendation h.261 à huit-t," tech. rep., 1993.
- [13] : "H.263 : video coding for low bit rate communication. première recommandation h.263 à huit," tech. rep., 1996.

[14] : “H.263+ : video coding for low bit rate communication. deuxième recommandation h.263 à huit,” tech. rep., 1998.

[15] : La compression vidéo numérique, comment s'y retrouver dans cet océan de standards et de protocoles... article de René St-Pierre,2007