



# Communication Multimédia

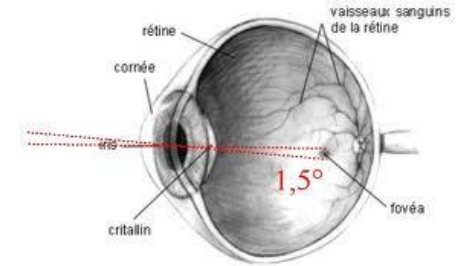
Mustafa Ali Hassoune  
Département d'informatique  
Université des Sciences et de la Technologie d'Oran



# III- Traitement multimédia dans les communications



# III-4 Codage vidéo



L'œil humain a la faculté de former des images précises sur les cellules photosensibles de la rétine pendant quelques millisecondes.

Si une séquence d'images défile à 25 images par seconde ou plus, l'œil ne perçoit pas qu'il s'agit d'images distinctes.  
Les systèmes vidéo et cinématographiques utilisent ce principe pour restituer le mouvement.

Une caméra permet de convertir une image optique en image électronique. Cette dernière est balayée pour obtenir un signal électrique appelé signal vidéo. L'objectif est de capturer un nombre suffisant d'images par seconde afin de donner l'illusion du mouvement.

Les images sont représentées par une grille rectangulaire d'éléments d'image, plus communément appelés pixels (picture elements).

Chaque pixel est codé par un mot binaire représentant l'information de couleur. C'est une combinaison de couleurs suivant l'espace colorimétrique utilisé.

Un signal vidéo numérique se compose d'une suite d'images. Ces images sont ensuite restituées au rythme approprié sur un écran.

# III-4 Codage vidéo

## Pourquoi encoder la vidéo?

La vidéo est constituée de deux composantes essentielles qui sont le son et l'image.

si certains films animés peuvent être visionnés à une cadence de 12 ips (images par secondes), le cinéma lui exige 24 ips, le standard vidéo PAL (Europe et Asie) en exige 25 et le standard NTSC (Amérique et Japon) en impose 30.

Prenons un exemple : une image de qualité DVD au standard PAL a une résolution de 720×576 pixels, chacun de ces pixels étant codés sur 16 voire 24 bits afin d'assurer un rendu optimal des couleurs. On arrive à une taille d'environ 1,2 Mo pour **une** image, soit près de 30 Mo pour une seconde de vidéo.

Ajoutons-y le son : en qualité CD et sans aucune forme de compression, une seconde de son prend 172 Ko, ce qui paraît négligeable comparé au poids de la vidéo, mais atteindra tout de même 1,18 Go pour un film de deux heures.

Heureusement, les diverses techniques d'encodage audios (AC3, MP3, WMA, Ogg Vorbis, etc.) sont capables de faire tomber ce chiffre à environ 280 Mo. Ces deux chiffres paraissent toujours bien faibles par rapport aux 210 Go requis uniquement pour stocker les images sans compression.

En effet, 210 Go suffisent amplement à remplir un disque dur "grand public" et équivalent à plus de 20 DVDs. Il est donc nécessaire de trouver une solution pour réduire le poids de la vidéo : l'encodage vidéo.

# III-4 Codage vidéo



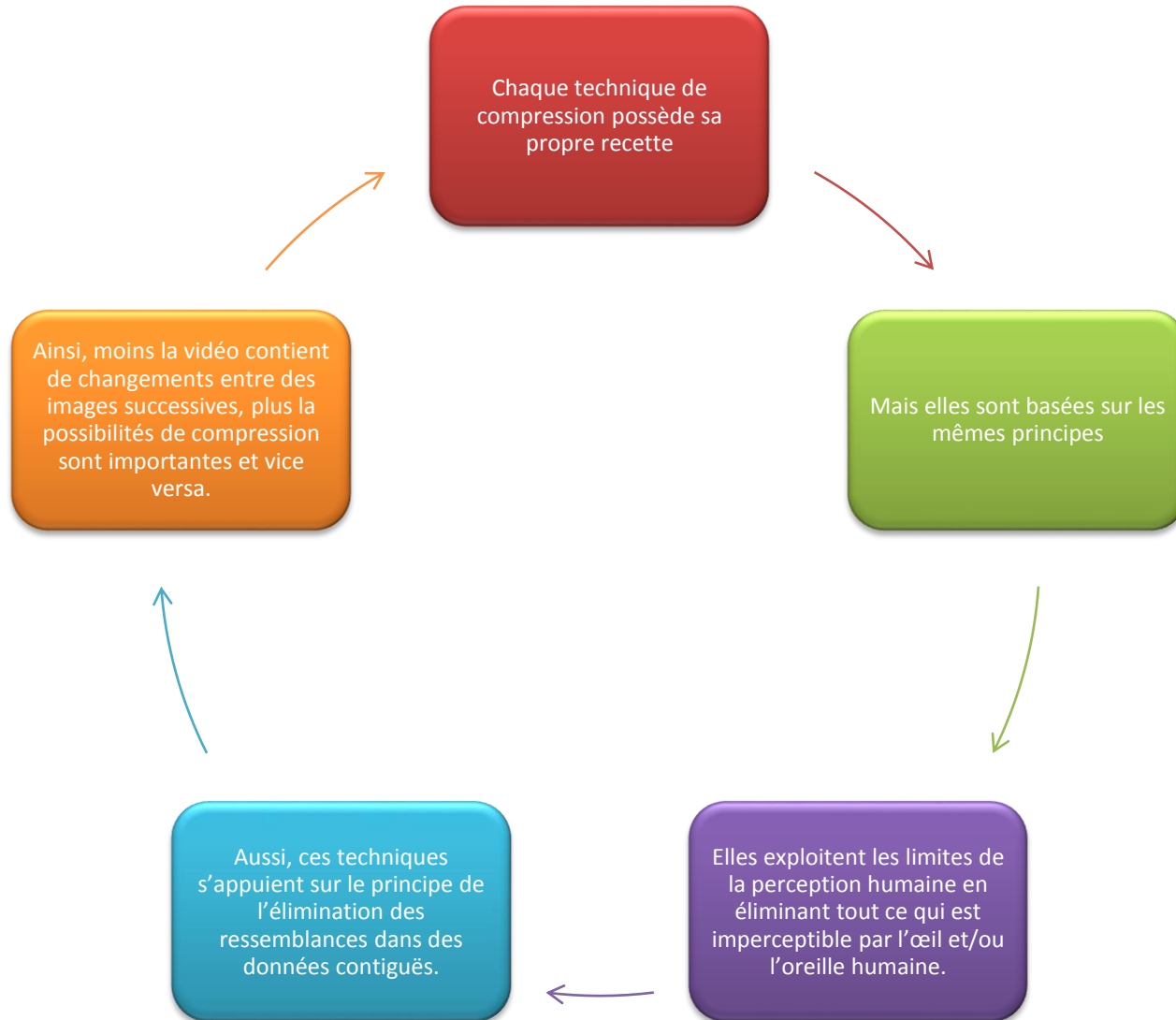
**Décomposition Hiérarchique d'une vidéo**

## III-4 Codage vidéo (Compression)

# Compression Vidéo

La compression vidéo est une méthode de compression de données, qui consiste à réduire la quantité de données, en minimisant l'impact sur la qualité visuelle de la vidéo. L'intérêt de la compression vidéo est de réduire les coûts de stockage et de transmission des fichiers vidéo.

# III-4 Codage vidéo (Compression)



# Principes de l'encodage vidéo

## Principes de base

Sans  
Perte

Avec  
perte



# Principes de l'encodage vidéo

## Sans perte

n'apportent en moyenne qu'un gain de 50 % en terme de poids. Ils sont donc surtout utilisés dans le milieu professionnel (exemple : montage vidéo) et par certains puristes ou alors dans le cadre d'utilisations spécifiques.

On citera pour exemple les codecs HuffYUV, AviZLIB et MotionSZH).

Cette compression lossless repose sur des principes similaires à ceux de la compression de fichiers quelconques (compressions ZIP, TAR.GZ, etc.), incluant l'élimination optimale des redondances dans les données.

Par exemple, si on trouve dans un fichier trente octets d'affilée ayant la même valeur, il est beaucoup plus économique de remplacer ces trente octets par un caractère spécial suivi du nombre de redondances (ici trente) et enfin de la valeur concernée.

L'optimisation de ce principe pour des données vidéo aboutit à un codec de compression vidéo lossless. Exemple : si une image contient 8 pixels côte à côte ayant exactement la même couleur, le 1<sup>er</sup> pixel est codé normalement et il suffit ensuite d'indiquer que les 7 suivants ont la même couleur.

# Principes de l'encodage vidéo

## Avec perte

Il suffit de réduire la valeur d'une ou plusieurs caractéristiques de cette vidéo pour cela.

On peut diminuer le nombre d'images par secondes ; on a ainsi moins d'images à stocker, d'où un gain de place conséquent ; en contrepartie, la lecture apparaîtrait saccadée en dessous de 16 ips pour les personnes les plus tolérantes.

# Principes de l'encodage vidéo

## Avec perte

Il suffit de réduire la valeur d'une ou plusieurs caractéristiques de cette vidéo pour cela.

il est possible de diminuer le nombre de couleurs utilisées de 16 777 216 à 65 536 voire 256 couleurs mais là encore la perte serait visible au niveau des dégradés.

# III-4 Codage vidéo (Compression)

## Niveaux (Levels)

On peut également jouer sur sur la résolution de l'image.

Le MPEG a ainsi défini quatre résolutions d'images standards, appelés aussi Niveaux ou Levels. Ces "Levels" peuvent également porter sur le débit vidéo.

Le premier d'entre eux, appelé "Low Level" correspond à une résolution de 352×288 en PAL ou 352×240 en NTSC. Il est souvent utilisé en MPEG-1 et correspond également à la résolution d'un VCD.

Le second, le "Main Level" (720×576 en PAL ou 720×480 en NTSC), est aussi le plus répandu, d'autant qu'il est utilisé pour la réalisation des DVD-Video.

Les deux autres formats, le "High Level" (1920×1080) et le "High 1440 Level" (1440×1152), sont destinés à la TVHD.

À partir des différents niveaux, on peut en déduire le débit maximal correspondant :

4 Mbps pour le Low Level (abrégé LL)

15 Mbps pour le Main Level (abrégé ML)

60 Mbps pour le High 1440 level

80 Mbps pour le High Level (abrégé HL)

Bien sûr, ces Niveaux ne sont que des recommandations qui n'empêchent nullement l'utilisation d'une résolution supérieure. En effet, la résolution maximale théorique du MPEG est de 16383×16383, c'est-à-dire de quoi remplir la surface de 256 moniteurs de PC.

# III-4 Codage vidéo (Compression)

## Notion de bitrate

Le bitrate est le nombre de bits qu'une seconde de vidéo devra occuper après encodage ; ce paramètre est défini par l'utilisateur avant la compression. Le bitrate s'exprime généralement en kilobits par seconde(kbps) ou Megabits par seconde (Mbps).

Exemple : une vidéo d'une durée de 2 heures (7200 secondes) encodée avec un bitrate constant de 900 kbps aura une taille de 772 Mo :

$$\text{Taille finale} = \text{Durée (en secondes)} \times \text{bitrate (en kbps)}$$

Le bitrate peut être spécifié de manière constante (CBR pour Constant BitRate) ou de manière variable (VBR pour Variable Bitrate). Dans le cas du MPEG-2, le CBR est généralement préféré au VBR.

Le bitrate étant le rapport entre la taille (donc la qualité) de la vidéo finale et sa durée, il influe lourdement sur la qualité de la vidéo.

The screenshot shows the 'Nic's Mini Calc' application window. It contains several sections for configuring video encoding:

- Length:** Film Length: 2.00 (with a dropdown for 25.000 FPS). A note below says '(e.g. Type in 1.50 for 1hr 50m) (FPS)'.
- Audio:** Radio buttons for 'Audio Bitrate: 0' (selected) and 'Audio Filesize: ...'.
- Subs Size:** 'Subtitle Filesize: ...'.
- Final Size:** 'Final Desired Size: 776' (with a dropdown).
- Overhead:** Radio buttons for '1 CBR MP3' (selected), '1 VBR MP3', and '1 AC3'. A checkbox for 'AVI Frame Overhead' is checked. A dropdown shows '1' and '40ms'.
- Final Results:** XviD Video Size: 790315; Size in MB: 772 MB; Avg Bitrate: 899 kBit/s.

Buttons for 'Generate' and 'Done' are at the bottom.

**4K DCI (4096x2160)**

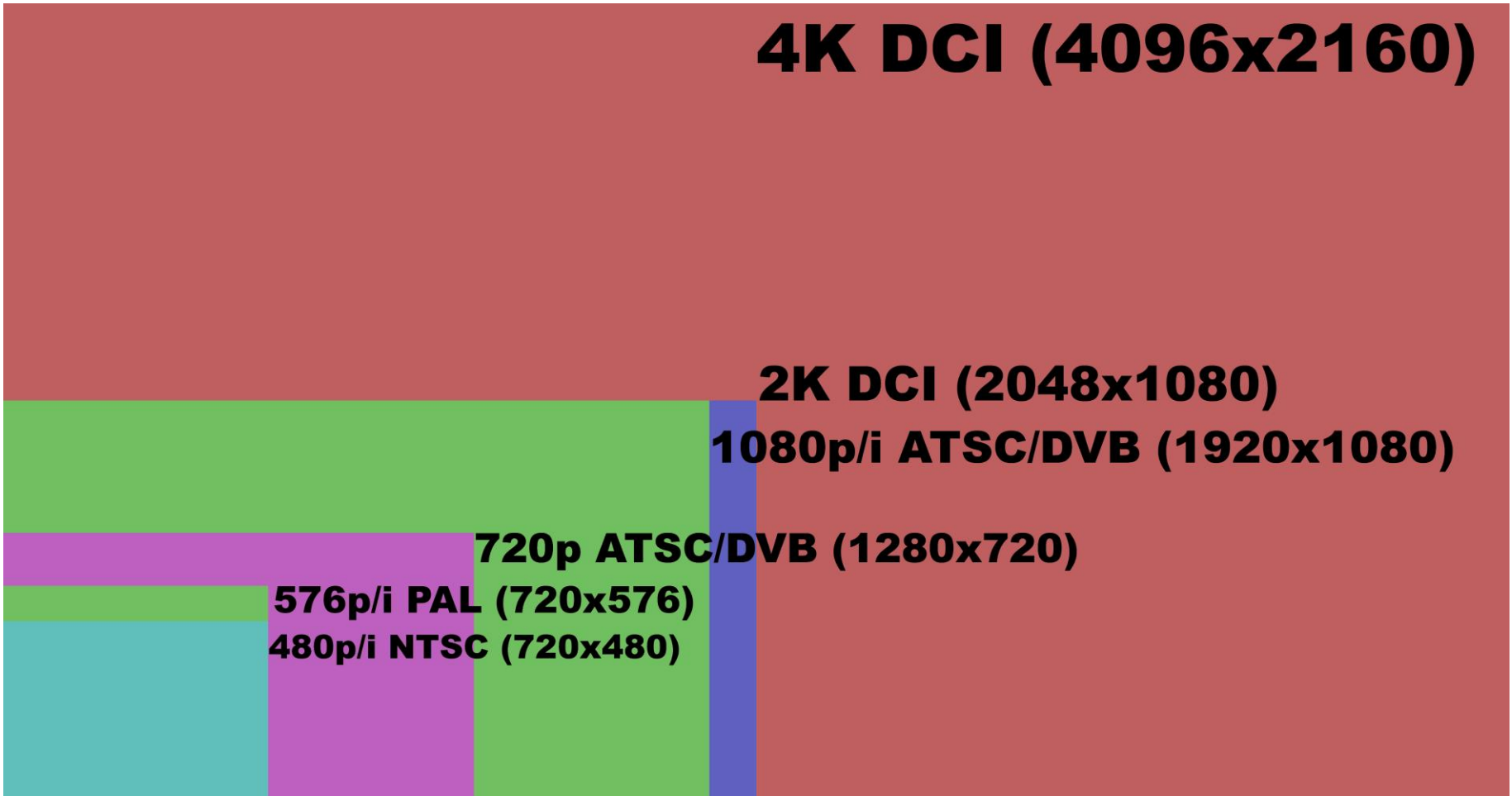
**2K DCI (2048x1080)**

**1080p/i ATSC/DVB (1920x1080)**

**720p ATSC/DVB (1280x720)**

**576p/i PAL (720x576)**

**480p/i NTSC (720x480)**



# Exercice

- Calculer la taille d'une séquence vidéo après compression sachant que sa taille avant compression était : 500 Mo , Qualité HD: 1080 x 1280, Couleurs : 24 bits, 24 fps
- La compression utilisée: 560 x 640 , 6 fps , Couleurs 16 bits

# III-4 Codage vidéo (Compression)

## Redondances spatiales et temporelles

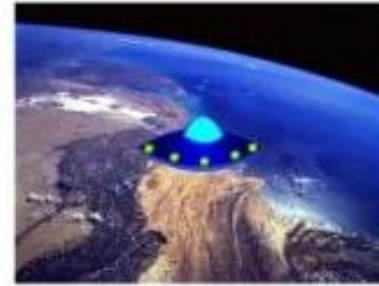
Les deux types de redondances visées par l'encodage MPEG-2 mais aussi par la plupart des codecs vidéos sont la redondance spatio-temporelle et la redondance spatiale :

la redondance spatio-temporelle est la réapparition de certains éléments d'une image donnée dans l'image suivante. Exemple : un objet se déplace sur la longueur de l'image ; l'encodage consistera à coder la première image de la scène en entier, avec une compression classique de type JPEG (Joint Photographic Expert group) puis à ne stocker pour les images suivantes que les différences par rapport à la première image

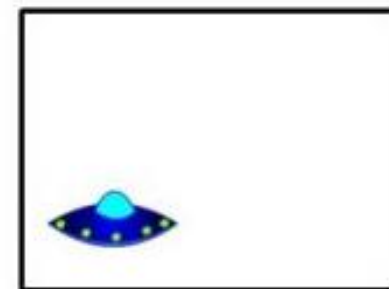
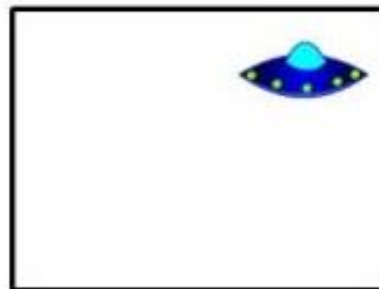
la redondance spatiale est la réapparition d'une partie de l'image sur cette même image ; la redondance peut être exactement la même ou alors présenter une légère différence. Selon le "bitrate" alloué à l'encodage de la vidéo, ces redondances seront plus ou moins assimilées comme étant la même.



# III-4 Codage vidéo (Compression)

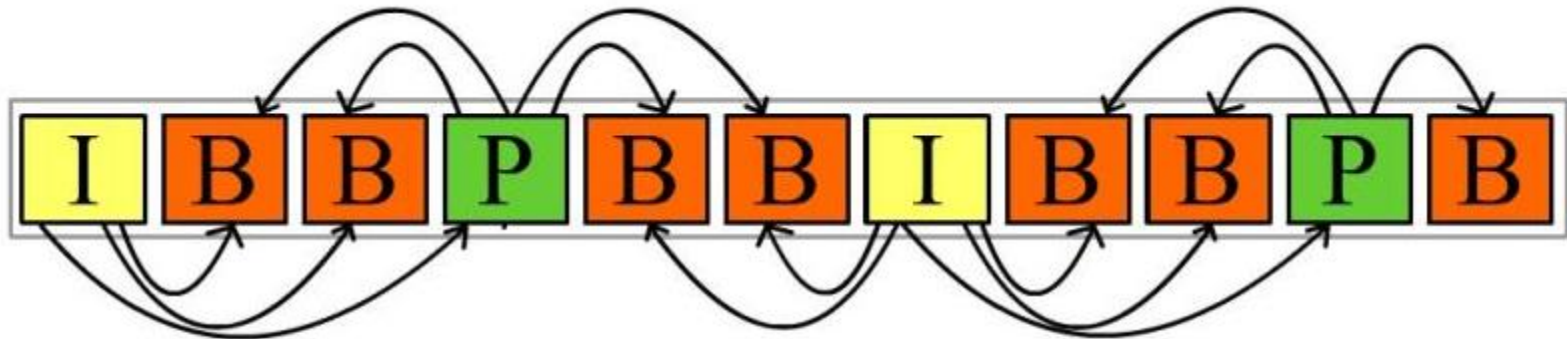


*Plutôt que de stocker intégralement les 4 images ci-dessus, on préférera stocker la première image puis ne retenir pour les suivantes que les différences par rapport à cette première image :*



Une autre possibilité aurait consisté à compresser chaque image de la vidéo en JPEG puis à rassembler toutes les images ; malheureusement le gain est insuffisant et cette technique, utilisée pour le MJPEG, demande encore beaucoup d'espace mais offre la possibilité d'accéder rapidement à chaque image.

# III-4 Codage vidéo (Compression)



Chaque image **P** est produite à partir d'une image **I**

Chaque image **B** est produite à partir d'une image **I** et d'une image **P**

# III-4 Codage vidéo (Compression)

Trois types de trames sont considérées :

## Trames I :

Ces trames sont codées sans aucune référence à autre image de la séquence vidéo. Les trames I permettent de réaliser l'accès aléatoire et les fonctionnalités FF/FR, bien qu'elles ne permettent qu'un très bas taux de compression.

## Trames P :

Ces trames sont codées avec une référence à l'image précédente (trame I ou trame P). Ces trames sont utilisées pour la prédiction de trames futures ou passées et elles ne peuvent pas être utilisées pour réaliser l'accès aléatoire et les fonctionnalités FF/FR.

## Trames B :

Elles ont besoin des trames futures et passées comme référence pour être codées. Elles sont utilisées pour obtenir un très haut taux de compression. Elles ne sont jamais utilisées comme référence.

# III-4 Codage vidéo (Compression)

## Images I, B et P et GOP

Lors de l'élimination des redondances spatio-temporelles, l'image de référence, c'est-à-dire la première image qui est entièrement codée, est appelée "image I" (de l'anglais Intra Frame).

Les autres types d'images sont P et B. Les images P (Prédites) sont composées à partir de l'image I ou P qui les précède. Les images B (Bidirectionnelles) sont construites à partir des images I et P qui les entourent.

Chacune de ces images est elle-même compressée avec une technique proche du JPEG.

En moyenne :

une image P (Prédite) représente 50 % du poids d'une image I

une image B (Bidirectionnelle) représente 15 % du poids d'une image I

C'est donc grâce à ces types d'images que l'on parvient à réduire le poids d'une vidéo.

Il est également possible de faire varier le nombre d'images I, B et P qui entrent dans la composition de la vidéo.

Cette valeur s'appelle un GOP (Group Of Pictures = Groupe d'Images), et est fixe sur toute la longueur du film. Un GOP commence par une image de référence I et est suivi par un certain nombre d'images P et B. La taille de ce GOP influe sur le poids de la vidéo finale : plus le GOP est long (donc plus il y a d'images P et B) et plus le fichier sera léger...

On trouve des GOP d'une seule image (de type I), de deux images (sur le modèle IB ou IP), de six images (IBBPBB), de huit images (IBBPBBBP), voire de 12 images comme pour le MPEG1 (IBBPBBBPBB). Il n'y a pas vraiment de règles précises dans la mesure où c'est le fabricant de l'encodeur MPEG qui définit le GOP comme bon lui semble.

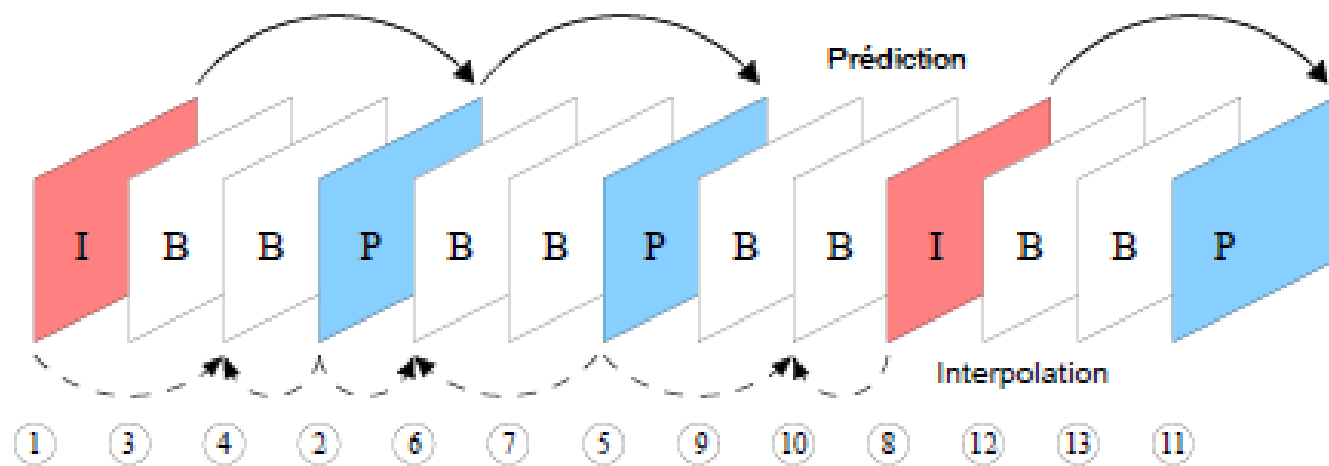


Figure 1.4 – Ordre de codage dans le modèle IPB de la norme MPEG-1.

# III-4 Codage vidéo (Compression)

## MPEG-1

L'utilisateur peut arranger la séquence des différents types de trame selon le besoins de l'application. Généralement une séquence vidéo codée en utilisant seulement des trames I (I I I I . . .) donne un haut degré d'accès aléatoire, de FF/FR et d'édition, mais un taux très bas de compression.

Une séquence vidéo codée seulement avec des trames P (I P P P P P P I P P P P . . .) permet un degré moyen d'accès aléatoire et de FF/FR.

Si on utilise les trois types de trames (I B B P B B P B B I B B P . . .) on arrive à un grand taux de compression et un raisonnable degré d'accès aléatoire et de FF/FR, mais on augmente beaucoup le temps de codage. Pour des applications comme la vidéotéléphonie ou la vidéoconférence ce temps peut devenir intolérable.

# Exercice

- Nous proposons une séquence vidéo de 2 min (25 i/s) dont la taille du fichier est 60 Mo.
- Nous utilisons MPEG1 pour compresser cette vidéo, dont B=20% d'une image avant compression et P=50% et I=100 % , en utilisant le codage:  
IBBPBBIBBP
- Quelle est la taille du GOP utilisé?
- Calculez la taille du fichier généré (compressé)

# III-5 Codage vidéo Hiérarchique

L'encodage hiérarchique d'une séquence vidéo en multi couche se traduit par une décomposition en ondelettes.

Il consiste à compresser un contenu en plusieurs couches correspondant à différents niveaux de qualité:

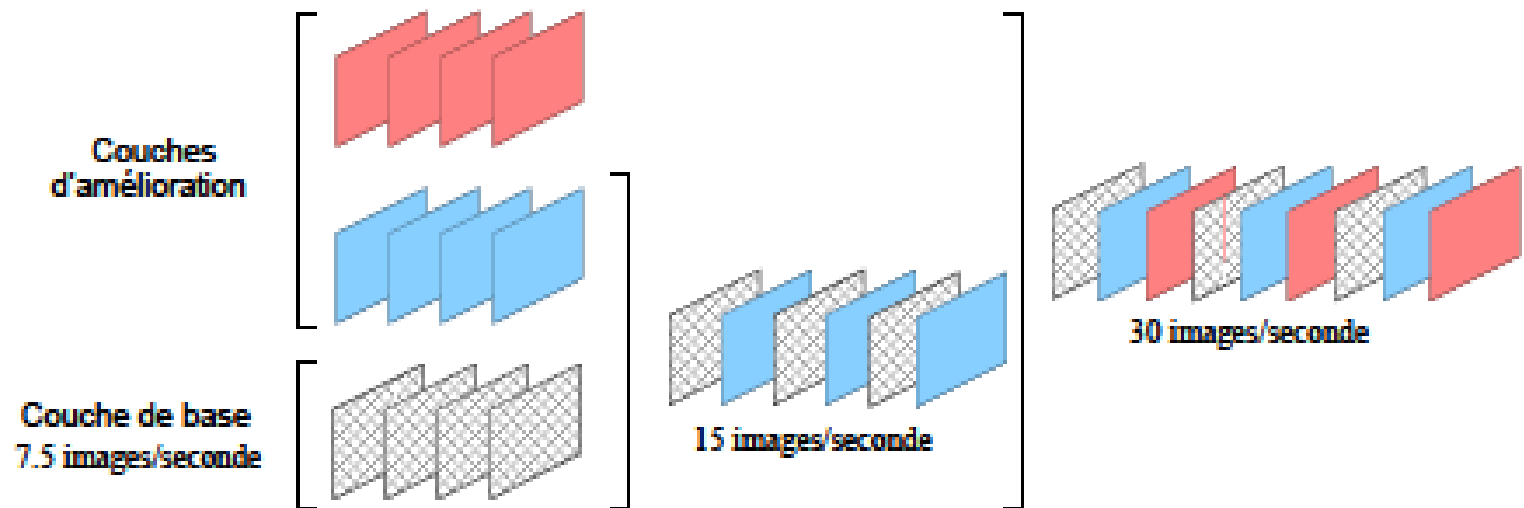
Une couche de base BL (Base Layer)

Et une ou plusieurs couches d'amélioration EL (Enhancement Layer) que l'on peut utiliser l'une après l'autre pour raffiner la qualité de la couche de base.

Notons que l'amélioration de la qualité peut prendre des dimensions spatiales ou temporelles, voire une combinaison des deux.



# III-4 Codage vidéo (Compression)



(b) Hiérarchisation temporelle

# SVC (Scalable Video Coding)

*Scalable Video Coding (SVC)* est le nom donné à une norme de compression vidéo développée conjointement par UIT-T et l'ISO. Les deux groupes ont créé le *Joint Video Team (JVT)* pour développer la norme H.264, ou MPEG-4 AVC (ITU-T Rec. H.264 | ISO/IEC 14496-10 AVC). L'objectif de SVC est d'offrir un contenu adaptable ou *échelonnable (scalable)*, c'est-à-dire que le contenu peut être converti une fois et offrir ensuite différents débits avec différentes qualités.

# Principes de la scalabilité

La *scalabilité* (évolutivité ou adaptabilité, en bon français) est la possibilité de représenter un signal à différents niveaux d'information. Le signal est codé dans un seul flux binaire de manière à offrir la possibilité de décoder un flux de base et des flux englobants dont la qualité augmente successivement. Pour ce faire, trois types de scalabilité sont définis :

la scalabilité spatiale  
qui permet d'offrir  
plusieurs niveaux de  
résolution,

la scalabilité  
temporelle qui  
permet d'offrir  
plusieurs fréquences  
temporelles du  
signal,

la scalabilité en  
qualité qui permet  
d'offrir différentes  
qualités d'image.

# Scalabilité spatiale

- La scalabilité spatiale correspond au choix de la résolution des images reconstruites (par exemple SD, QCIF ou CIF ). La scalabilité spatiale est obtenue au moyen d'une décomposition pyramidale en différents niveaux spatiaux.

# Scalabilité temporelle

- La scalabilité temporelle correspond à la fréquence des images du flux vidéo décodé. Les différentes fréquences sont obtenues à l'aide d'une structure hiérarchisée d'images.

# Scalabilité en qualité

- La scalabilité SNR (Signal to Noise Ratio, ou Rapport signal sur bruit), ou en qualité, consiste à augmenter le rapport signal sur bruit d'une couche, c'est-à-dire à réduire la distorsion de quantification entre l'image originale et l'image reconstruite.

**Scalable  
bit stream**



QCIF



CIF



4CIF

Spatial  
scalability



15 Hz



30 Hz

Temporal  
scalability



Low quality



High quality

SNR (quality)  
scalability

# CIF

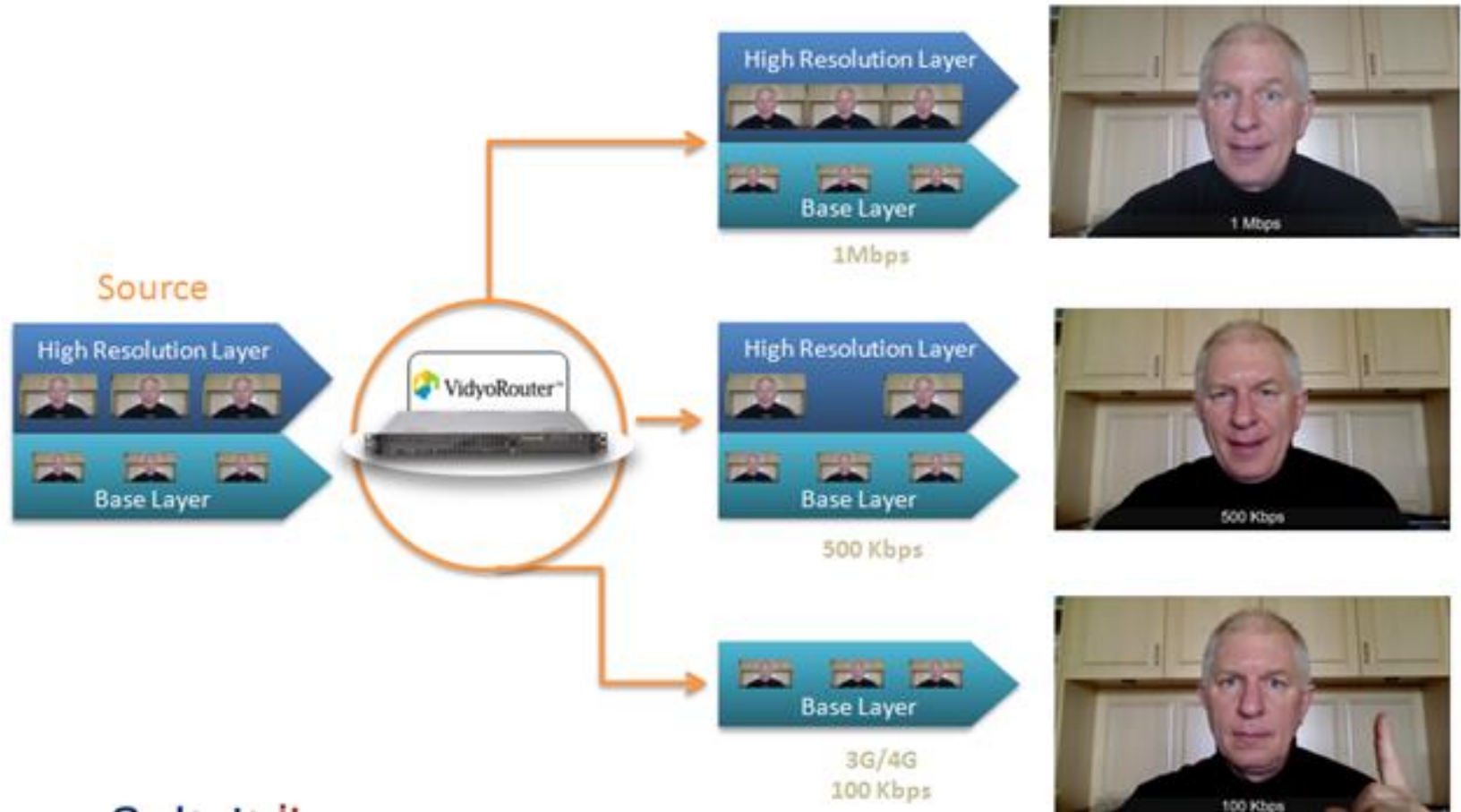
- **CIF**, abréviation de ***Common Intermediate Format***, est une définition standardisée d'image numérique définie par l'Union internationale des télécommunications (ITU) : 352 × 288 pixels.



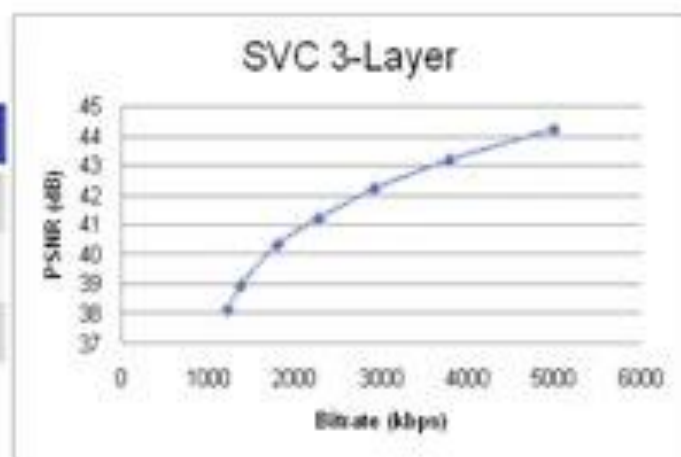
# Définitions des formats dérivés

<b>Abréviation</b>	<b>Définition (en pixels)</b>
SQCIF (Sous-Quart de CIF)	128 × 96
QCIF (Quart de CIF)	176 × 144
CIF	352 × 288
4CIF (4 × CIF)	704 × 576
9CIF (9 × CIF)	1056 × 864
16CIF (16 × CIF)	1408 × 1152

# SVC - Dynamically Adapts to User Environment



Resolution	Bitrate	Avg. Y-PSNR
HDTV	2.92 Mbps	42.24 dB
SDTV	1.05 Mbps	40.98 dB
LDTV	270 Kbps	40.00 dB



3 Spatial Layer : 1280x704  
Platform : PC with HDTV

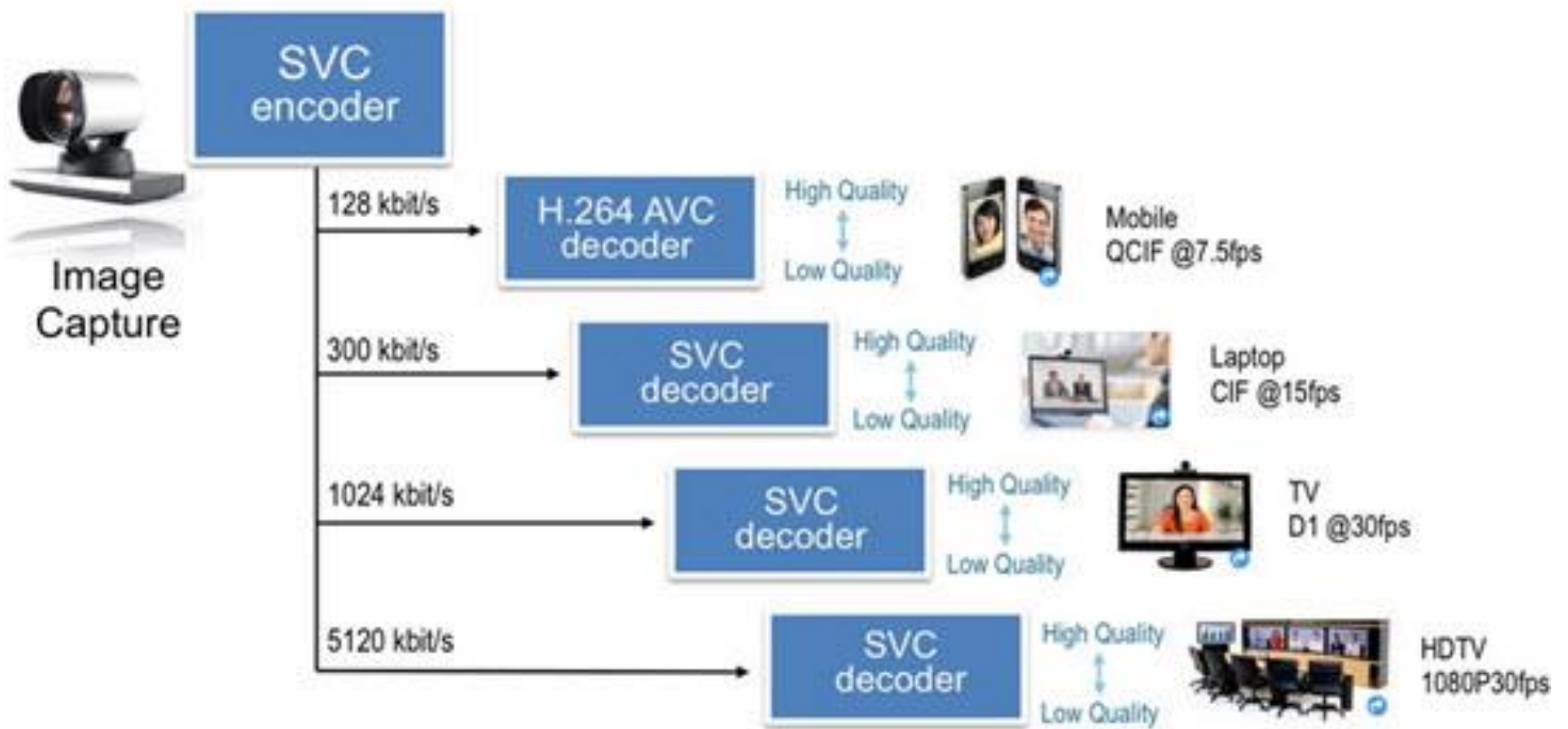


2 Spatial Layer : 640x352  
Platform : Laptop

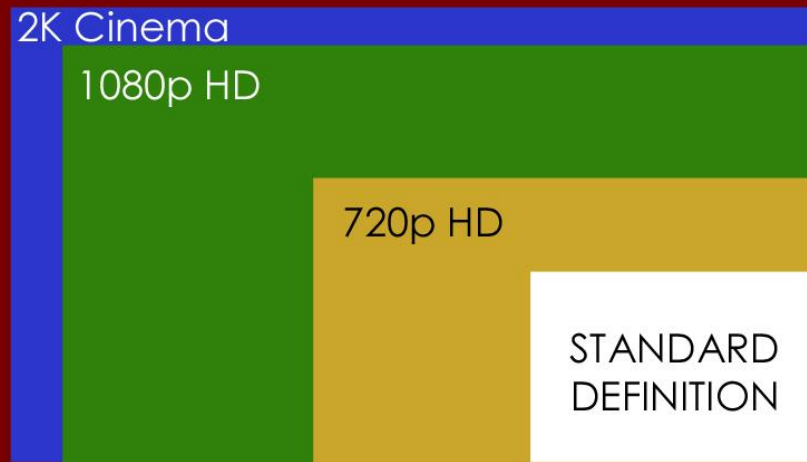


1 Spatial Layer : 320x176  
Platform : PDA

Fig. 2



Red 4K Images!!!



# RED RESOLUTION COMPARISON