

Analyse de la vidéo

Chapitre 5 - Compression de la vidéo

10 mars 2015

Plan de la présentation

- 1 Standards de compression
 - JPEG
 - MPEG
 - MPEG2
 - H.263
 - MPEG4
 - H.264/MPEG4 Part 10 AVC
- 2 Formats conteneurs (avi, mkv, mov, mp4, ...)

Standards de compression vidéo

- JPEG
 - Joint Photographic Experts Group
 - Compression d'images fixes
- MPEG1
 - Moving Picture Experts Group
 - Compression de vidéo pour CD/Internet
- MPEG2
 - Compression de vidéo pour télévision digitale
- MPEG4
 - Compression de vidéo pour applications mobiles
- H.261/H.263
 - Compression de vidéo pour téléconférence

Standards de compression vidéo

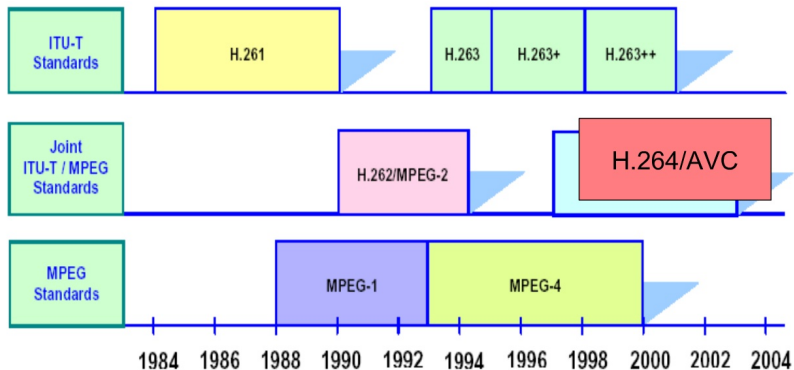


Figure 1. Progression of the ITU-T Recommendations and MPEG standards.

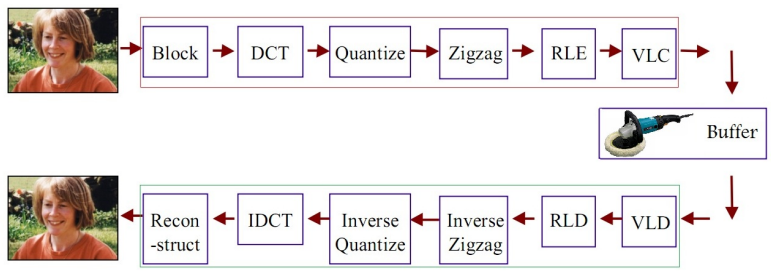
ITU-T VCEG (Video Coding Expert Group)

Standards pour des systèmes de compression vidéo avancés axés sur les applications concrètes de télé-conférence et communication.

ISO/IEC MPEG (Moving Picture Experts Group)

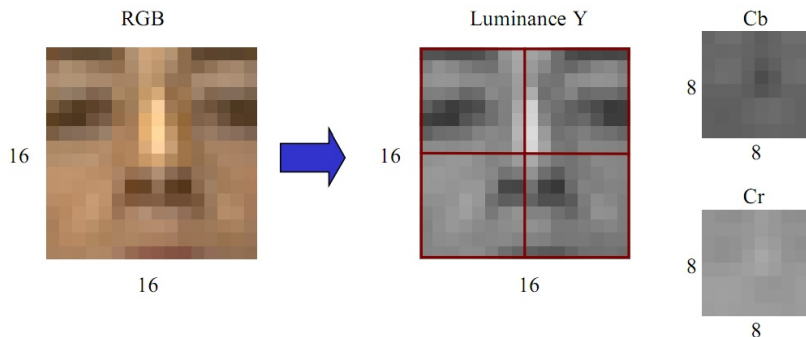
Standards de compression, encodage, traitement, représentation de la vidéo et de l'audio.

codec JPEG



Décomposition de l'image en blocs

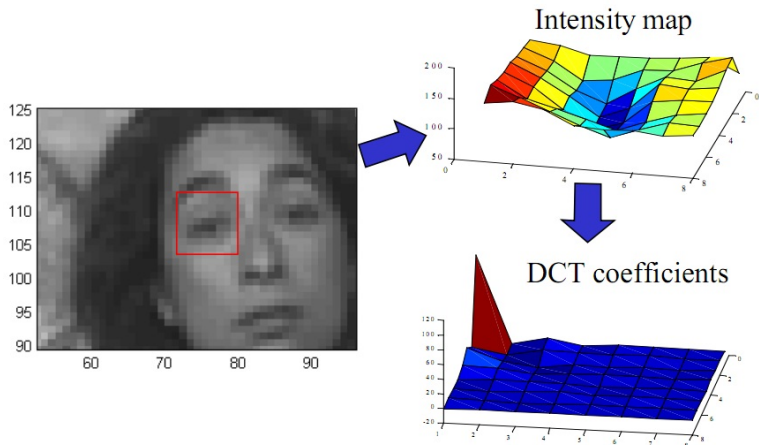
- On subdivise l'image en blocs 8×8 .
 - On exprime l'image RGB (8 :8 :8) dans l'espace YCbCr (4 :2 :2).
- On utilise une demi-résolution pour les canaux chromatiques (Système visuel humain)



DCT

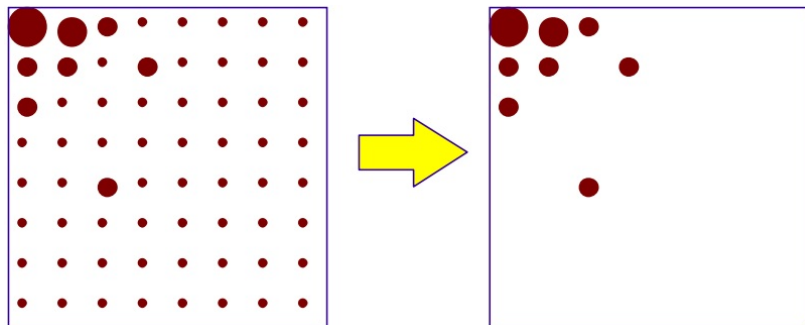
Chaque bloc 8×8 est exprimé en coefficients du domaine fréquentiel.

- L'énergie se concentre dans les coefficient de basse fréquence.



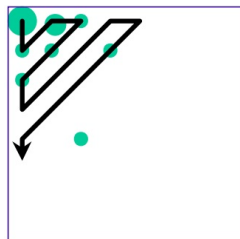
Quantification

- On coupe les valeurs faibles de la DCT.
- Perte de précision.
- Peu de valeurs restantes, concentrées dans les basses-fréquences.




Encodage entropique : Parcours en zigzag, RLE, VLC

- On organise les données à l'aide d'un parcours en zigzag.
- On encode les données à l'aide d'un *Run-Length Encoding*.
- On encode les pairs en exploitant leur redondance par un *Variable length code* (Huffman).



RLE

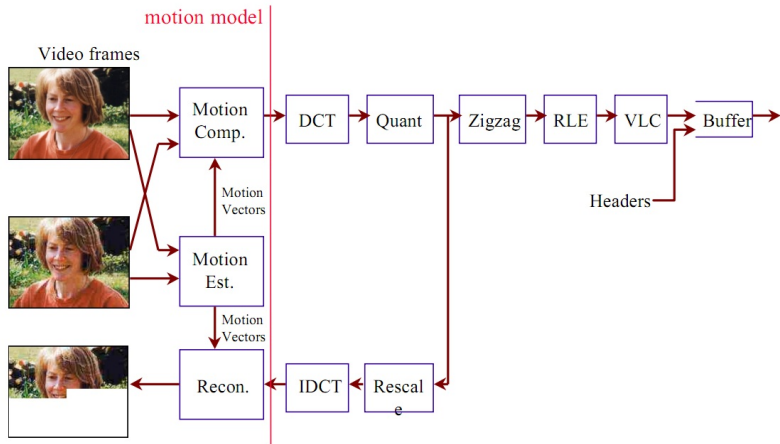
	
data	14,3,4,0,0,-3,0,0,0,0,14,...
(Run,level)	(0,14)(0,3)(0,4)(2,-3)(5,14)...

Redondance spatiale

- On souhaite exploiter la **redondance spatiale de l'information** dans une vidéo.
- On ajoute une modélisation du mouvement à l'encodeur de l'image.

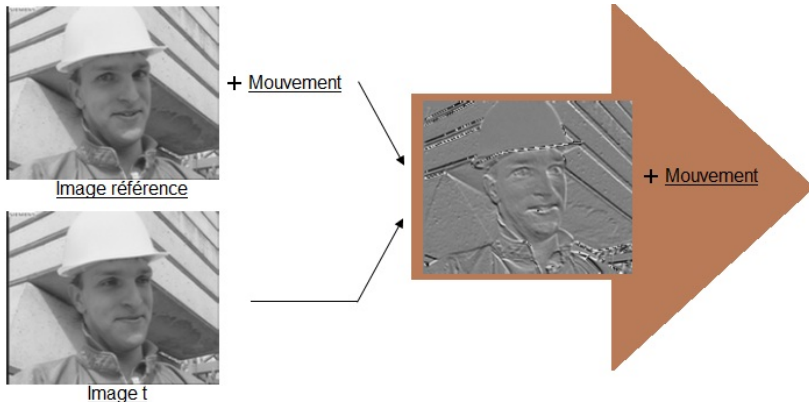


Encodeur video



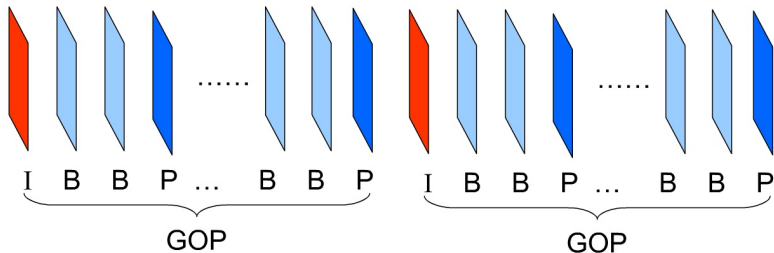
Prédiction et compensation du mouvement

- Pour chaque macrobloc 16x16 de l'image, on recherche dans l'image précédente un macrobloc identique ou semblable.
- On calcule ensuite la différence entre les deux macroblocs.
- On encode les vecteurs de mouvements et les macroblocs de différence.



MPEG

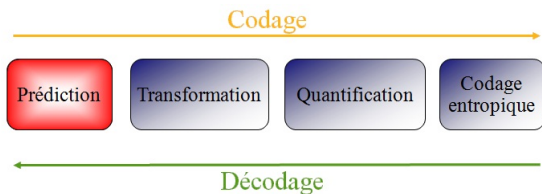
- Video progressive seulement
- Représentation hiérarchique :
 - Sequence d'image
 - Groupe d'image (GOP)
 - Image
 - Macrobloc
 - Bloc



Prediction - Compensation du mouvement

Pour coder les redondances temporelles, MPEG distingue trois types d'images :

- **Les Images I (Intra frames)** : Images entières, qui seront codées en jpeg comme vu précédemment. Elles sont également appelées "Key-frame" car elles constituent des images de référence du flux.
- **Les images P (Predictive frame)** : Elle sont les différences par rapport à une image antérieure. Les images P contiennent des vecteurs de mouvement et des macroblocs.
- **Les images B (Bi-directionnal frame)** : Semblable aux images P, mais utilisent à la fois l'image de référence précédente et la suivante pour sa construction. Une image P peut être considérée comme une image de référence.



Prediction - Compensation du mouvement

La fréquence des images I est réglable lors de l'encodage. Un faible taux d'image I :

- Permet une compression efficace ;
- Est sensible aux erreurs de prédiction.

→ Si une erreur survient sur une Image I, toutes les images P et B faisant référence à cette frame seront affectées.

Ex : L'imagerie par satellite est exposée aux erreurs de transmission et requiert un fort taux d'image I.

Prediction - Compensation du mouvement

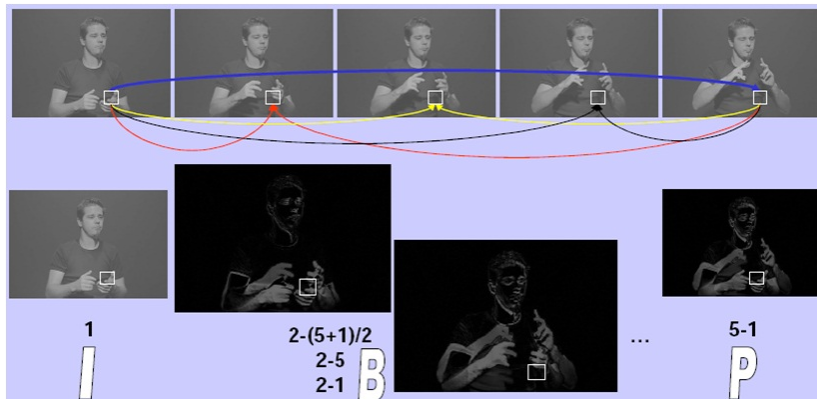
Exemple de disposition des frames dans un fichier MPEG :

IBBPBBPBBPBBIBBPBBPBBP...

- **I** : 8 frame, 14388 octets
- **P** : 22 frame, 6622 octets
- **B** : 56 frame, 2462 octets

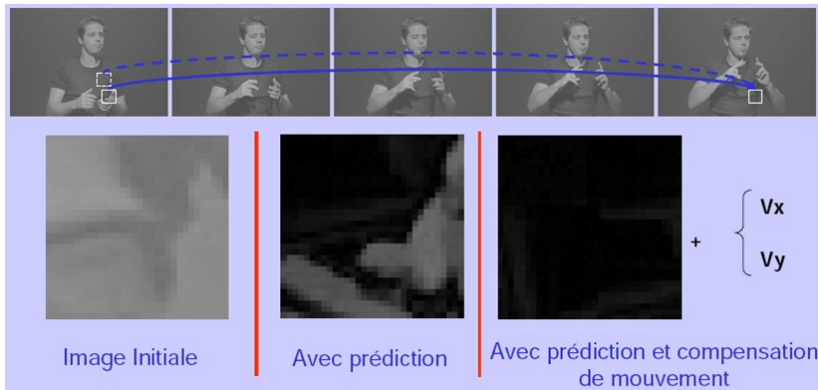
Prediction - Compensation du mouvement

Voyons une chaîne d'image sans compensation du mouvement :

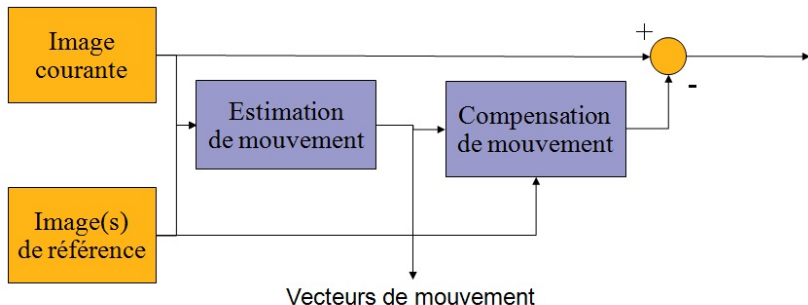


Prediction - Compensation du mouvement

La compensation de mouvement permet de minimiser au maximum l'information à encoder :



Prediction - Compensation du mouvement



Comparaison entre H.261 et MPEG

H.261

- 1990
- Le plus jeune standard basé sur la DCT
- Format d'image CIF (352×288) et QCIF
- Bit rate de $p \times 64$ kbps ($p = \{1, \dots, 30\}$), maximum 1.93 Mbps
- Pas d'image B

MPEG

- 1991
- Format d'image SIF (352×288)
- Bit rate de 1.93 Mbps
- Pas de filtre d'effet de bloc
- Précision du mouvement à demi-pixel
- Image B
- Encodage de l'audio en 2 canaux.

MPEG-2 en résumé

Finalisé en 1994 :

- Optimisé pour la diffusion de télévision numérique
 - Entrelacement de frames possible.
 - Transformation (DCT) et prédiction peuvent être soit inter-champ , soit inter-image (c-a-d sur des blocs obtenus après mélange des 2 frames).
- Débits :
 - 3 à 15 Mbits/sec pour TV, 15 à 30 pour HDTV.
 - Utilisée dans les DVDs et TV numérique.

Différences entre MPEG-1 et MPEG-2

- MPEG-2 permet le mode **progressif ou entrelacé**, alors que MPEG-1 gère uniquement le mode progressif.
→ Gestion des "demi-images".
- MPEG-2 possède un **ajustement du framerate** (temporal scalability) qui permet d'adapter le framerate (fps) au flux.
→ Séquences rapides : framerate important (gain en qualité), Séquences lentes : framerate faible (gain en compression).
- MPEG-2 apporte le son *Surround* et *Dolby AC3*, et des canaux pour plusieurs langues. MPEG-1 code le son en mp3.

Différences entre MPEG-1 et MPEG-2

- Possibilité de différents parcours des coefficients DCT.
- Possibilité de précision des coefficients DC sur 9, 10 ou 11 bits au lieu de 8 bits.
- MPEG-2 a un ajustement spatial : un frame peut être lu à des qualités différentes (bitrates différents). On adapte le bitrate à la complexité de l'image. C'est l'**encodage en deux passes** qui permet cette technique.
- Introduction dans le MPEG2 des **profiles** et des **niveaux**.

Encodage en deux passes

L'encodage en deux passes se déroule de la manière suivante :

1 Première passe : Encodage à un bitrate constant.

L'algorithme accumule des statistiques sur la complexité des images. Pour chaque image, on fait d'une part une estimation grossière du mouvement : on calcule les variations de luminance d'une image sur l'autre. C'est "l'activité du macrobloc". D'autre part, on estime la complexité des textures. La somme des deux valeurs constitue la complexité de l'image.

2 Deuxième passe : Redistribution des bits disponibles sur chaque frame en fonction de la complexité mesurée lors de la première passe.

Le but → Non pas un bitrate constant mais une qualité visuelle constante. Il est évident qu'une scène très précise et rapide nécessite plus de bits qu'une scène très simple.

Profiles et niveaux

- **Profil** : "Spécialisation" du flux pour un mode donné : profils pour la vidéo simple, la vidéoconférence, etc. L'avantage est de prévoir un niveau de qualité adapté à l'utilisation et au transport du flux. → Scalabilité du bitstream, niveau...
- **Niveau** : "Échelle" de qualité de lecture du flux. Ils permettent d'adapter la lecture des flux MPEG-2 à la machine qui décode celui-ci. Un niveau bas permettra de jouer un film sur une machine peu performante. → Résolution

Notation : profile@level

Ex. MP@LL signifie main profile@low level

Profiles et niveaux

HAUT*		4:2:0 1920 x 1152 80 Mb/s I, P, B				4:2:0 4:2:2 1920 x 1152 100 Mb/s I, P, B
HAUT-1440		4:2:0 1140 x 1152 60 Mb/s I, P, B			4:2:0 1140 x 1152 60 Mb/s I, P, B	4:2:0 4:2:2 1140 x 1152 80 Mb/s I, P, B
PRINCIPAL	4:2:0 720 x 576 15 Mb/s I, P	4:2:0 720 x 576 15 Mb/s I, P, B	4:2:2 720 x 608 50 Mb/s I, P, B	4:2:0 720 x 576 15 Mb/s I, P, B		4:2:0 4:2:2 720 x 576 4 Mb/s I, P, B
BAS		4:2:0 352 x 288 4 Mb/s I, P, B		4:2:0 352 x 288 4 Mb/s I, P, B		
NIVEAU PROFIL	SIMPLE	PRINCIPAL	PROFIL 4:2:2	SNR	SPATIAL	HAUT

H.263

- Dérivé de H.261
- Développé pour des **applications à très bas-débit**
 - Meilleure qualité à 20 kbps que H.261 à 64 kbps
 - Utilisé dans MSN Messenger
- Peu atteindre des **hautes résolutions** (16CIF : 1408 x 1152)
- **Précision demi-pixel**
- Options d'encodage supplémentaires
 - **Codage arithmétique**
 - Vecteurs de **mouvements non-restreints** (sortir de l'image)
 - **Prédiction du mouvement avancé**
 - *Run-Length-Last* encoding (RLL)
 - Filtre récursif anti-bloc

H.263

Prédiction du mouvement avancée

- Au lieu d'avoir un mouvement par macrobloc 16×16 , on a un mouvement pour les 4 blocs 8×8 .
- On l'applique dans les cas où le mouvement est complexe.
- On détermine ces cas en comparant les résultats avec 1 et 4 vecteurs de mouvements.
- On aura besoin de plus de bits pour encoder ces mouvements supplémentaires.

H.263

Encodage Run-Length-Last

On encode les coefficient de l'image à l'aide d'un RLL :

LAST

Valeur binaire nous indiquant si le coefficient encodé non-nul est le dernier de la séquence.

RUN

Nombre de 0 précédent le coefficient courant.

LEVEL

Valeur du coefficient.

MPEG4

Avec **MPEG4**, Au départ, on voulait **réduire le bitrate pour des applications mobile**. Finalement, deux standards en ont découlé :

- Basé sur H.263
- Nouveaux concepts plutôt que de nouveaux algorithmes
- Gère plusieurs types de média : audio, objet, image, texture, ...

De tous les formats MPEG4, deux importants en découle :

Partie 2 Visual

- Indexation et l'interactivité sur le contenu audiovisuel
- Encodage basé sur les objets
- *Fine granular scalability (FGS)*

Partie 10 "AVC" (Advanced VideoCoding)

- Recentré sur l'efficacité de codage.
- Abandonne l'interactivité et l'indexation

MPEG4 Part 2 Visual

Objets audiovisuels

- Un codage d'objets : **Objets Audiovisuels (AVO)**.
- Chaque scène est composée de plusieurs objets (segmentation effectuée avant)
- Chaque objet est décrit selon la forme, texture, couleur, mouvement (descripteurs locaux ou globaux)
- AVO naturels :
 - Arrière-plan panoramique (sprite)
 - Image (objet statique)
 - Vidéo (objet mobile)
- AVO synthétiques :
 - 2D-Mesh
 - 3D-Mesh
 - Interaction visuelle

MPEG4 Part 2 Visual

Objet VS Sprite

Il y a deux types de mode : Le mode Objet et le mode Sprite

Mode objet

- On segmente préalablement des objets ;
- On encode séparément les objets et l'image d'arrière plan pour chaque frame.



Mode sprite

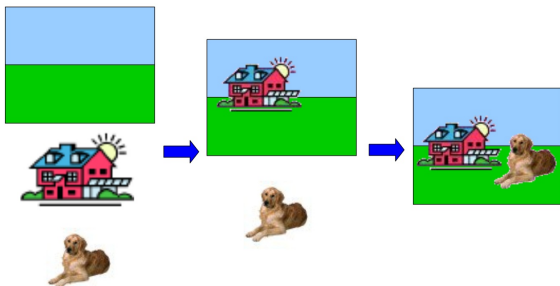
- **Image panoramique (mosaïque)** représentant l'arrière-plan ;
- On encode seulement ce qui est différent de l'a-p.



MPEG4 Part 2 Visual

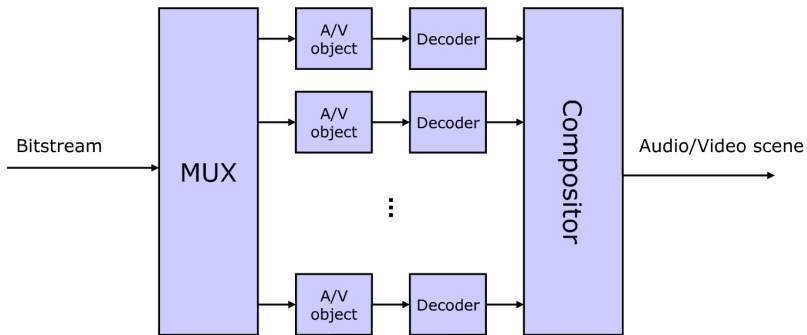
Composition de scène - mode sprite

Le décodeur peut composer différentes scènes en décidant quels objets seront décodés.

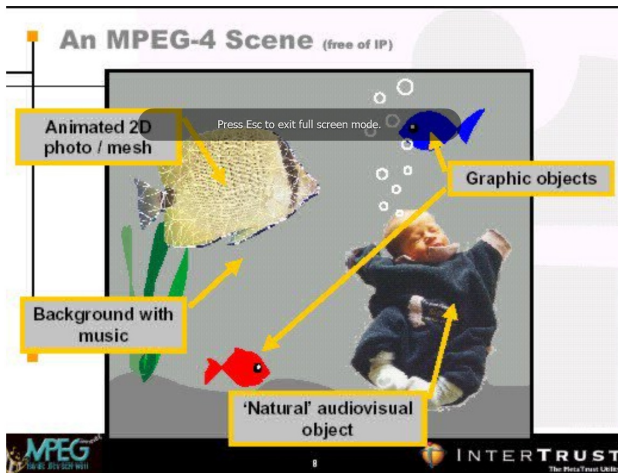


- L'image mosaïque n'est transmise qu'une seule fois au départ. Les objets qui bougent sont alors transmis pour chaque frame.
- Au décodage, on extrait cette image partie par partie pour représenter chaque frame
→ Par cropping et warping

MPEG4 Part 2 Visual



MPEG4 Part 2 Visual



Sommaire des standards

Standard	Digitisation format	Compressed rate	Example applications
H. 261	CIF/ QCIF	X 64 kbps	Video conferencing over LANs
H. 263	S-QCIF/ QCIF	<64kbps	Video conferencing over low bits rate channels
MPEG 1	SIF	<1.5Mbps	VHS quality video storage
MPEG 2			
Low	SIF	<4Mbps	VHS quality video recording
Main	4:2:0	<15Mbps	Digital video broadcasting
	4:2:2	<20Mbps	
High 1440	4:2:0	<60Mbps	High definition TV (4/3)
	4:2:2	<80Mbps	
High	4:2:0	<80Mbps	High definition TV (16/9)
	4:2:0	<100Mbps	
MPEG 4	Various	5kbps - tens Mbps	Versatile multimedia coding standard
H.264	Various	Various	Various

SIF: Standard Interchange Format, 352x240 pixels at 30 Hz.

H.264/MPEG4

- Bitrate de 64 kbps à 240 Mbps
- Transmission par câble, satellite, DSL, ...
- Video sur demande, messagerie

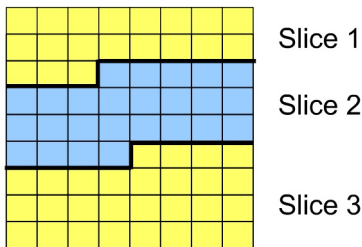
On améliore considérablement les méthodes de prédiction, compensation, compression de la vidéo :

- Encodage par tranche/couche
- Prédiction intra-image (à l'intérieur de l'image)
- Prédiction inter-image (à l'aide des autres images)
- Deux codages entropiques :
 - CAVLC - *Context-adaptive variable length coding*
 - CABAC - *Context-adaptive binary arithmetic coding*
- Encodage du ratio signal-bruit.
- Possibilité d'encodage stéréovision (3D)

H.264/MPEG4

Encodage par tranche

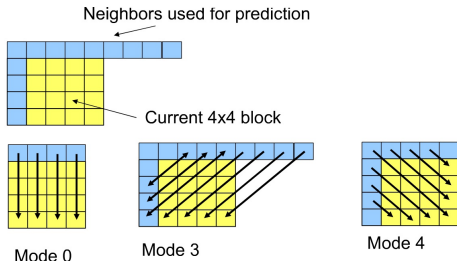
- Les tranches ont différentes grandeurs/formes
- Chaque tranche est indépendante
- Utile pour :
 - Isoler les erreurs dans des régions spécifiques ;
 - Traitement en parallèle.



H.264/MPEG4

Prédiction intra-image

- Trois modes de prédictions :
 - Brut : Régions très texturées
 - Intra 4×4 : Régions texturées, prédire chaque microbloc 4×4 (9 modes)
 - Intra 16×16 : Régions plus uniformes, prédire chaque macrobloc 16×16 (4 modes)

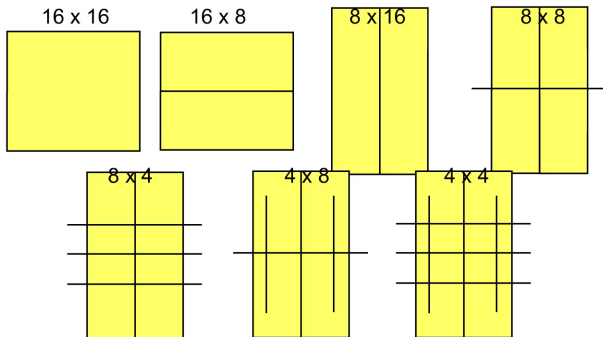


H.264/MPEG4

Prédiction inter-image

Les macroblocs de **P** peuvent être partitionnés en plus petites régions :

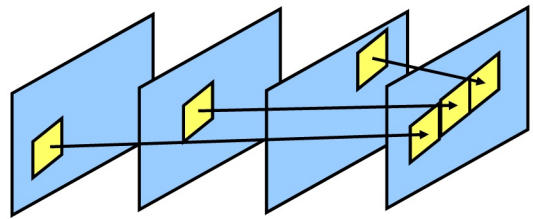
- On peut estimer jusqu'à 16 vecteurs de mouvement par macrobloc
- Chaque vecteur de mouvement est encodé différemment
- Un long processus itératif d'optimisation décide de la modélisation idéale



H.264/MPEG4

Prédiction inter-image

Plusieurs image de référence **P** et **I** sont utilisés pour la prédiction :



H.264/MPEG4

Nouveau codage entropique

Deux techniques de codage orientées pour les objets sont utilisés pour coder les coefficient des DCT :

CABAC (Context Adaptive Binary Arithmetic Coding)

- Exploite la corrélation des symboles par l'utilisation de contexte.
- Utilisation de nombre non entier de bits par symbole (codage arithmétique).

CAVLC (Context Adaptive Variable Length Coding)

- Exploite la corrélation des symboles par l'utilisation de contexte.
- Utilisation de nombre entiers de bit par symbole

H.264/MPEG4

Modélisation du contexte

On utilise un apprentissage *a priori* d'un dictionnaire de possible valeurs (contexte C).

- Chaque valeur possible possède sa distribution probabiliste $p(x|C_i)$
- Quatre différents types de contexte, où les deux principaux :
 - Utilisation des blocs voisins (spatial)
 - Utilisation des blocs précédents (temporel)

H.264/MPEG4

Résumé

<u>Outils</u>	<u>MPEG-2</u>	<u>MPEG-4 AVC</u>
Image I et P	Oui	Oui
Images B	Oui	Oui
En-tête compact d'image	Non	Oui
Compensation mouvement 4x4	Non	Oui
Compensation mouvement 8x8	Non	Oui
Prédiction des vecteurs de mouvement	Basique	Adaptative
Compensation mouvement 1/4 pixel	Non	Oui
Support videos entrelacées	Oui	Oui
Prédiction des coefficients DC	Basique	Adaptative
Prédiction des coefficients AC	Non	Adaptative
Transformée avec taille de blocs adaptative	Non	Oui
Codage efficace du quantificateur	Non	Oui
Codage arithmétique adaptatif	Non	Oui

Plan de la présentation

1 Standards de compression

- JPEG
- MPEG
- MPEG2
- H.263
- MPEG4
- H.264/MPEG4 Part 10 AVC

2 Formats conteneurs (avi, mkv, mov, mp4, ...)

Conteneurs multimédias

Un **conteneur** multimédia est une enveloppe qui englobe plusieurs médias différents, ainsi qu'un entête (meta-information) décrivant son contenu.

- Le conteneur vidéo contient des pistes audio, des séquences vidéos, des pistes de sous-titre, . . .
- Il permet de relier tous ces médias ensemble (DVD, blu-ray, ...)
- Le conteneur ne décrit pas comment ses médias sont encodés.
- Comme **le contenu** du conteneur **peut varier**, les appareils ne peuvent pas systématiquement décoder tous les conteneur d'un type donné (ex : avi, mkv)

MPEG4 part 14 (.MP4) et Quicktime (.mov)

Quicktime est un conteneur développé par Apple, qui spécialement pour les fichiers vidéos MPEG4.

MPEG part 14 est un conteneur développé par MPEG (ISO), spécialement pour les fichiers vidéos MPEG4. Il est basé directement sur Quicktime

- Permet la diffusion par internet ;
- Une piste de meta-information secondaire permet de contrôler les différents paramètres de contrôle des médias
- Video : MPEG-4 Part 10 (H.264/MPEG-4 AVC), MPEG-4 Part 2
- Si le conteneur ne possède qu'une piste audio (ACC), il aura l'extension ".m4a"
- Basé directement sur Quicktime (".mov").

Comme Quicktime est un format propriétaire, alors que MPEG4 part 14 est un format ISO, ce dernier est plus **flexible** face aux appareils électroniques.

Audio Video Interleave (.avi)

Audio Video Interleave est un conteneur introduit par Microsoft en 1992. Il est très flexible :

- Permet la diffusion par internet ;
- Navigation et recherche rapide ;
- Facilement extensible ;
- Méta-information de menu, choix de sous-titre, de piste audio ;
- Video : MPEG-4 Part 10 (H.264/MPEG-4 AVC), MPEG-4 Part 2, MPEG2, MPEG1, DivX, Xvid, ...

Comme le format est vieux, il a quelques problèmes :

- Il n'est pas possible de choisir un format de sortie flexible ;
- L'indexation temporellement des séquences vidéo n'est pas constante ;
- On ne peut faire varier le framerate ou le bitrate ;

Matroska (.mkv)

Matroska est le conteneur opensource le plus utilisé. Il peut contenir presque n'importe quel fichiers multimédia. Il règle aussi les problèmes du conteneur .avi

- Permet la diffusion par internet ;
- Navigation et recherche rapide ;
- Facilement extensible ;
- Méta-information de menu, choix de sous-titre, de piste audio ;
- Video : MPEG-4 Part 10 (H.264/MPEG-4 AVC), MPEG-4 Part 2, MPEG2, MPEG1, DivX, Xvid, ...

FIN!

