

Standards de codage vidéo

Béatrice Pesquet-Popescu

Département Traitement du Signal et des Images
TELECOM ParisTech

15 mai 2012

Plan

- 1 Introduction
- 2 Les principes de la compression vidéo
- 3 La famille MPEG
 - M-JPEG et MJ2K
 - MPEG-1
 - MPEG-2
 - MPEG-4

Plan

- 1 Introduction
- 2 Les principes de la compression vidéo
- 3 La famille MPEG
 - M-JPEG et MJ2K
 - MPEG-1
 - MPEG-2
 - MPEG-4

Motivations

Exemple 1 : Librairie de photos numériques

- Images à 1 Megapixel
- Trois composantes couleur
- Un octet par composante
- Occupation mémoire : 3 Mo par photo
- Publication sur le Web ?

Motivations

Exemple 2 : La télévision

- système analogique
 - ⇒ bande de fréquence : 6 MHz
- système numérique
 - 1 composante de luminance 576×720
 - 2 composantes de chrominance 288×360
 - quantification sur 8 bits
 - 25 images par seconde
 - $R \approx 125$ Mbps
 - ⇒ bande de fréquence ?
- 2 heures de film > 100 Go

Motivations

Exemple 3 : Musique

Son mono en qualité CD :

- Échantillonnage à 44.1 kHz
- Représentation des échantillons sur 16 bits
- Transmission en temps réel :

$$16 \text{ bit} \cdot 44100 \text{ s}^{-1} \approx 700 \text{ kbps}$$

- 1 heure de musique nécessite :

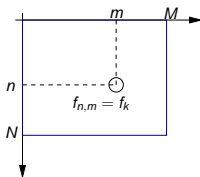
$$16 \text{ bit} \cdot 44100 \text{ s}^{-1} \cdot 3600 \text{ s} \approx 320 \text{ Mo}$$

- Stéréo : les besoins doublent !

Représentation des images numériques

Images en niveaux de gris

- Grille discrète, image $N \times M$ pixels
- A chaque pixel (m, n) , on associe un ordre de traitement k
- Généralement, balayage ligne par ligne unilatéral :
 $k = (n - 1)M + m$
- On notera indifféremment $f_{n,m}$ ou f_k



Représentation des images numériques

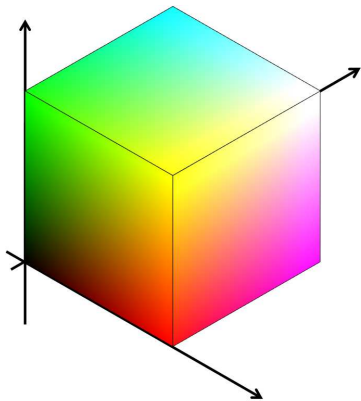
Images couleurs : Format RVB

Images en couleurs : trois composantes, chacune représentée comme une image en niveaux de gris.

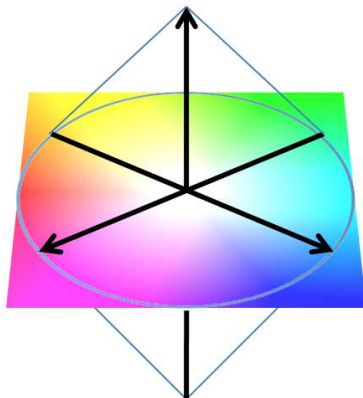


Représentation des images numériques

Espaces de couleurs



Espace RGB

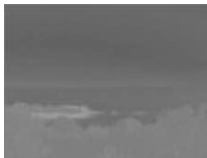


Espace HSV

Représentation des images numériques

Images couleurs : Format YUV

Images en couleurs : une composante de luminance et deux de chrominance (sous-échantillonnées).



Représentation de la vidéo numérique

- Séquence d'images numériques
- On rajoute la dépendance du temps
- Trois composantes dans le cas de la vidéo couleur
- Représentation RVB ou luminance/chrominance
- Sous-échantillonnage des composantes couleur

Fondements de la compression

POURQUOI EST-IL POSSIBLE DE COMPRIMER ?

- Redondance statistique des données
 - homogénéité des images
 - similitude entre les images successives
- Redondance psychovisuelle
 - sensibilité aux basses fréquences
 - effets de masquage
 - autres limites du système visuel humain
- Un algorithme de compression (ou codage) doit exploiter au maximum la redondance des données

Algorithmes de compression

Types d'algorithmes

- Algorithmes sans perte (*lossless*)
 - Reconstruction parfaite
 - Basés sur la redondance statistique
 - Faible taux de compression
- Algorithmes avec perte (*lossy*)
 - Image reconstruite \neq image originale
 - Basés sur la quantification
 - Redondance psychovisuelle : “visually lossless”
 - Taux de compression élevé

Algorithmes de compression

Types d'algorithmes

- Algorithmes symétriques
 - Même complexité à l'encodeur et au décodeur
 - Pas de compensation de mouvement
 - Faible taux de compression
 - Temps réel
- Algorithmes asymétriques
 - Codeur beaucoup plus complexe que le décodeur
 - Estimation/compensation du mouvement
 - Taux de compression élevé
 - Usage "off line", distribution sur supports de mémoire

Critères de performance

Débit

Rapport (taux) de compression

- $T = \frac{B_{in}}{B_{out}} = \frac{R_{in}}{R_{out}}$

Débit de codage

- Image : $R = \frac{B_{out}}{NM}$ [bpp]
- Video, son : $R = \frac{B_{out}}{T}$ [bps]

Codage d'image sans perte : $T \leq 3$

Codage d'image avec perte : $T \approx 5 \rightarrow ?$

Codage vidéo avec perte : $T \approx 20 \rightarrow ?$

Critères de performance

Qualité et distorsion

Seul le débit n'est pas suffisant pour évaluer un algorithme avec pertes

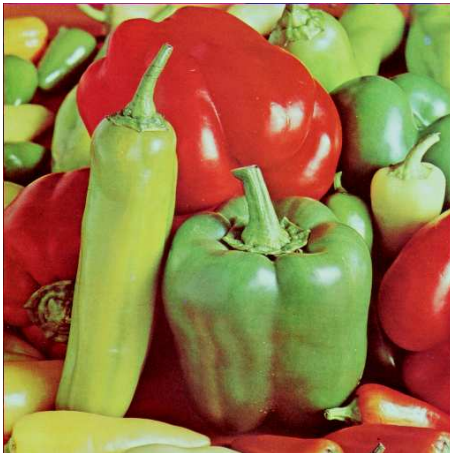
Il faut déterminer la qualité ou la distorsion de l'image reconstruite

- Les **Critères objectifs** sont des fonctions mathématiques de
 - $f_{n,m}$: l'image d'origine ; et de
 - $\tilde{f}_{n,m}$: l'image reconstruite après compression
- Critères objectifs non perceptuels
 - Erreur quadratique moyenne (EQM = MSE) :
$$\mathcal{D} = \frac{1}{NM} \sum_{n=1}^N \sum_{m=1}^M (f_{n,m} - \tilde{f}_{n,m})^2$$
 - Rapport signal sur bruit crête : $\text{PSNR} = 10 \log_{10} \left(\frac{255^2}{\mathcal{D}} \right)$
- Critères objectifs perceptuels
 - On utilise des modèles du SVH pour prendre en compte la sensibilité aux fréquences, les masquages, ...

Évaluation de la qualité

Aspects psychovisuels : Exemple

Image Originale, 24 bpp



Évaluation de la qualité

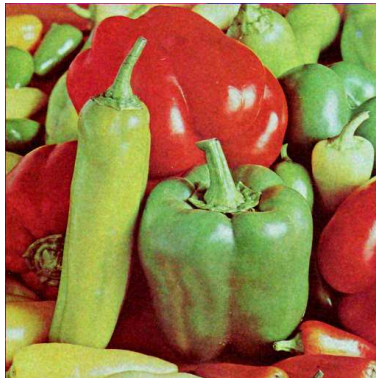
Aspects psychovisuels

Bruit blanc $\sigma = 4$

Image Originale, 24 bpp



Erreur dispersée, MSE = 16.00

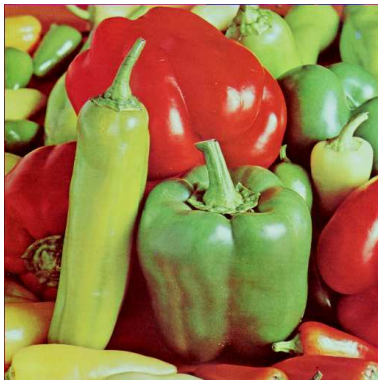


Évaluation de la qualité

Aspects psychovisuels

Bruit concentré sur 100×100 pixels

Image Originale, 24 bpp



Erreur concentrée, MSE = 16.00



Évaluation de la qualité

Aspects psychovisuels

Bruit concentré sur les contours (estimation par filtre de Sobel)

Contours de l'image



Erreur concentrée sur les contours, MSE = 16.00



Évaluation de la qualité

Aspects psychovisuels

Bruit dans les hautes fréquences spatiales

Image Originale, 24 bpp



Erreur en fréquence, MSE = 16.00

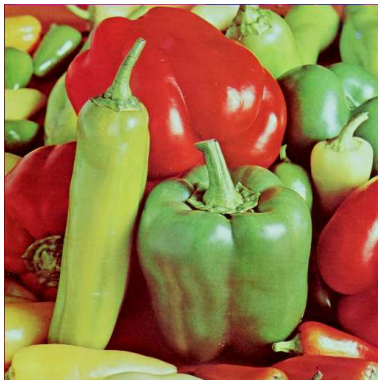


Évaluation de la qualité

Aspects psychovisuels

Sous-échantillonnage dans l'espace des couleurs

Image Originale, 24 bpp



Sous-échantillonnage en couleur, MSE = 21.27



Critères de performance

Qualité et distorsion

- Les **Critères subjectifs** sont basés sur l'évaluation de la qualité des images faite par des humaines
 - Difficulté de créer un bon modèle du SVH
 - Analyse statistique des résultats
 - Évaluations longues, difficiles et coûteuses
- En conclusion, souvent on se limite à utiliser les critères objectifs non perceptuels :
 - Simplicité
 - Interprétation géométrique (norme euclidienne)
 - Optimisation analytique
 - Relation avec la qualité perçue ?

Critères de performance

Complexité, retard et robustesse

- La complexité d'un algorithme de codage peut être limitée par :
 - contraintes liées à l'application (temps réel)
 - limites du matériel (hardware)
 - coût économique
- Le retard est normalement mesuré au codeur
 - Lié à la complexité
 - Influencé par l'ordre de codage
- Robustesse: sensibilité de l'algorithme de compression/reconstruction à des petites altérations du code comprimé (erreurs de transmission)

Critères de performance

Bilan

Besoins contradictoires :

↑↑	Qualité	↓↓	Débit
↑↑	Robustesse	↓↓	Complexité
		↓↓	Retard

Outils fondamentaux pour la compression

- Codage entropique
- Quantification
- Transformée
- Prédiction

Plan

- 1 Introduction
- 2 Les principes de la compression vidéo
- 3 La famille MPEG
 - M-JPEG et MJ2K
 - MPEG-1
 - MPEG-2
 - MPEG-4

Principes de la compression vidéo

- Redondance spatiale
 - Les régions dont les images se composent sont homogènes
- Redondance temporelle
 - Les images dont une séquence se compose sont similaires les unes aux autres
- Un codeur performant doit éliminer les deux types de redondance

Principes de la compression vidéo

Redondance spatiale



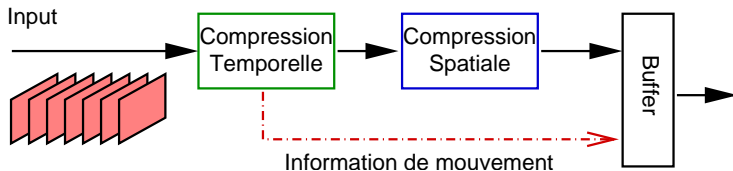
Principes de la compression vidéo

Redondance temporelle



Principes de la compression vidéo

Schéma générique d'un codeur vidéo



Classification des codeurs vidéo

	Espace	Temps	+ et -	Applications cibles	Standards
Motion Picture	Transformée	-	Simplicité d'implémentation, faible coût de calcul mais <i>faible taux de compression</i>	Cinéma numérique, chat vidéo	M-JPEG, M-JPEG2000
Hybride	Transformée	Prédiction	Très bon taux de compression, <i>complexité élevée</i>	Distribution sur support mémoire et sur réseaux	MPEG-n, H.26x
Transformée 3D	Transformée	Transformée	Très bonne scalabilité, <i>taux de compression un peu inférieur aux codeurs hybrides</i>	Distribution sur réseaux	-

La prédiction dans le codage vidéo

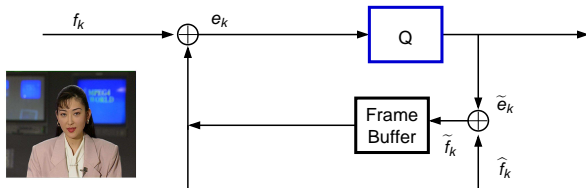
DPCM

- Les images successives se ressemblent beaucoup
- Prédiction : $\hat{f}_{n,m,k} = \tilde{f}_{n,m,k-1}$

Image courante



Erreur



La prédiction dans le codage vidéo

Conditional replenishment

- Faire la prédiction seulement si c'est utile
- Prédiction :

$$\hat{f}_{n,m,k} = \begin{cases} f_{n,m,k-1} & \text{si } |f_{n,m,k} - f_{n,m,k-1}| < \gamma \\ 0 & \text{sinon} \end{cases}$$

Problème :

- *side information* : un bit par pixel
- on préfère considérer des blocs de pixels

La prédiction dans le codage vidéo

Conditional replenishment

Mesure de ressemblance des blocs :

$$d(B_1, B_2) = \sum_{\mathbf{p}} |B_1(\mathbf{p}) - B_2(\mathbf{p})|^k$$

Si $d(B_k^{(\mathbf{p})}, B_h^{(\mathbf{p})}) < \gamma$

- *refine* : on transmet l'erreur de prédiction
- *skip* : on ne transmet aucun bit

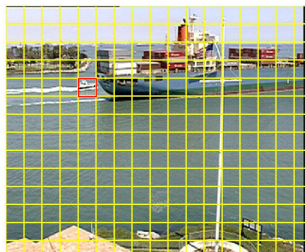
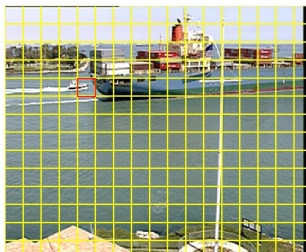
Si $d(B_k^{(\mathbf{p})}, B_h^{(\mathbf{p})}) \geq \gamma$

- *new* : on transmet le bloc de pixels

Choix de γ ? Choix de la taille des blocs ?

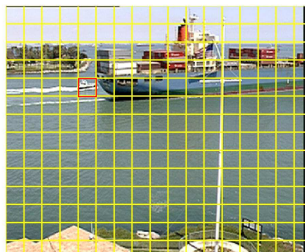
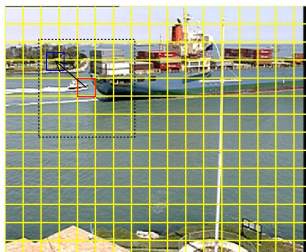
La prédiction dans le codage vidéo

Estimation du mouvement



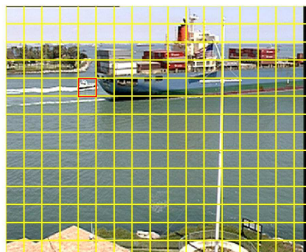
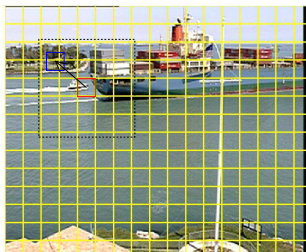
La prédiction dans le codage vidéo

Estimation du mouvement



La prédiction dans le codage vidéo

Estimation du mouvement



On compare $B_k^{(p)}$ et $B_h^{(p+v)}$

La prédiction dans le codage vidéo

Estimation du mouvement

- Test ME :

$$d(\mathbf{v}) = d(B_k^{(\mathbf{p})}, B_h^{(\mathbf{p}+\mathbf{v})})$$

- Vecteur estimé :

$$\mathbf{v}^* = \arg \min_{\mathbf{v}} d(\mathbf{v})$$

- Info transmise :

$$B_k^{(\mathbf{p})} - B_h^{(\mathbf{p}+\mathbf{v})}$$

- Au décodage on reconstruit la prédiction de $B_k^{(\mathbf{p})}$ à l'aide des vecteurs de mouvement et de l'image de référence : c'est la *compensation du mouvement*

La prédiction dans le codage vidéo

Régularisation de l'estimation du mouvement

- Dans les régions homogènes l'EM peut donner des résultats chaotiques
- On ajoute un terme de régularisation

$$J(\mathbf{v}) = d(\mathbf{v}) + \lambda r(\mathbf{v})$$

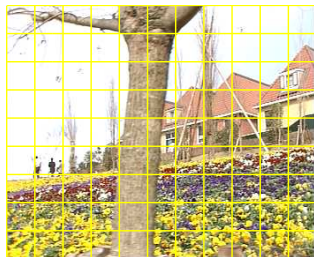
- Vecteur estimé :

$$\mathbf{v}^* = \arg \min_{\mathbf{v}} J(\mathbf{v})$$

- λ gère le compromis entre fidélité et régularité
- $r(\mathbf{v})$: coût de codage ; régularité géométrique ...

La prédiction dans le codage vidéo

Exemple d'estimation de mouvement



La prédiction dans le codage vidéo

Vecteurs estimés

Débit MVs : 1.675 Kbits Qualité prédiction : 21.69 dB



MVF non régularisé

La prédiction dans le codage vidéo

Vecteurs estimés

Débit MVs : 1.675 Kbits Qualité prédiction : 21.69 dB



MVF non régularisé

Débit MVs : 1.430 Kbits Qualité prédiction : 21.69 dB



MVF régularisé

La prédiction dans le codage vidéo

Estimation du mouvement

Compromis complexité/efficacité

Soit n le côté de la fenêtre de recherche

- Méthode *full search* : Toutes les n^2 positions sont contrôlées
- Méthode *cross search* : On contrôle d'abord le déplacement horizontal, ensuite le vertical ; $2n$ positions sont contrôlées
- Méthode *log search* : On contrôle 8 position à distance $2^m - 1$; on choisit la direction et on continue avec un pas de $2^{m-1} - 1$ pixels $\approx 8 \log_2 n$ positions sont contrôlées
- Méthode *diamond search* : On contrôle 4 directions, mais on réduit le pas seulement quand on a choisi le centre ; cette méthode est très populaire

Robustesse aux erreurs

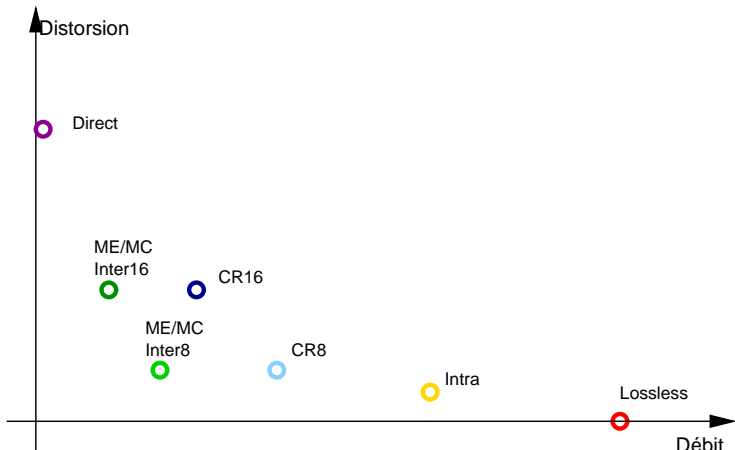
- Mitiger la propagation des erreurs
 - Protection inégale (en-têtes mieux protégés)
 - Limitation aux modes prédictifs
- Rendre facile la récupération de la synchronisation
 - Utilisation des marqueurs de synchronisation
 - *Marker emulation prevention*
- Limiter l'impact visuel des erreurs
 - Reconnaître les images erronées
 - Afficher une image précédente plutôt que une image erronée

Le codage hybride

- Codage par macroblocs
- Modes de codage
 - Intra:** Pas de prédiction temporelle, codage par transformée
 - Inter:** ME/MC pour la prédiction temporelle, codage par transformée
 - Direct:** Vecteur de mouvement déduit des voisins ; copie du bloc de référence
 - Lossless:** Codage sans perte

Codeurs hybrides

Exemple des performances des modes



Codeurs hybrides

Choix du mode de codage

- Objectif : minimiser D pour un R donné :

$$D = \sum_{k=1}^K D_k(i_k, Q) \quad R = \sum_{k=1}^K R_k(i_k, Q)$$

- Le pas de quantification Q est donné
- L'ensemble des modes $\mathbf{i} = \{i_k\}_{k=1}^K$ doit être choisi de façon à minimiser :

$$J(\mathbf{i}, Q, \lambda) = \sum_{k=1}^K D_k(i_k, Q) + \lambda \sum_{k=1}^K R_k(i_k, Q)$$

Codeurs hybrides

Choix du mode de codage

- La minimisation conjointe sur i est trop complexe
- On préfère une minimisation sous-optimale
- Pour chaque MB k , on choisit le mode de façon à minimiser :

$$J_k(i_k, Q, \lambda) = D_k(i_k, Q) + \lambda R_k(i_k, Q)$$

- On minimise séparément chaque terme de la somme J
- Le mode choisi dépend donc de Q et λ

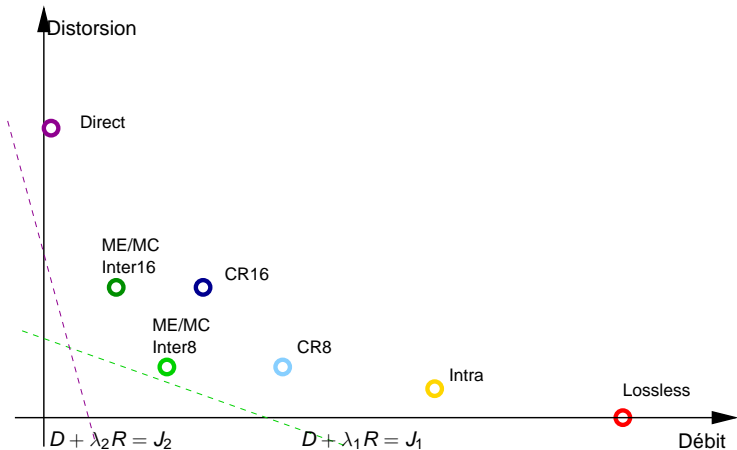
Codeurs hybrides

Choix du mode de codage

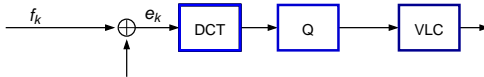
- Le pas de quantification Q est considéré comme un *input*
- Pour chaque Q (débit) il existe une valeur optimale de λ , déterminée empiriquement
 - MPEG-2 : $\lambda = aQ^2 + b$
 - H.264 : $\lambda = c2^{dQ+e}$
- Avec le λ donné, minimiser J_k revient à trouver une droite dans le plan RD

Codeurs hybrides

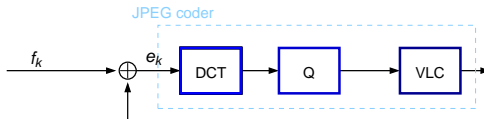
Exemple de performances des modes



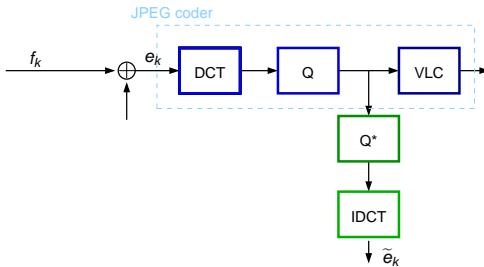
Le codeur hybride



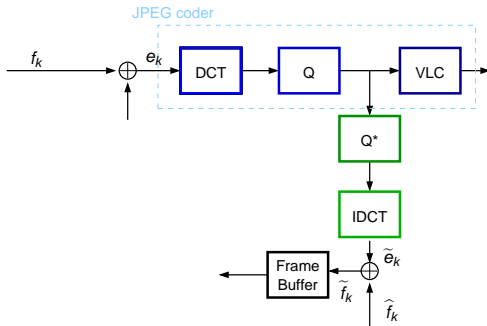
Le codeur hybride



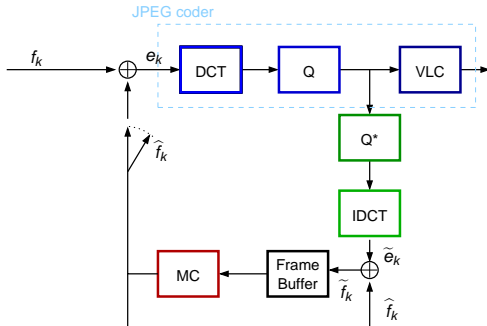
Le codeur hybride



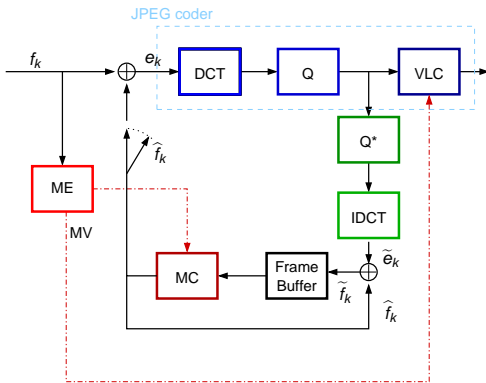
Le codeur hybride



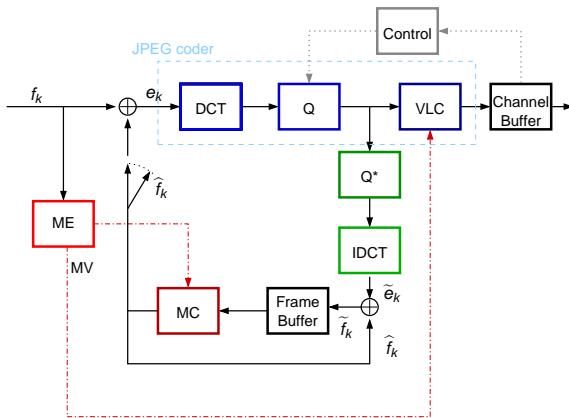
Le codeur hybride



Le codeur hybride

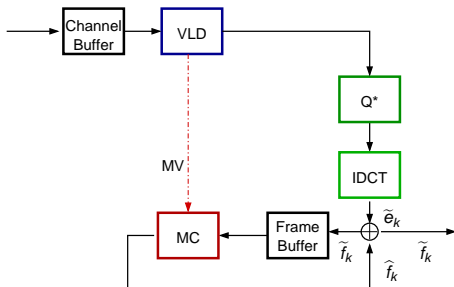


Le codeur hybride



Le décodeur hybride

Schéma asymétrique !



Plan

- 1 Introduction
- 2 Les principes de la compression vidéo
- 3 La famille MPEG**
 - M-JPEG et MJ2K
 - MPEG-1
 - MPEG-2
 - MPEG-4

M-JPEG et MJ2K

- Pas de ME/MC : faible complexité, faible taux de compression
- M-JPEG : chaque image codée avec JPEG
 - Ce n'est pas un standard
 - Utilisé pour le mode vidéo des appareils photo-numériques
- M-JP2K : chaque image codée avec JPEG2000
 - Partie 3 du standard JPEG2000
 - Utilisé pour le cinéma numérique et la communication vidéo en temps réel (video chat)
 - Meilleure qualité par rapport à M-JPEG
 - Scalabilité

Les standards vidéo

Groupes de normalisation

Organismes de normalisation :

ISO International Standardization Organization

IEC International Electrotechnical Commission

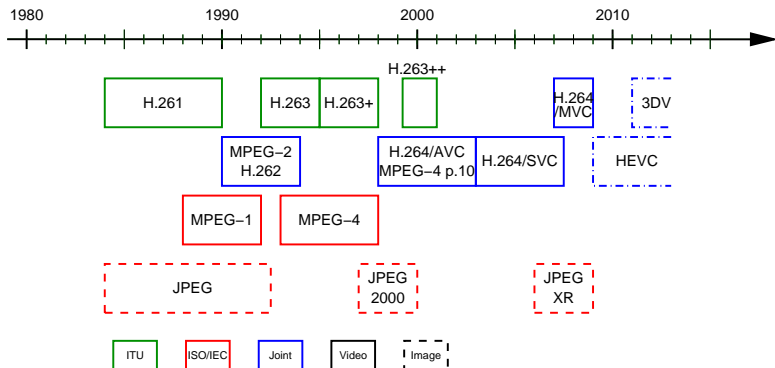
ITU International Telecommunication Union

Groupes de travail

- MPEG (1988) : ISO/IEC Moving Picture Expert Group
- VCEG (1997) : ITU Video coding Expert Group
- *Joint Video Team*: H.264 et MPEG-4/Part 10 (JVT);
extension scalable de H.264 (SVC)

Les standards vidéo

Développement des standards



Le standard MPEG-1

- Développé entre 1988-1992
- Parties du standard
 - 1 Systèmes
 - 2 Vidéo
 - 3 Audio
 - 4 Conformance test
 - 5 Software simulation

Le standard MPEG-1

Partie 2 (Vidéo)

- Codeur hybride avec ME/MC
- Entrée : max 720×576 pixel @ 30 fps
- Débit ≤ 1.86 Mbps (qualité VHS)
- Applications asymétriques, VoD, CD vidéo, jeux vidéo

Nouveautés techniques

- Types d'images
- ME/MC à précision sous-pixellique

Le standard MPEG-1

Types d'images

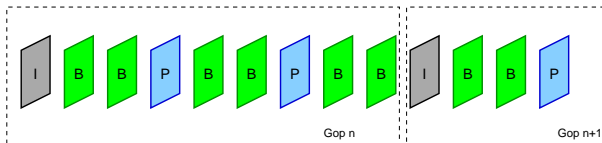
- Trames I (Intra coded)
- Trames P (Prédictives)
- Trames B (Bi-directionnelles)
- Trames D (DC coded)

Trames I et P : Anchor Frames

Le standard MPEG-1

Group of Pictures

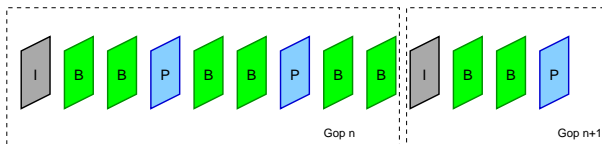
- Trames organisées en GOP
- Première image : Intra
- Structure :
 - intervalle entre I
 - intervalle entre AF



Le standard MPEG-1

Trames I

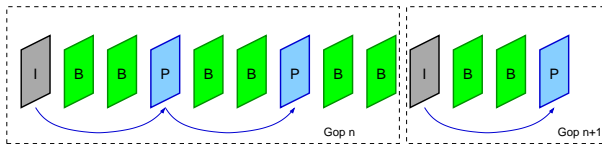
- Codée indépendamment des autres
- Codage JPEG
- Faible complexité, faible taux de codage
- Utilisée pour :
 - Fast forwards
 - Random access
 - Robustesse aux erreurs



Le standard MPEG-1

Trames P

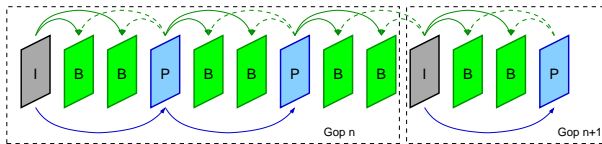
- Prédite de l'AF précédente
- Complexité élevée (ME)
- Taux de compression élevé



Le standard MPEG-1

Trames B

- Prédite des AFs précédente et successive
- Complexité très élevée (double ME)
- Taux de compression élevé

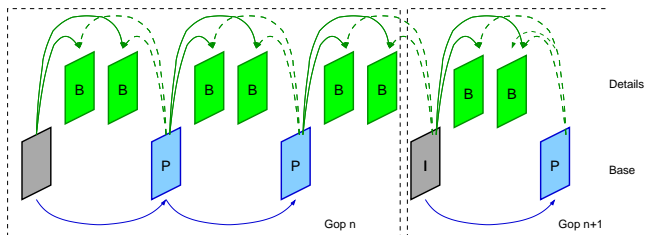


Standard MPEG-1

Ordre de codage des trames

I → AF → Trames B → AF → Trames B ...

Retard ?



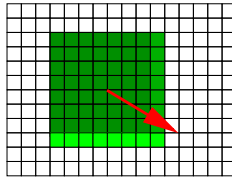
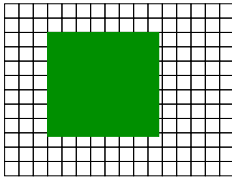
Standard MPEG-1

ME/MC à précision sous-pixellique

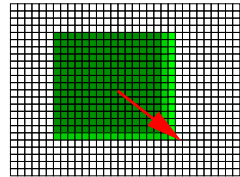
- Le mouvement ne correspond pas à la grille des pixels
- Interpolation pour améliorer la précision
- Augmentation de la complexité
- Très bonnes performances de codage

Standard MPEG-1

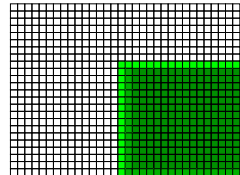
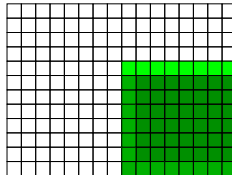
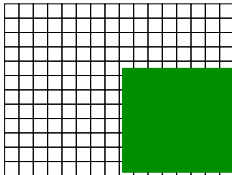
ME/MC à précision sous-pixellique



$v=(5,3)$



$v=(5.5,2.5)$



Standard MPEG-1

Résumé

Codeur hybride classique (H.261), plus :

- 1 Trames B et D
- 2 Vecteurs à précision sous-pixellique
- 3 Gamme de résolutions étendue
- 4 Groupe d'images avec structure flexible

Le standard MPEG-2

- Développé entre 1990-1994
- Parties du standard
 - 1 Systèmes
 - 2 Vidéo
 - 3 Audio
 - 4 Conformance test
 - 5 Software simulation
- Les parties 6-9 spécifient certaines fonctionnalités secondaires
- MPEG-3, à l'origine prévu pour l'HDTV, a été fusionné avec MPEG-2
- *Équivalent* à la norme H.262

Standard MPEG-2

- Codeur hybride
- Débit ≤ 15 Mbps (HDTV)
- Profils et niveaux
- Support pour la vidéo entrelacée
- Support pour la scalabilité

Standard MPEG-2

Profils et niveaux

Il est possible de mettre en oeuvre une partie des fonctionnalités (profils) pour une certaine gamme de paramètres (niveaux), tout en restant compatible avec le standard

Niveau	width [pixel]	height [pixel]	frame rate [frame/s]	bit rate [Mbps]
Low	352	288	30	4
Main	720	576	30	15
High-1440	1440	1152	60	60
High	1920	1152	60	80

Profil	fonctionnalités
Simple	Pas de scalabilité; vidéo entrelacée; pas de trames B
Main	Simple + Trames B
SNR scalable	Main + Deux/trois niveaux de scalabilité en qualité
Spatial scalable	SNR + Deux/trois niveaux de scalabilité en résolution
High	Space + Chroma sur-échantillonnée

Standard MPEG-2

Profils et niveaux

Seulement certaines couples profil/niveau sont admissibles

Niveau	Profil				
	Simple	Main	SNR scalable	Spatial scalable	High
Low		•	•		
Main	•	•	•		•
High-1440		•		•	•
High		•			•

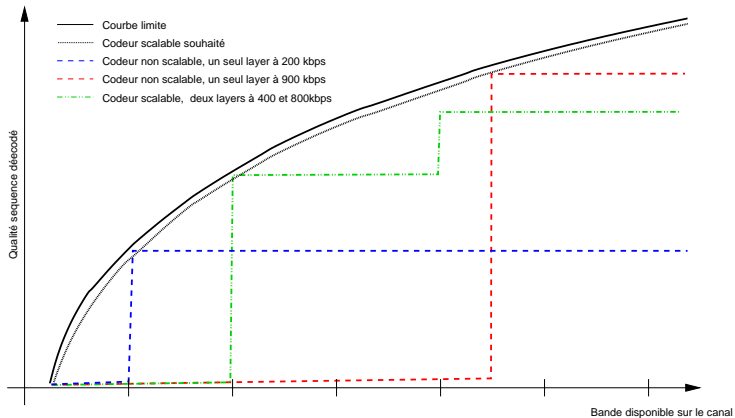
Standard MPEG-2

Scalabilité

Encode once, decode many!

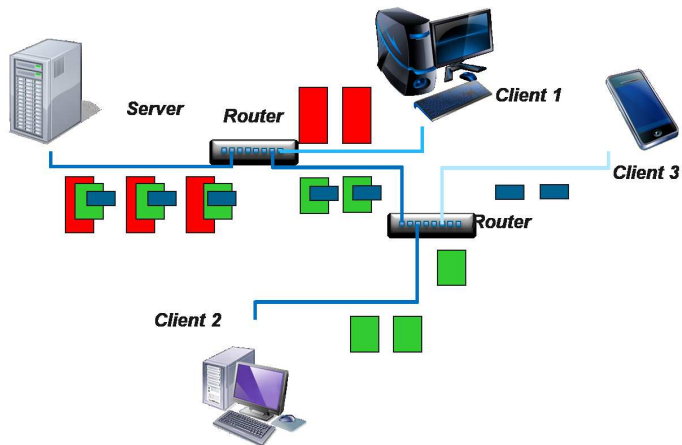
- Train binaire formé par :
 - Un flux *base*
 - Des *détails*
- Le flux de base peut être décodé seul
- Les détails augmentent la qualité du flux de base
- Les détails seuls sont inutiles
- Selon ses nécessités, le client peut demander le flux de base ou les 2 flux

Codeurs scalables et non scalables



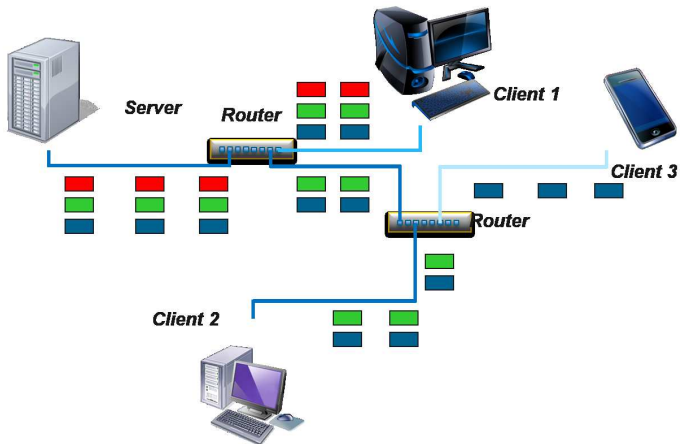
Scalabilité : exemple

Distribution de la vidéo sans scalabilité



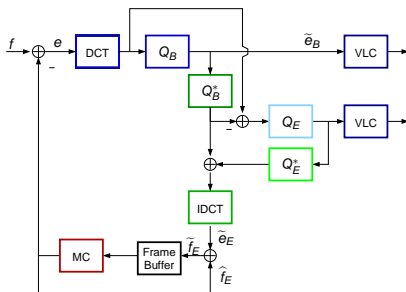
Scalabilité : exemple

Distribution de la vidéo avec scalabilité



Standard MPEG-2

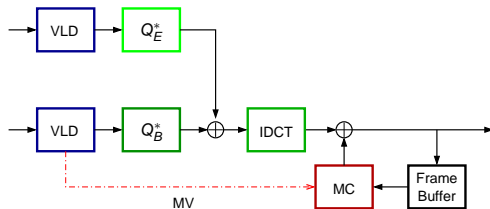
Scalabilité SNR : codeur



- Raffinement des coefficients DCT
- *Drift* : décodeur au niveau de base désynchronisé
- Bonne qualité du niveau d'amélioration
- Qualité du niveau de base pas toujours satisfaisante

Standard MPEG-2

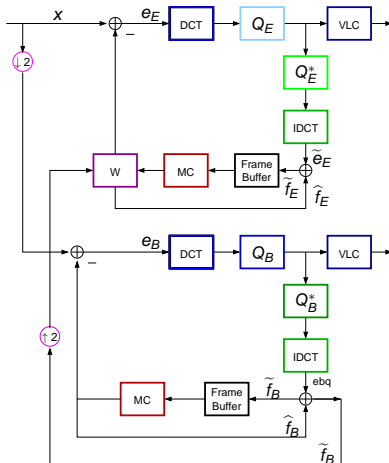
Scalabilité SNR : décodeur



- Pas de contrôle du drift
- Un seul champ de vecteurs de mouvement

Standard MPEG-2

Scalabilité en résolution : codeur



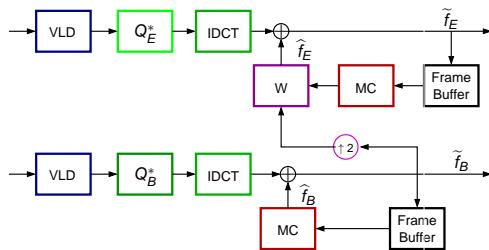
Standard MPEG-2

Scalabilité en résolution : codeur

- Double boucle : pas de drift
- La vidéo d'entrée est filtrée et sous-échantillonnée
- La prédiction du niveau d'amélioration est la somme pondérée de :
 - L'image du niveau de base interpolée
 - La prédiction ME/MC
- Le choix du poids est fait MB par MB et cette info est codée dans le train binaire

Standard MPEG-2

Scalabilité en résolution : décodeur



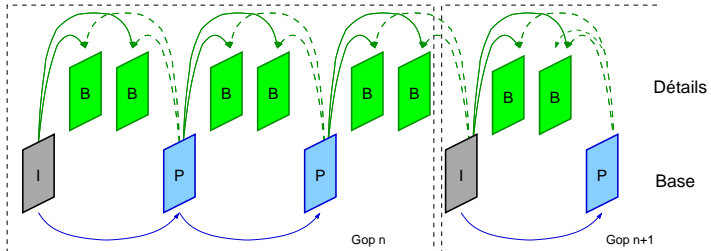
Standard MPEG-2

Scalabilité en résolution : décodeur

- L'image à basse résolution est interpolée et pondérée
- Interpolation linéaire
- On obtient les informations nécessaires pour le décodage du niveau d'amélioration

Standard MPEG-2

Scalabilité temporelle



Standard MPEG-2

Scalabilité hybride, 1/2

- Les formes de scalabilité disponibles peuvent être combinées entre elles, jusqu'à 3 niveaux au total
- Trois types de scalabilité hybride sont possibles
- SNR + spatiale
 - 1 SDTV/CIF, basse qualité
 - 2 HDTV/SDTV, basse qualité
 - 3 HDTV/SDTV, haute qualité

Standard MPEG-2

Scalabilité hybride, 2/2

- spatiale + temporelle
 - 1 SDTV entrelacée
 - 2 HDTV entrelacée
 - 3 HDTV progressive
- SNR + temporelle
 - 1 HDTV entrelacée, basse qualité
 - 2 HDTV entrelacée, haute qualité
 - 3 HDTV progressive, haute qualité

Standard MPEG-2

Data partitioning

- Outil pour la robustesse aux erreurs
- Les blocs de 64 coefficients DCT sont divisés en sous-ensembles (DC+basses fréquences, hautes fréquences)
- Le sous-ensemble plus critique peut être envoyé sur un canal plus robuste (Unequal error protection)
- Il est nécessaire que le décodeur supporte le Data partitioning (pas de *backward compatibility*)

Le standard MPEG-4

- Développé entre 1993-1998
- Parties du standard
 - 5 parties principales (comme MPEG-1 et 2)
 - 18 parties supplémentaires, couvrant une variété de problèmes
 - P.e. MPEG4/part 10 coïncide avec H.264/AVC

Standard MPEG-4

Fonctionnalités

- Codeur hybride
- Interactivité
 - Manipulation du train binaire basée sur le contenu et sans transcodage
 - Codage hybride de données naturelles et synthétiques
 - Accès aléatoire amélioré
- Compression
 - Performance de compression améliorée
 - Codage du flux données en parallèle (stéréo TV)
- Accès universel
 - Robustesse en environnement bruité
 - Scalabilité basée objet

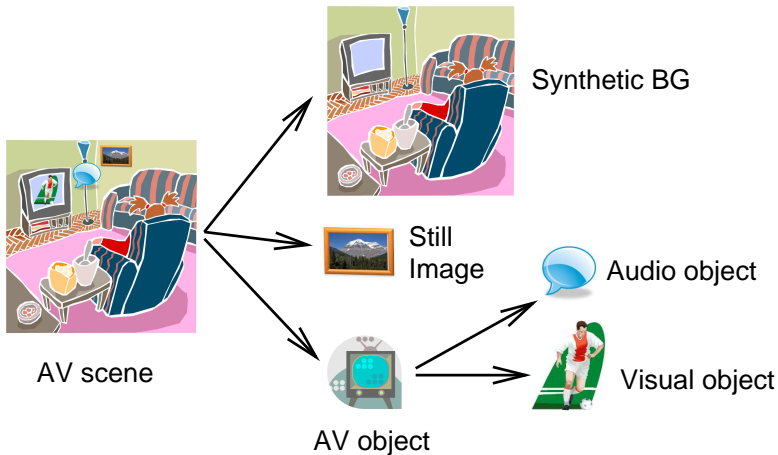
Standard MPEG-4

Représentation basée objet

- Une représentation basée objet est nécessaire pour atteindre les objectifs fixés
- Audiovisual object (AVO)
 - Différents AVOs codés dans des trains binaires différents
 - Composé d'une partie audio (mono, stéréo, synthétique, ...) et/ou une vidéo (naturelle, synthétique, ...)
- Plusieurs AVOs composants une *AV scene*
- MPEG-4 définit la syntaxe pour la description de la scène

Standard MPEG-4

Audiovisual scene



Standard MPEG-4

Visual coding

MPEG-4 fournit 4 types d'instruments de codage

Video object coding : codage d'un objet de forme arbitraire, naturel ou synthétique

Mesh object coding : codage d'un objet représenté par une grille de points

Model-based coding : animation du visage et du corps humain

Still texture coding : codage d'images fixes (*background*)

Standard MPEG-4

Video object coding

Un *video object* (VO) est une succession de *video object planes* (VOP) constitués par :

- Mouvement
- Texture
- Contours (Forme)

Si l'objet a une forme rectangulaire, le codage est pratiquement MPEG-2, car les objets sont divisés en macroblocs, codés ensuite par un codeur hybride

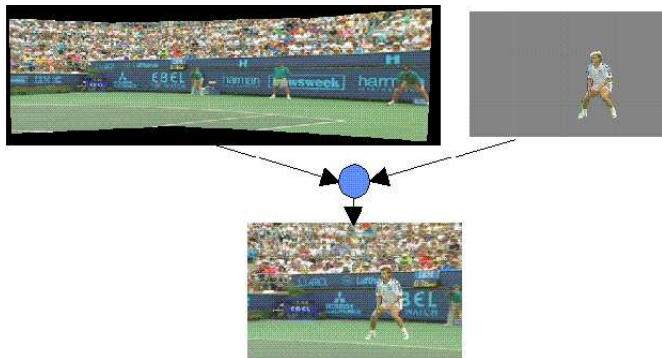
Standard MPEG-4

Video object coding

- Le mouvement est codé par prédiction des vecteurs voisins
- Les vecteurs peuvent sortir de l'image et ont une précision double par rapport à MPEG-2 (1/4 de pixel)
- La texture est codée en mode Inter (prédiction + DCT/Q de l'erreur) ou Intra (TCD ou transformée adaptée à la forme)
- Forme codée par un *alpha-plane*
 - Codage simple et efficace d'une image "en niveaux de gris"
 - Codage lossless (codage arithmétique avec prédiction temporelle en Intra) et lossy (sous-échantillonnage et codage lossless) de l'alpha-plane
- Sprite coding : représentation de l'arrière-plan

Standard MPEG-4

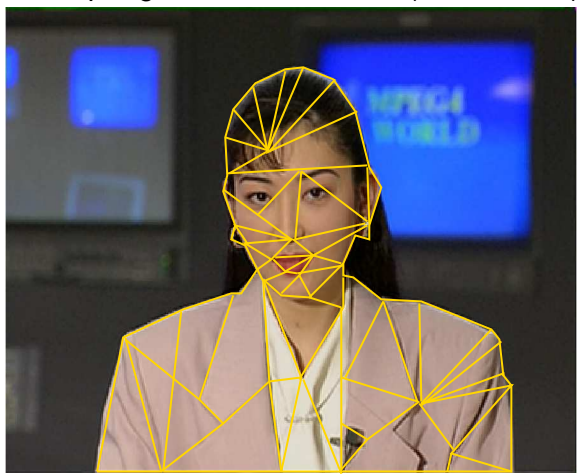
Codage par sprites et VO



Standard MPEG-4

Mesh object coding

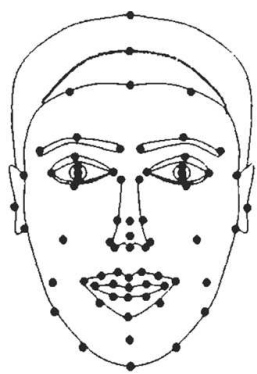
Représentation par grille du mouvement (déformation)



Standard MPEG-4

Model-based coding

Animation des points de contrôle pour représenter les émotions



tongue



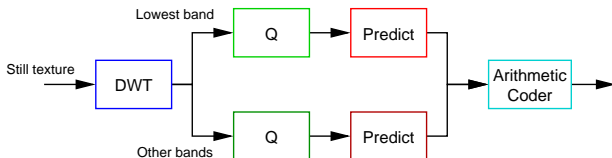
teeth



Standard MPEG-4

Still texture coding

- TO Daubechies ou TO adaptée à la forme
- Codage DPCM pour les coefficients LL
- Algorithme de codage basée sur les arbres de zéros pour les coefficients hautes fréquences
 - Exploite la corrélation *inter*-bande
 - Faible complexité et excellente compression
- Codage arithmétique adaptatif
- Scalabilité en qualité et en résolution



Standard MPEG-4

Scalabilité

- Frame-rate et résolution : comme en MPEG-2
- Qualité : *fine grain scalability* grâce au codage par plans de bits
- Objet : la scène peut être composée en utilisant un sous-ensemble des AVOs disponibles

Standard MPEG-4

Robustesse aux erreurs

La robustesse aux erreurs est obtenue à l'aide de trois types de techniques :

- Resynchronisation (resynchronization)
- Partition des données (data partitioning)
- Récupération des données (data recovery)

Standard MPEG-4

Robustesse aux erreurs

Resynchronisation

- Nécessaire pour recommencer le décodage lors d'une erreur
- Aide à la récupération des données et à la dissimulation des erreurs (error concealment)
- Marqueurs de re-synchronisation
- Information supplémentaire pour chaque paquet (numéro de MB et quantificateur)

Standard MPEG-4

Robustesse aux erreurs

Partition des données

- Séparation entre mouvement et texture
- Marqueurs supplémentaires
- Si la texture est corrompue, on peut utiliser le mouvement pour cacher l'erreur (copie du bloc de l'image de référence)

Récupération des données

- Technique de codage sans perte moins performante mais plus robuste
- Décodage possible "en avant" et "en arrière" (RVLC)

Standard MPEG-4

Profils et niveaux

Comme en MPEG-2, il est possible de limiter les fonctionnalités (profils) pour des plusieurs résolutions (niveaux)

- ➊ **Simple Visual Profile** : codage efficace pour AVOs rectangulaires. Adapté aux réseaux mobiles
- ➋ **Simple Scalable Visual Profile** : rajoute le support à la scalabilité spatiale et temporelle
- ➌ **Core Visual Profile** : rajoute le codage d'objets de forme arbitraire
- ➍ **Main Visual Profile** : codage entrelacé, sprite, codage du plan alpha
- ➎ **N-Bit Visual Profile** : pixel-depth de 4 à 12 bits

D'autres profils ont été ajoutés aux versions plus récentes du standard

Conclusions

Pour approfondir

- 1 A. Bovik. *Handbook of image and video processing*
- 2 I. Richardson. *H.264 and MPEG-4 Video Compression*
- 3 Special issues de *IEEE Trans. Circ. Video Tech.*

Conclusions

MERCI DE VOTRE ATTENTION !

Questions ?

Contact : cagnazzo@telecom-paristech.fr