

TELECOM
ParisTech



Institut
Mines-Telecom

Video coding principles

F. Dufaux, M. Cagnazzo

Σ 207



Plan

La prédiction temporelle

La structure du GOP (Group of Pictures)

Le codeur hybride

La norme MPEG-1

Plan

La prédiction temporelle

La structure du GOP (Group of Pictures)

Le codeur hybride

La norme MPEG-1

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 performante doit éliminer les deux types de redondance

Principes de la compression vidéo

Redondance spatiale



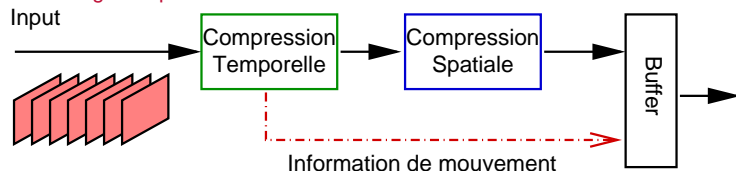
Principes de la compression vidéo

Redondance temporelle



Principes de la compression vidéo

Schéma générique d'un codeur vidéo



Compression spatiale et temporelle

- ▶ Les principes de la compression spatiale sont ceux de la compression d'image fixe
 - ▶ Transformée, quantification, codage sans pertes
 - ▶ La prédiction spatiale s'ajoute à cela
 - ▶ Comme l'on verra, ces techniques s'appliquent aussi bien à des images (Images "Intra") qu'à des erreurs de prédiction temporelle
- ▶ La compression spatiale est ce qui caractérise la compression vidéo
 - ▶ Même si des méthodes par transformée temporelle existent, l'approche la plus populaire et performante est basée sur la prédiction temporelle
 - ▶ On parle de codeur hybride (transformée pour la compression spatiale, prédiction pour la compression temporelle)

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

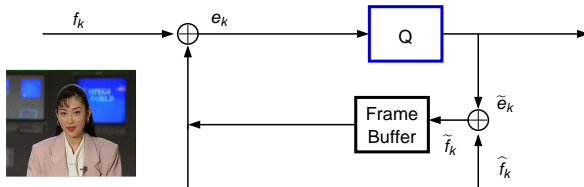


Image précédente

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 pixel

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

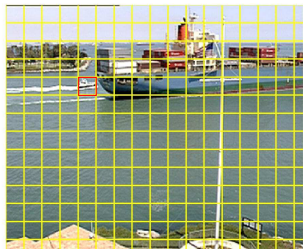
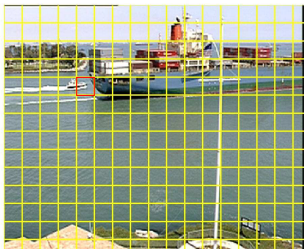
- ▶ *refine* : on transmet l'erreur de prédiction
- ▶ *skip* : on ne transmet aucun bit

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

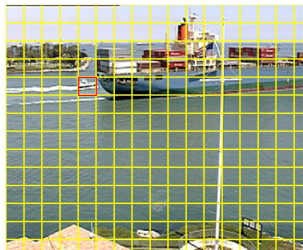
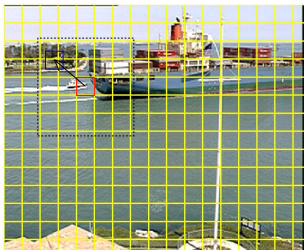
- ▶ *new* : on transmet le bloc de pixels

Choix de γ ? Choix de la taille des blocs ?

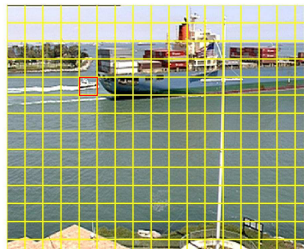
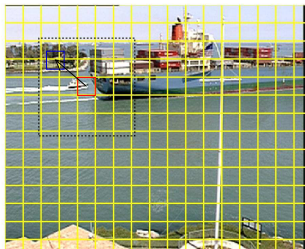
Estimation du mouvement



Estimation du mouvement



Estimation du mouvement



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

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*

Estimation du mouvement

Fonctions de coût

Plusieurs choix sont possibles pour la fonction $d(\cdot, \cdot)$:

- ▶ SAD (Sum of Absolute Differences)

$$d(B_1, B_2) = \sum_{n,m} |B_1(n, m) - B_2(n, m)|$$

- ▶ SSD (Sum of Squared Differences)

$$d(B_1, B_2) = \sum_{n,m} [B_1(n, m) - B_2(n, m)]^2$$

- ▶ ZN-SSD (Zero-mean Normalized SSD)

$$d(B_1, B_2) = \frac{\sum_{n,m} [\bar{B}_1(n, m) - \bar{B}_2(n, m)]^2}{\sum_{n,m} \bar{B}_1^2(n, m) \sum_{n,m} \bar{B}_2^2(n, m)}$$

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 une 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



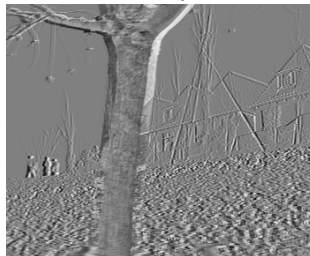
La prédiction dans le codage vidéo

Exemple d'estimation de mouvement

Current image

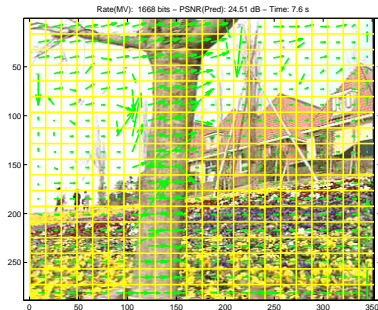


Difference image



La prédiction dans le codage vidéo

Vecteurs estimés

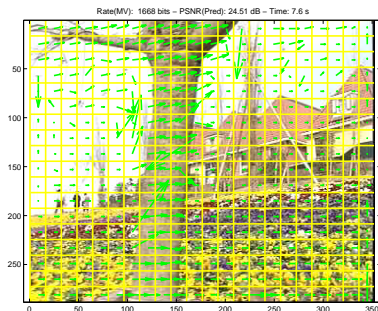


MVF non régularisé

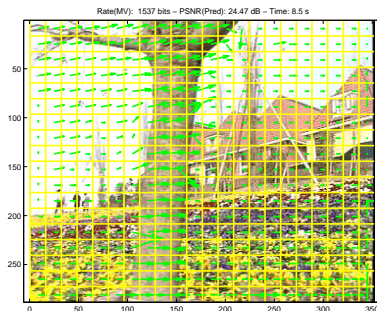
MVF régularisé

La prédiction dans le codage vidéo

Vecteurs estimés



MVF non régularisé



MVF régularisé

La prédiction dans le codage vidéo

Vecteurs estimés

Motion-compensated image



MVF régularisé, image compensée

MVF régularisé, erreur de compensation

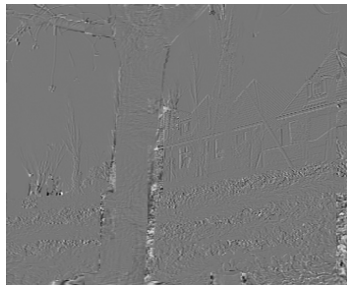
La prédiction dans le codage vidéo

Vecteurs estimés

Motion-compensated image



MC error

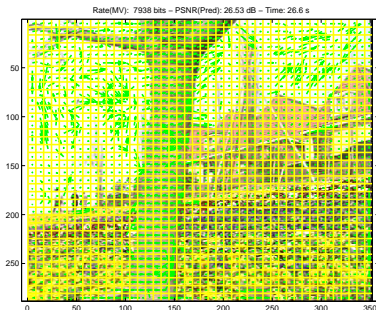


MVF régularisé, image compensée

MVF régularisé, erreur de compensation

La prédiction dans le codage vidéo

Vecteurs estimés

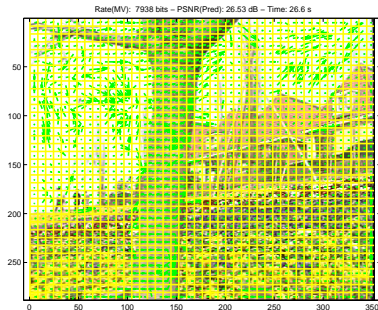


MVF non régularisé

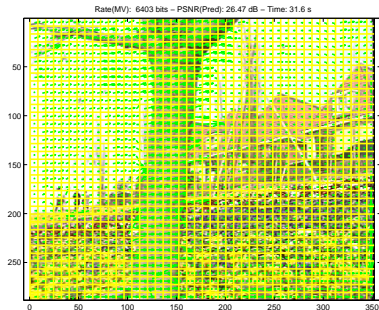
MVF régularisé

La prédiction dans le codage vidéo

Vecteurs estimés



MVF non régularisé



MVF régularisé

La prédiction dans le codage vidéo

Vecteurs estimés

Motion-compensated image



MVF régularisé, image compensée

MVF régularisé, erreur de compensation

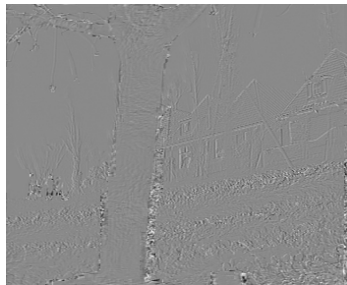
La prédiction dans le codage vidéo

Vecteurs estimés

Motion-compensated image



MC error

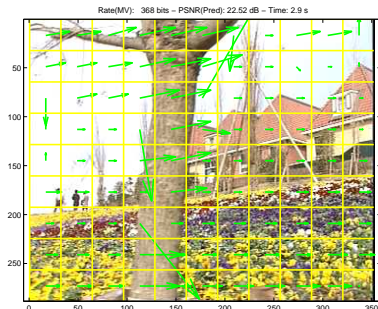


MVF régularisé, image compensée

MVF régularisé, erreur de compensation

La prédiction dans le codage vidéo

Vecteurs estimés

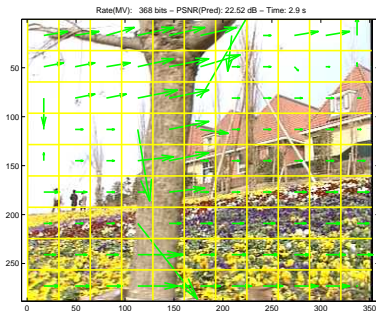


MVF non régularisé

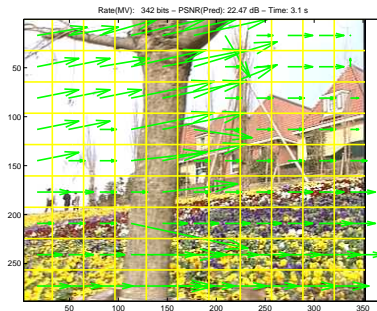
MVF régularisé

La prédiction dans le codage vidéo

Vecteurs estimés



MVF non régularisé



MVF régularisé

La prédiction dans le codage vidéo

Vecteurs estimés

Motion-compensated image



MVF régularisé, image compensée

MVF régularisé, erreur de compensation

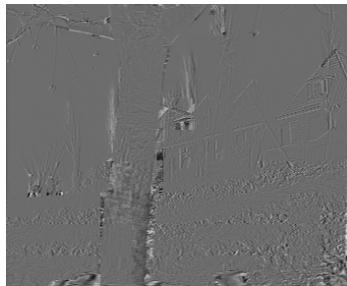
La prédiction dans le codage vidéo

Vecteurs estimés

Motion-compensated image



MC error



MVF régularisé, image compensée

MVF régularisé, erreur de compensation

La prédiction dans le codage vidéo

Exemple d'estimation de mouvement

Reference image



Current image



La prédiction dans le codage vidéo

Exemple d'estimation de mouvement

Current image

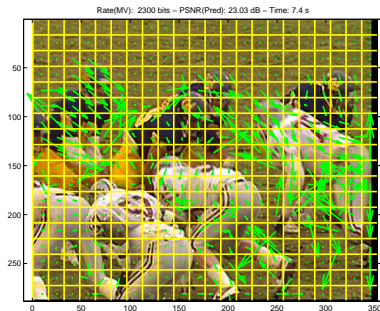


Difference image



La prédiction dans le codage vidéo

Vecteurs estimés

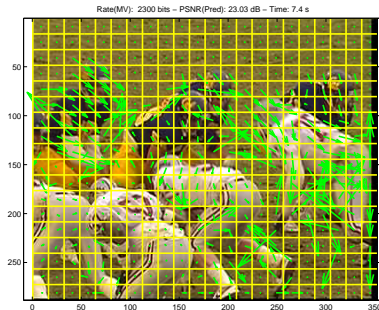


MVF non régularisé

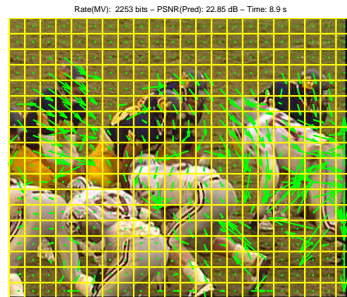
MVF régularisé

La prédiction dans le codage vidéo

Vecteurs estimés



MVF non régularisé



MVF régularisé

La prédiction dans le codage vidéo

Vecteurs estimés

Motion-compensated image



MVF régularisé, image compensée

MVF régularisé, erreur de compensation

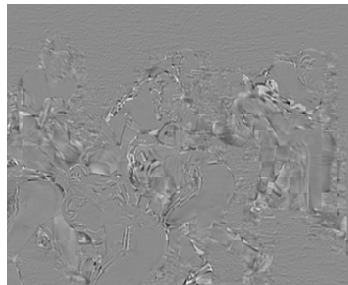
La prédiction dans le codage vidéo

Vecteurs estimés

Motion-compensated image



MC error



MVF régularisé, image compensée

MVF régularisé, erreur de compensation

Estimation du mouvement

Stratégie de recherche : compromis complexité/efficacité

Soit n le coté de la fenêtre de recherche

- ▶ Méthode *full search* : Toutes les n^2 position sont contrôlées
- ▶ Méthode *cross search* : On contrôle d'abord le déplacement horizontal, en suite le vertical ; $2n$ positions sont contrôlées
- ▶ Méthode *log search* : On contrôle 8 position à distance $2^m - 1$; on choisi 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

Estimation du mouvement

Bilan

- ▶ Méthode très efficace de prédiction temporelle
- ▶ Utilisée dans la plus part des codeurs vidéo
- ▶ Compromis : précision - coût de codage - complexité
- ▶ Degrés de liberté :
 - ▶ Fonction de coût (SAD, SSD, régularisation, ...)
 - ▶ Modèle de mouvement (forme et taille des blocs, fenêtre de recherche, ...)
 - ▶ Stratégie de recherche (Full-search, Log, Diamond, ...)

Plan

La prédiction temporelle

La structure du GOP (Group of Pictures)

Le codeur hybride

La norme MPEG-1

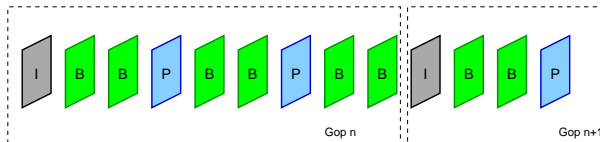
Types d'images

- ▶ Trames I (Intra coded)
- ▶ Trames P (Prédictive)
- ▶ Trames B (Bi-directionnelles)

Trames I et P : Anchor Frames

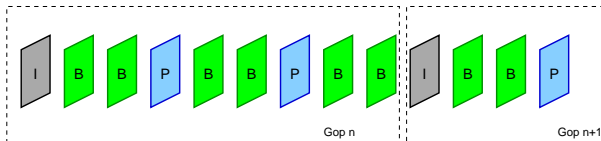
Group of Pictures

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



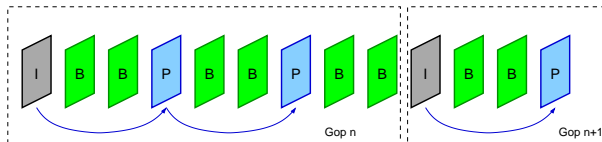
Trames I

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



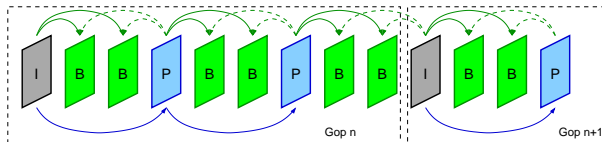
Trames P

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



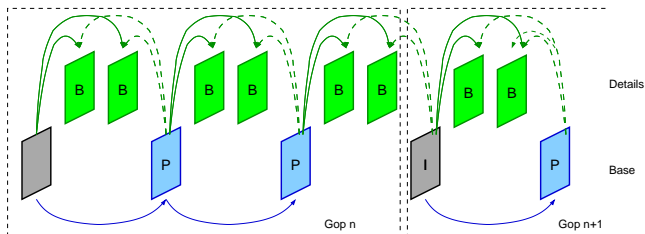
Trames B

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



Ordre de codage des trames

I → AF → Trames B → AF → Trames B ...
Retard ?





Plan

La prédiction temporelle

La structure du GOP (Group of Pictures)

Le codeur hybride

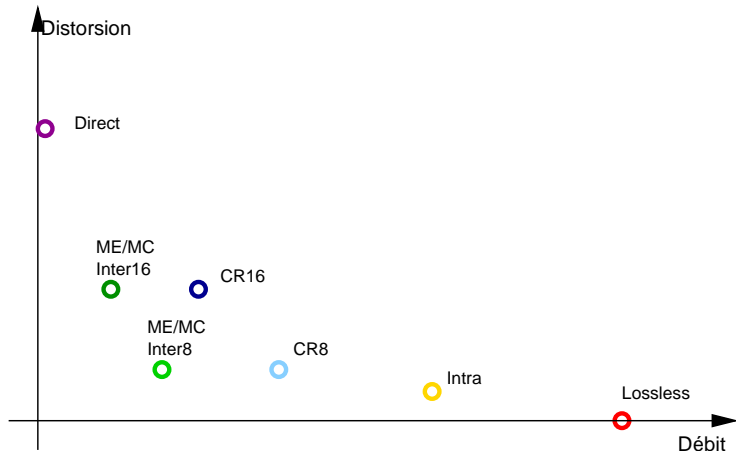
La norme MPEG-1

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 pertes

Codeurs hybrides

Exemple de performances des modes



Codeurs hybrides

Choix du mode de codage

- Objectif : minimiser D pour un R donné :

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

- Le pas de quantification Q est donné
- L'ensemble des modes $\mathbf{i} = \{i_k\}_{k=1}^K$ doit être choisi en sorte que on puisse 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 \mathbf{i} est trop complexe
- ▶ On préfère une minimisation sous-optimale
- ▶ Pour chaque MB k , on choisit le mode en sorte de 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

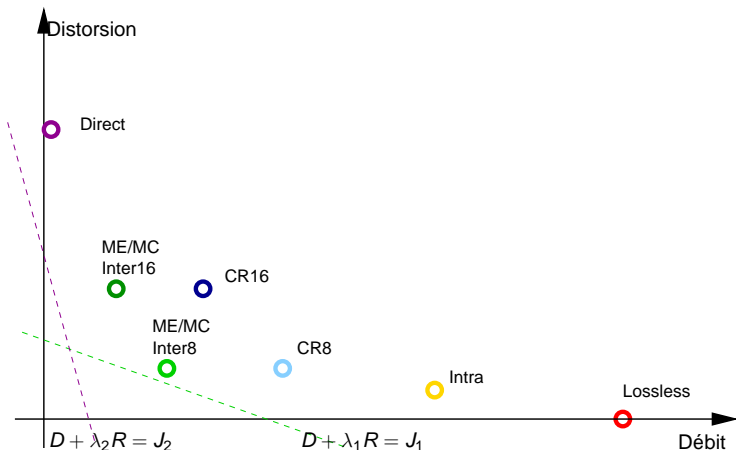
Codeurs hybrides

Choix du mode de codage

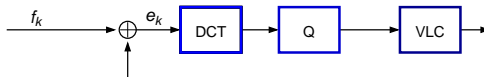
- ▶ Le pas de quantification Q est considéré comme un *input*
- ▶ Pour chaque Q (débit) il existe une valeur optimale de λ , déterminée empiriquement
 - ▶ MPEG-2 : $\lambda = aQ^2 + b$
 - ▶ H.264 : $\lambda = c2^{dQ+e}$
- ▶ Avec le λ donné, minimiser J_k revient à trouver une droite sur le plan RD

Codeurs hybrides

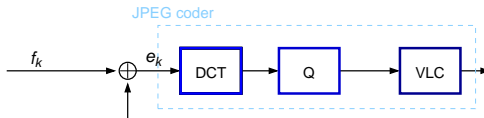
Exemple de performances des modes



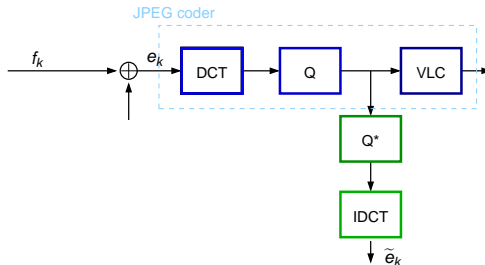
Le codeur hybride



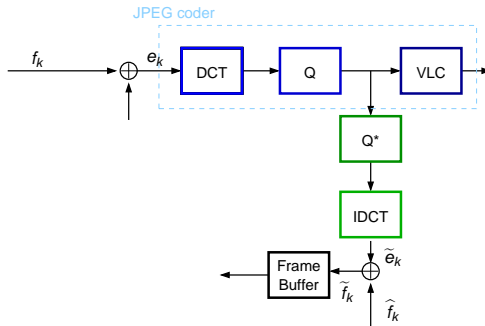
Le codeur hybride



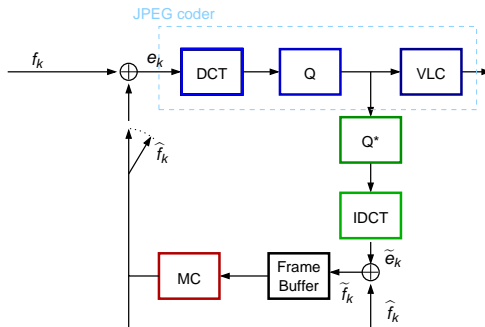
Le codeur hybride



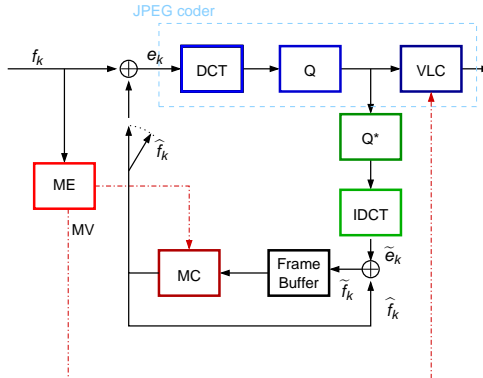
Le codeur hybride



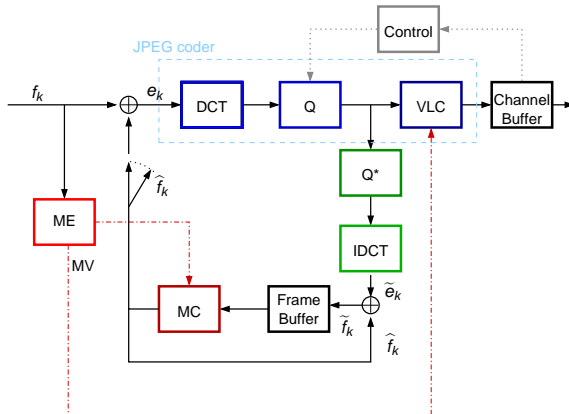
Le codeur hybride



Le codeur hybride

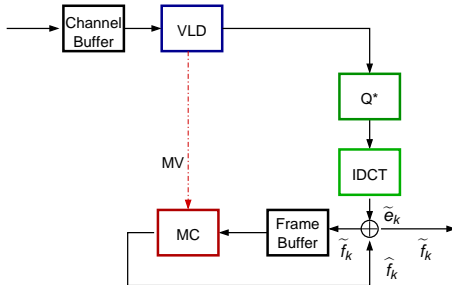


Le codeur hybride



Le décodeur hybride

Schéma asymétrique !



Plan

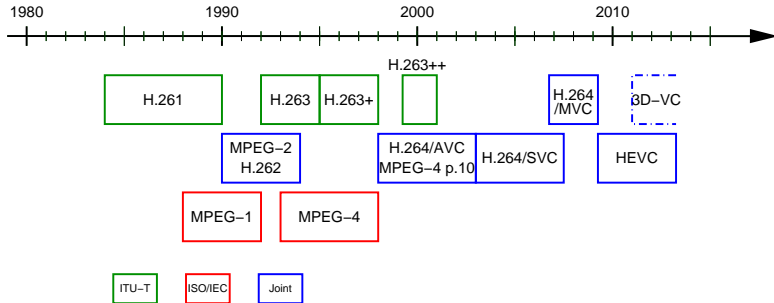
La prédiction temporelle

La structure du GOP (Group of Pictures)

Le codeur hybride

La norme MPEG-1

Les standards vidéo : chronologie



Les standards vidéo

Groupes de standardisation

Organismes de standardisation :

ISO International Standardization Organization

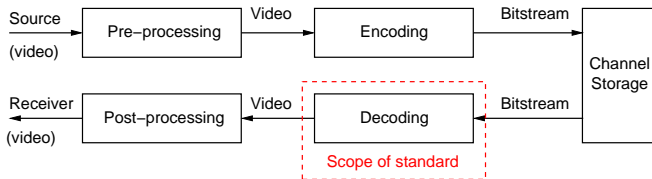
IEC International Electrotechnical Commission

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); standards récents

Scope des standards



- ▶ Le standard ne définit que la syntaxe du train binaire et les outils de décodage
- ▶ But : interopérabilité, concurrence
- ▶ Pédagogie : description de l'encodage

Le standard MPEG-1

- ▶ Développé en 1988-1992
- ▶ Parties du standard
 1. Systèmes
 2. Video
 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, vidéo CD, jeux vidéo

Nouveautés techniques

- ▶ Types d'images
- ▶ ME/MC à précision sous-pixelique

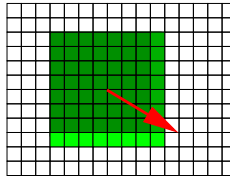
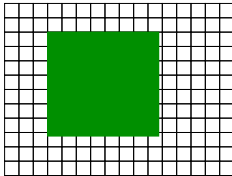
Standard MPEG-1

ME/MC à précision sous-pixelique

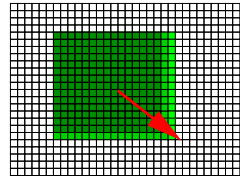
- ▶ Le mouvement ne correspond pas à la grille des pixels
- ▶ Interpolation pour améliorer la précision
- ▶ Ulérieure augmentation de la complexité
- ▶ Très bonnes performances de codage

Standard MPEG-1

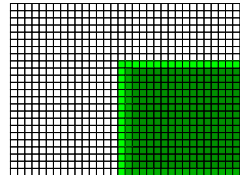
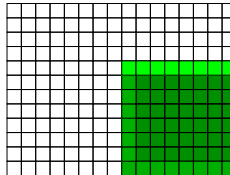
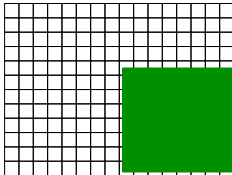
ME/MC à précision sous-pixelique



$v=(5,3)$



$v=(5.5,2.5)$



Standard MPEG-1

Résumé

Codeur hybride classique (H.261), plus :

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