



Technologies du multimédia

Traitement multimédia, standards de compression

Grégoire Mercier

Institut TELECOM / TELECOM Bretagne / dpt ITI
CNRS Lab-STICC 3192, Equipe CID
MTS 443 – C3-C4



Stratégie de compression des images dynamiques
Gestion des flux multimédias

H.261
MPEG 1 & 2

Contenu

- 1 Stratégie de compression des images dynamiques
 - H.261
 - MPEG 1 & 2

- 2 Gestion des flux multimédias
 - Contraintes
 - MPEG 4





Images dynamiques

Stratégie de codage

Définition :

$\nrightarrow \neq$ volume 3D

$\nrightarrow \neq 2D \oplus 1$

$\nrightarrow = 2D \oplus \text{temps}$

Conséquence :

\longrightarrow Notion de mouvement

\nrightarrow Suivre des blocs carrés rigides le long du temps

\nrightarrow Utiliser un modèle de mouvement

(cf. translation_{|blocs})

\nrightarrow Estimation du mouvement

(mise en correspondance de blocs)

\nrightarrow Compensation de mouvement

\nrightarrow Contraintes de temps réel

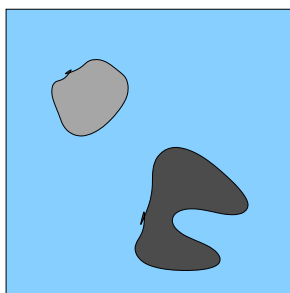
\longrightarrow au décodage



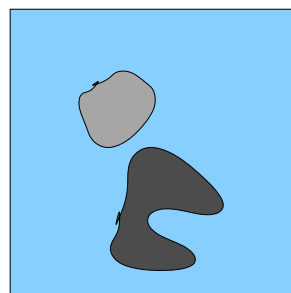
Codage différentiel

Implémentation simpliste :

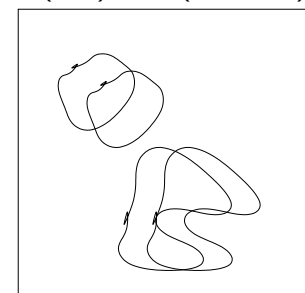
L'erreur de prédiction :



image($t - 1$)



image(t)



image(t)
 \ominus image($t - 1$)

Avantages :

- Réduction de la dynamique

Inconvénients :

- Occlusions
- Nouvelles structures

Estimation de mouvement

Formulation :

estimer le déplacement de blocs au cours du temps

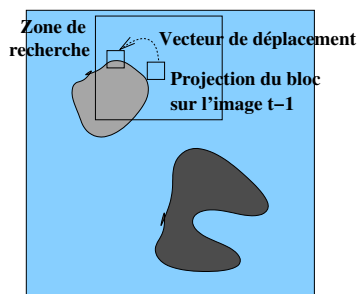


image $t - 1$

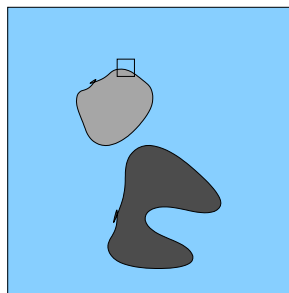
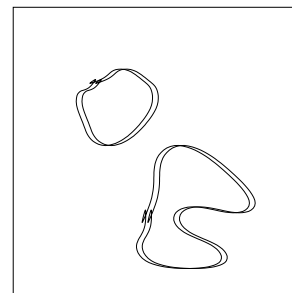


image t



image(t)
 \ominus image ($t - 1$)

Avantages :

- Meilleure estimation

Inconvénients :

- Occlusions
- Nouvelles structures
- Définition du modèle de prédiction

La norme H.261

But de la normalisation :

Services audiovisuels à $p \times 64$ kbits/s , $p \in \{1, 2, \dots, 30\}$

cf. réseau numéris (RNIS)

Format d'image CIF :

Common Intermediate Format

- ↪ Nb ligne en luminance : 288
- ↪ Nb ligne en chrominance : 144
- ↪ Nb points par ligne en luminance : 360
- ↪ Nb points par ligne en chrominance : 180
- ↪ Balayage non entrelacé
- ↪ cadence de répétition des images : 29,97 Hz

Format Q-CIF :

Quart-CIF

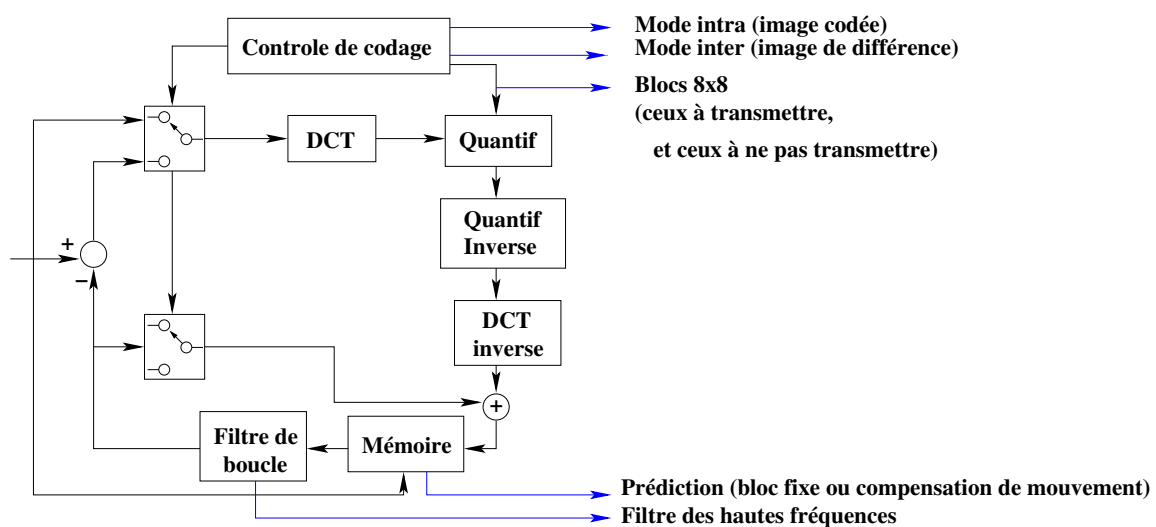
idem avec des dimensions à diviser par 2

La norme H.261 (suite)

Caractéristiques :

- ↪ Taux de compression : $\eta = 20 : 1$
- ↪ Effets de blocs dûs
 - À la quantification de blocs 8×8 transformés par DCT
 - Aux erreurs de transmission
- ↪ Utilise le circuit JPEG en interne
- ↪ Applications :
 - Visiophonie
 - Visioconf.
 - Travail coopératif

La norme H.261 (suite)





Les normes MPEG 1 & 2

Objectifs :

MPEG 1 :

- ↪ Enregistrement sur CD-ROM, DAT, Disques optiques
- ↪ Transmission sur réseaux large bande (RNIS asynchrones, TTA, LAN)
- ↪ 1.5 Mbits/s

MPEG 2 :

- ↪ TV & TVHD



Les normes MPEG

Estimateurs de mouvement

Recherche du mouvement

- ↪ On cherche à minimiser pour chaque macrobloc l'erreur de prédiction :

$$\sum_{(x,y) \in \text{bloc}} \|I(x, y, t) - I(x + dx, y + dy, t - 1)\|$$

par rapport au *vecteur de mouvement* (dx, dy) .

On utilise comme *critère d'erreur* $\|\cdot\|$ soit

- le carré de l'erreur (norme 2)
- la valeur absolue de l'erreur (norme 1)

On peut effectuer sur la zone de recherche :

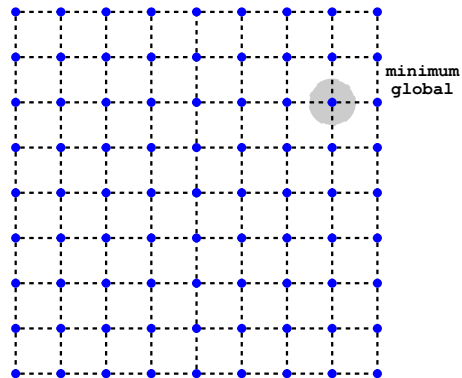
- Une minimisation globale par recherche exhaustive (lent,optimal)
- Une minimisation locale selon diverses heuristiques (rapide,sous-optimal)



Les normes MPEG

Estimateur exhaustif

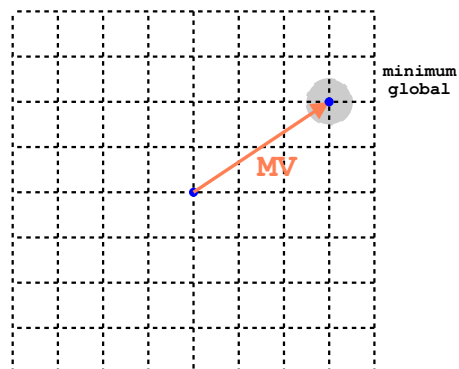
Recherche sur l'ensemble du voisinage



Les normes MPEG

Estimateur exhaustif

Recherche sur l'ensemble du voisinage



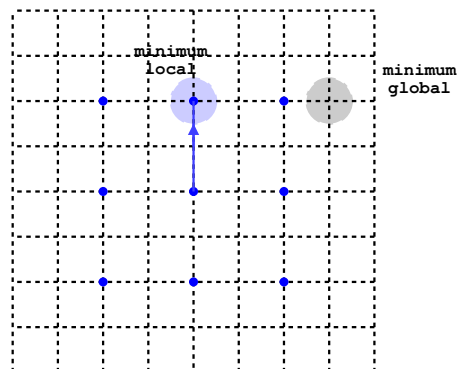
- + Optimal
- Complexité prohibitive

Les normes MPEG

Estimateur logarithmique

Recherche en plusieurs étapes successive sur 8 voisins

- nouvelle zone de recherche centrée sur le minimum local
- pas divisé par 2 à chaque étape

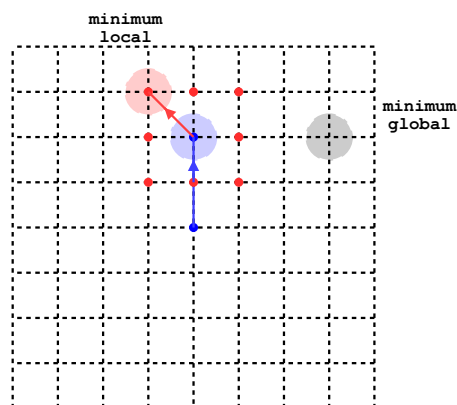


Les normes MPEG

Estimateur logarithmique

Recherche en plusieurs étapes successive sur 8 voisins

- nouvelle zone de recherche centrée sur le minimum local
- pas divisé par 2 à chaque étape

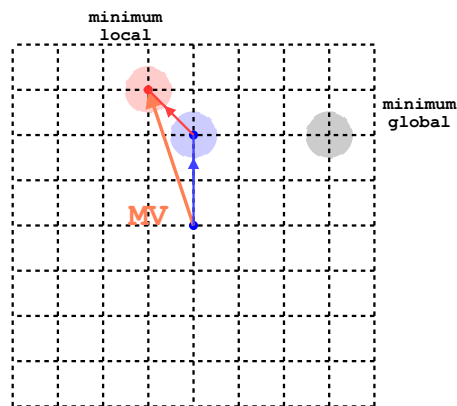


Les normes MPEG

Estimateur logarithmique

Recherche en plusieurs étapes successive sur 8 voisins

- nouvelle zone de recherche centrée sur le minimum local
- pas divisé par 2 à chaque étape



Les normes MPEG

Estimateur logarithmique

Recherche en plusieurs étapes successive sur 8 voisins

- nouvelle zone de recherche centrée sur le minimum local
- pas divisé par 2 à chaque étape

- + Complexité réduite
- Performances moyennes

Les normes MPEG

Estimateur avancé (PMVFAST)

Prédiction du vecteur de mouvement et recherche en diamant

- Prédiction par les vecteurs de mouvement d'un voisinage causal
- Taille de la zone de recherche fonction de la variance de ces prédicteurs

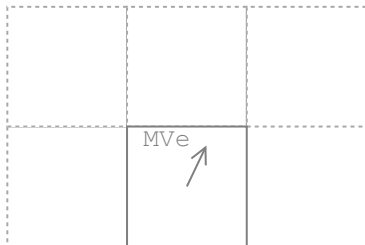


image précédente

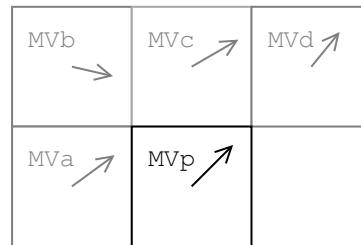


image courante

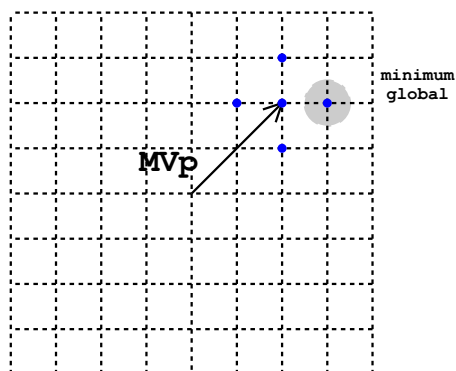
$$MVP = \begin{pmatrix} \text{mediane}(MVA_x, MVB_x, MVC_x, MVD_x, MVE_x) \\ \text{mediane}(MVA_y, MVB_y, MVC_y, MVD_y, MVE_y) \end{pmatrix}$$

Les normes MPEG

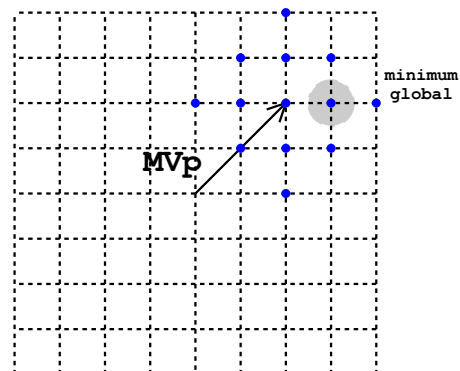
Estimateur avancé (PMVFAST)

Prédiction du vecteur de mouvement et recherche en diamant

- Prédiction par les vecteurs de mouvement d'un voisinage causal
- Taille de la zone de recherche fonction de la variance de ces prédicteurs



petit diamant



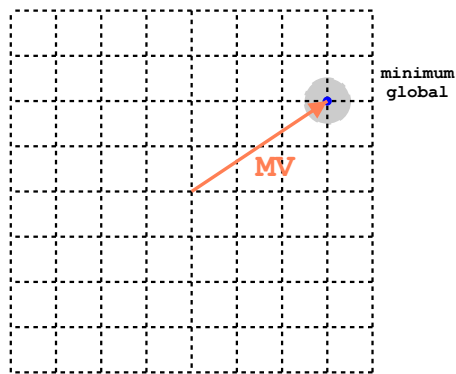
grand diamant

Les normes MPEG

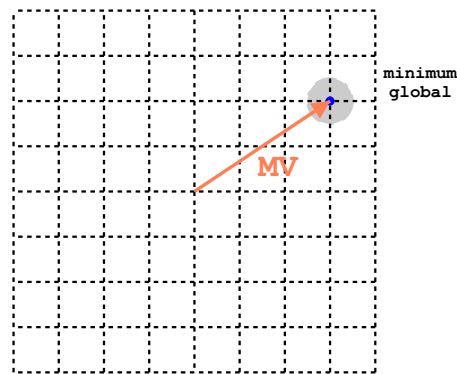
Estimateur avancé (PMVFAST)

Prédiction du vecteur de mouvement et recherche en diamant

- Prédiction par les vecteurs de mouvement d'un voisinage causal
- Taille de la zone de recherche fonction de la variance de ces prédicteurs



petit diamant



grand diamant

Les normes MPEG

Estimateur avancé (PMVFAST)

Prédiction du vecteur de mouvement et recherche en diamant

- Prédiction par les vecteurs de mouvement d'un voisinage causal
- Taille de la zone de recherche fonction de la variance de ces prédicteurs

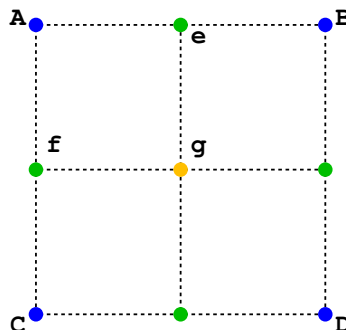
- + Complexité réduite
- + Performances proche de la recherche exhaustive

Les normes MPEG

Recherche sous-pixel

Le mouvement en pixels peut être fractionnaire

↪ recherche au demi pixel par interpolation de l'image de prédiction



$$e = (A + B + 1)/2$$

$$f = (A + C + 1)/2$$

$$g = (A + B + C + D + 2)/4$$

Les normes MPEG

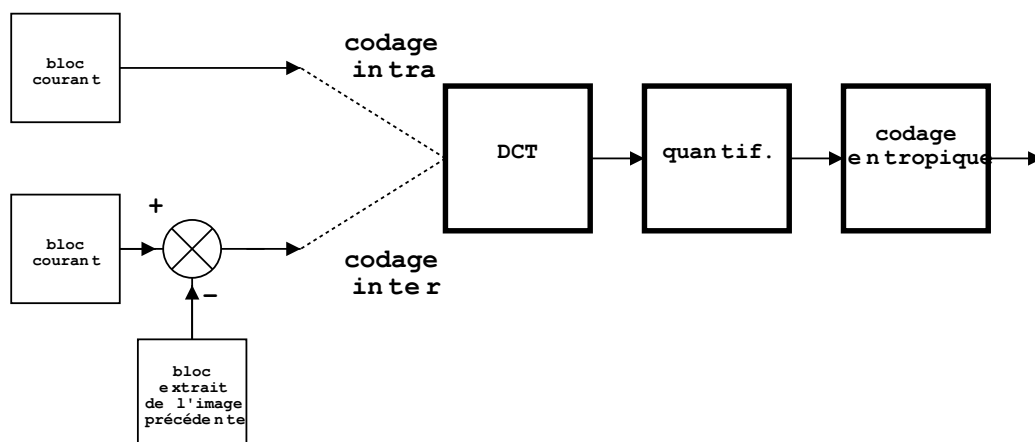
Types de macroblochs

On appelle :

- *intra* un macrobloc dont les 4 blocs sont codés indépendamment
- *inter* un macrobloc dont les 4 blocs sont prédits

↪ Le codage des blocs *intra* s'effectue comme pour JPEG.

↪ Pour les blocs *inter* le prédicteur est soustrait au préalable





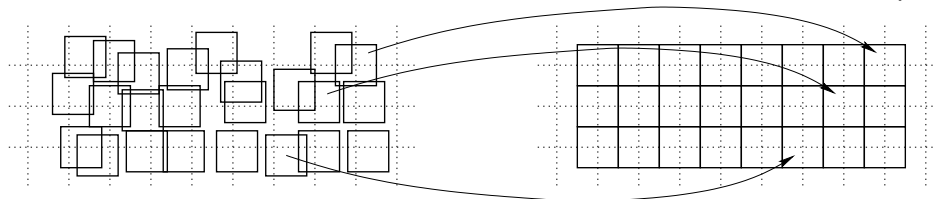
Les normes MPEG

Types de prédiction temporelle

Définition du sens de prédiction :

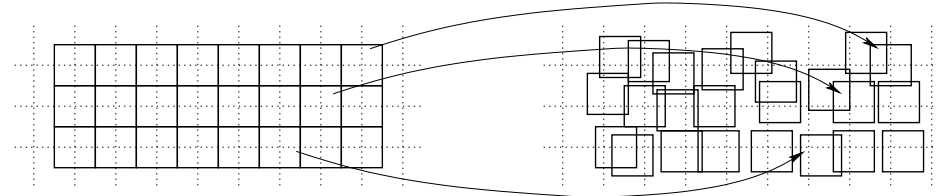
1 Avant

(forward)



2 Arrière

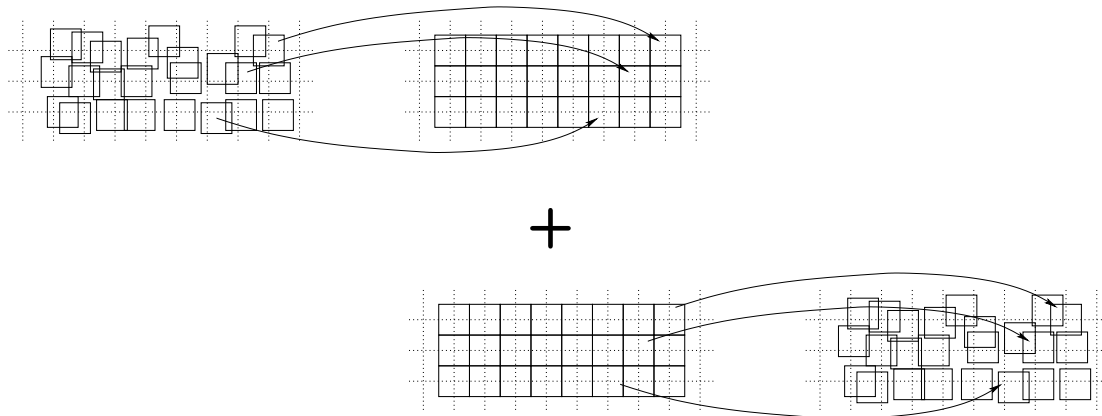
(backward)



Les normes MPEG

Types de prédiction temporelle

3 Bidirectionnelle



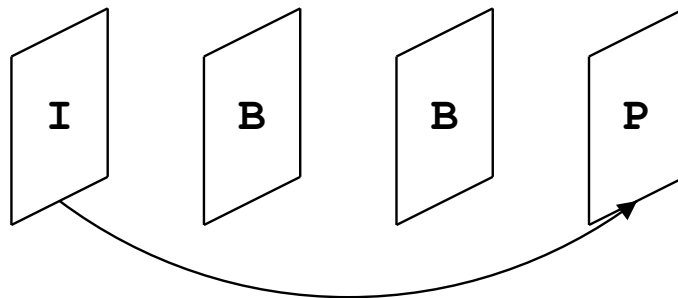
$$\text{prédicteur} = \frac{\text{avant} + \text{arrière}}{2}$$

Les normes MPEG

Types d'images

On définit 3 types d'images :

- **I** (*intra* ou *clé*) codée indépendamment des autres images
- **P** (*prédite*) codée par rapport à l'image précédente
- **B** (*bidirectionnelle*) codée par rapport à une image précédente et suivante

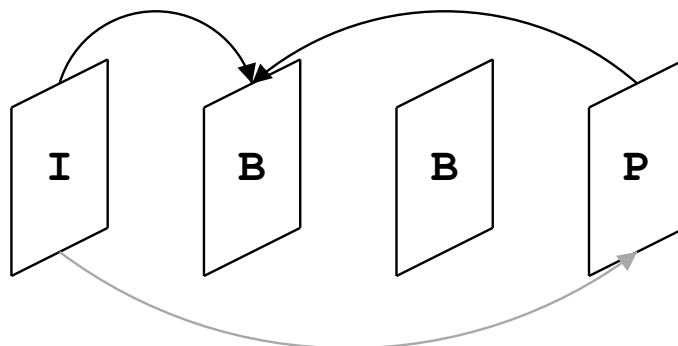


Les normes MPEG

Types d'images

On définit 3 types d'images :

- **I** (*intra* ou *clé*) codée indépendamment des autres images
- **P** (*prédite*) codée par rapport à l'image précédente
- **B** (*bidirectionnelle*) codée par rapport à une image précédente et suivante

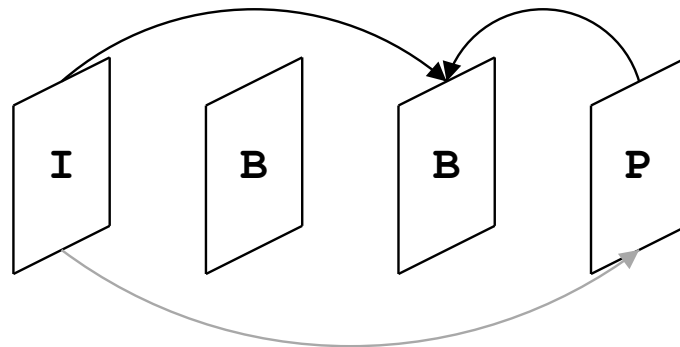


Les normes MPEG

Types d'images

On définit 3 types d'images :

- **I** (*intra* ou *clé*) codée indépendamment des autres images
- **P** (*prédite*) codée par rapport à l'image précédente
- **B** (*bidirectionnelle*) codée par rapport à une image précédente et suivante



Les normes MPEG

Types d'images

- On utilise toujours l'image **décodée** pour la prédiction
le décodeur n'a pas connaissance de l'image de prédiction d'origine
- Les images B ne servent pas à la prédiction d'autres images
permet une lecture rapide en ne décodant que les images I et P
- L'ordre de décodage n'est pas l'ordre temporel
l'image de prédiction arrière doit être décodée avant l'image B
IBBPBBP (temps) → **IPBBPBB** (flux compressé)

Les normes MPEG

Types d'images et blocs

Certains modes de codages sont autorisés en fonction du type d'image :

type	I (intra/JPEG)	P (prédite)	B (bidirectionnelle)
blocs intra	X	X	X
blocs inter avant		X	X
blocs inter arrière			X
blocs inter bidirectionnel			X

Pour chaque bloc, le mode minimisant l'énergie résiduelle à coder est choisi

On ajoute souvent une pénalité supplémentaire aux modes inter pour tenir compte du coût de codage des vecteurs de mouvement

Les normes MPEG

Contrôle de débit

On se trouve souvent dans des situations où le débit maximum est imposé

- transmission sur réseaux (Internet, satellite)
- DVD, Video CD, ...

→ contrôle du débit

↪ En faisant varier le pas de quantification Δ (pour l'image ou par bloc)

↪ Modèle débit-distortion $R = f(\Delta)$:

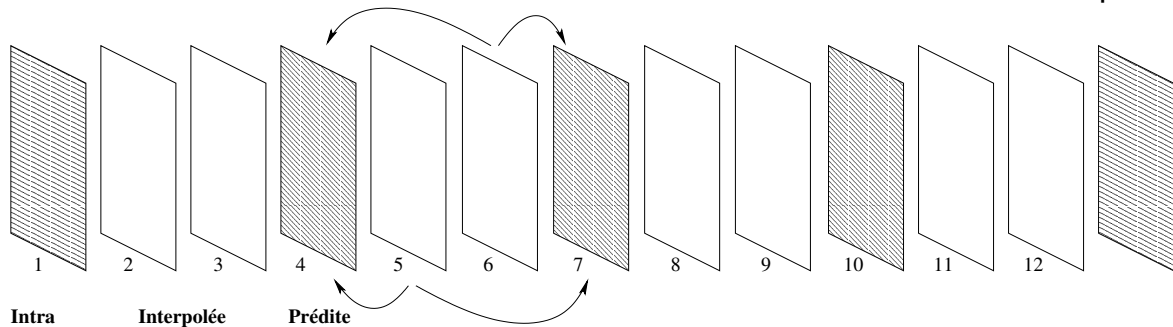
- paramétrique linéaire $R = \frac{a}{\Delta}$
- paramétrique quadratique $R = \frac{a}{\Delta} + \frac{b}{\Delta^2}$
- optimisation globale sous contrainte de débit

MPEG 1 & 2 : caractéristiques

- MPEG 1 :**
- ↪ cf. H.261
 - ↪ Mode de prédiction plus complexes \oplus images interpolées
 - ↪ Complexité/mémoire plus grande

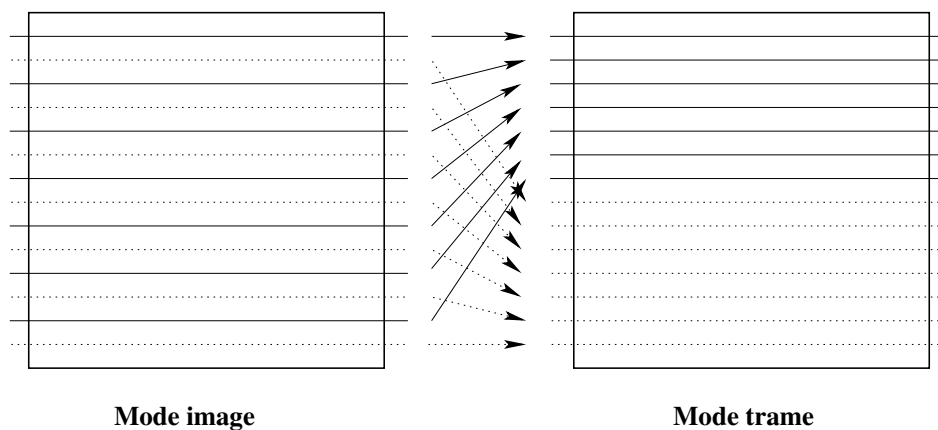
- MPEG 2 :**
- ↪ cf. H.263
 - ↪ Mode entrelacé
 - ↪ cf. MPEG 1 (compatibilité)
 - ↪ *Streaming*
 - ↪ Notion de boîte à outils

Définition de profils



Norme MPEG 2

Mode image / mode trame :



Mode image

Mode trame



Norme MPEG 2

Profils :

- Simple** : pas d'images interpolées
- Principal** : schéma de codage classique
 - SNR** : qualité d'image améliorée
- Spatial** : codage TVHD à partir du format TV par approche hiérarchique
- High** : codage 4:2:2

Niveaux :

- Low** : format 1/4 TV
- Main** : format TV
- High 1440** : TVHD en 1440 colonnes
- High** : TVHD (1920 colonnes)



Contenu

- 1 Stratégie de compression des images dynamiques
 - H.261
 - MPEG 1 & 2
- 2 Gestion des flux multimédias
 - Contraintes
 - MPEG 4

Panorama des grandes questions

Compression

- ↪ Les données multimédias nécessitent des capacités de **stockage** importantes.
- ↪ Les supports de stockage actuels, relativement lents, ne peuvent pas rejouer des données multimédias, surtout la vidéo, en **temps réel**.
- ↪ La bande passante des réseaux de transmission ne permet pas (en général) la transmission de vidéo en **temps réel**.

Description de scène

- ↪ Difficulté d'**indexer** le contenu de documents multimédia
- ↪ Difficulté de rendre la lecture d'un document multimédia **interactive**.
(changer la langue, le point de vue d'une scène 3D, le mixage d'un concert...)

Réseaux multimédias

- ↪ Débit ou bande passante élevée
cf. MPEG-1 $\approx 1,5$ Mbits/s,
MPEG-2 $\approx 4 \dots 10$ Mbits/s,
TVHD ≈ 20 Mbits/s

Panorama des grandes questions (suite)

Réseaux multimédias (suite)

- ↪ Erreurs tolérables / délais inacceptables
- ↪ Faible latence (délai) pour assurer les interactions
- ↪ Transmission synchronisée à faible gigue
- ↪ Communications multipoints
- ↪ mots-clés :
 - Ethernet, Ethernet commuté
 - Réseau Numérique à Intégration de Services (RNIS bande étroite, ISDN)
 - Réseau Numérique à Intégrations de Services large bande (RNIS large bande, B-ISDN)
en mode synchrone (STM) ou asynchrone (ATM)

Panorama des grandes questions (suite)

Synchronisation multimédia : associent divers médias (discrets / continus) pour créer des documents multimédias composites

↪ **Composition spatiale** qui permet de lier divers objets multimédias discrets en une seule entité via des règles de placement et de déformation ;

↪ **Composition temporelle** qui permet de créer une présentation multimédia par arrangement d'objets multimédias selon des relations temporelles

cf. **synchronisation**

Divers types de synchronisation

Synchronisation continue : pour les médias continus
ex. visiophonie – signaux audio et vidéo synchronisés à la réception

Synchronisation ponctuelle : pour les médias discrets et continus
ex. projection de diapos avec commentaires audio synchronisés

Synchronisation séquentielle : débit auquel les événements doivent arriver à l'intérieur d'un même flot de données
synchronisation **intramédia**

Synchronisation parallèle : séquençage relatif de deux flots de synchronisation séparés
synchronisation **intermédia**

Les systèmes multimédias

Avancées techniques récentes :

Puissance des stations de travail et des téléphones portables /
Capacité des systèmes de stockage et de transmission / Débit
des réseaux / Méthodes de traitement des images et de la vidéo
(dont aspects image de synthèse) / Traitement de la parole
(reconnaissance de locuteur et conversion texte vers parole) /
Compression de la parole, de l'audio, de l'image et de la vidéo

Éléments clés :

- ① Support Physique
- ② Système d'exploitation / Interface graphique
- ③ Logiciels multimédias / Systèmes auteurs

Applications et Services

Messagerie électronique multimédia : évolution du courrier électronique
actuel incluant des fonctionnalités étendues comme l'édition de
courrier vocal, utilisation de la vidéo.

Conférences multimédias : des participants distants échangent des
informations multimédias au travers des stations connectées par
un réseau à haut débit.

Travail coopératif : des membres distants d'un groupe de travail peuvent
discuter un problème et créer un produit ensemble, par le biais
d'outils de visiophonie/visioconférence et d'édition de
documents multimédias partagés



Evolution de la normalisation

MPEG-1 :

↔ Norme pour le stockage de l'information AV sur CD-Rom

MPEG-2 :

↔ Norme pour la distribution de télévision numérique
⇒ Normes orientées Terminal et décodeur

MPEG-4 :

↔ Norme pour les communications multimédia

MPEG-7 :

↔ Norme pour la description du contenu audiovisuel

ISO MPEG (ISO/IEC JTC1/SC29/WG11) Phase 4 :

Fondement normatif commun aux applications de communication multimédia



MPEG-4 : Évolution des modèles de communication

- ① Production
- ② Acheminement
- ③ Consommation



MPEG-4 : Évolution des modèles de communication

Production :

↪ Classique

- Modèle linéaire : capteur (audio, vidéo) → postproduction
- Format d'information monolithique

↪ Tendance :

- Usage généralisé d'ordinateur et logiciels d'édition de contenu
- Matériau AV naturels et synthétique (2D, 3D, son spatial)
- Format / multiplicité des sources, réutilisation de l'existant, plus riche et plus flexible



MPEG-4 : Évolution des modèles de communication

Acheminement : De réseaux homogènes (ex. RNIS) ...

↪ À des systèmes de plus en plus hétérogènes

- Réseaux mobiles
- Câble
- RTC

↪ À qualité de service (QoS) variable

- Internet

↪ Grâce à des technologies récentes.

- Compression
- Robustesse

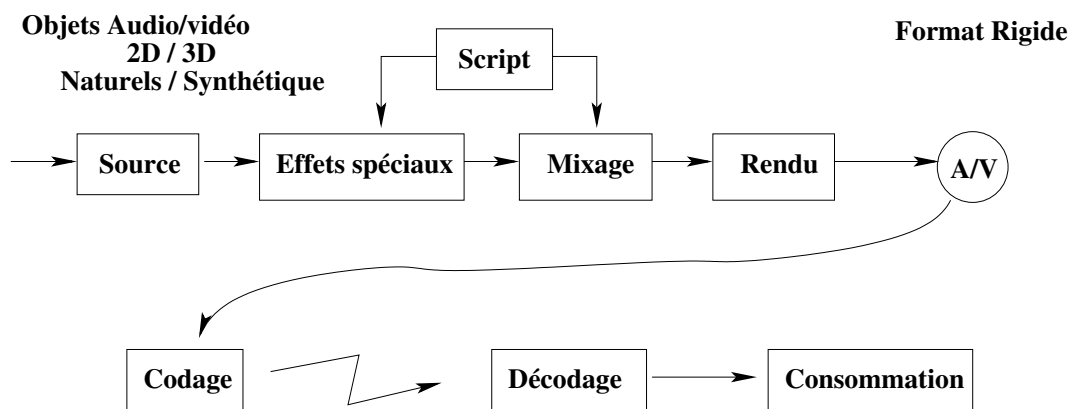
MPEG-4 : Évolution des modèles de communication

Consommation :

- ↪ Part plus grande d'info multimédia
- ↪ « *Consommation* » plus interactive
- ↪ Modes d'interaction variés
 - Choix de la langue d'un film
 - Interaction avec un serveur
 - Navigation de scène
 - Jeux en réseau
 - ...

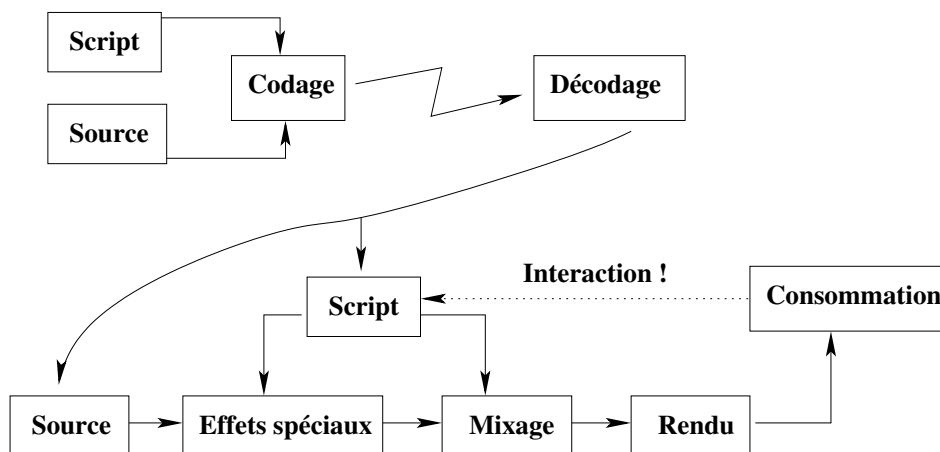
Production, Acheminement, Consommation

Paradigme classique :



Production, Acheminement, Consommation

Paradigme MPEG-4 :



Production, Acheminement, Consommation

	Classique	MPEG-4
Production	2D, Caméra & Microphone	2D/3D, génération par ordinateur, matériel pré-segmenté, Codage AV hybride naturel/synthétique
Acheminement	Quelques réseaux transportent l'info AV	Nouveaux réseaux transportant l'info AV (PSTN, sans fil, Internet), hétérogènes
Consommation	Consommation passive. Quelques médias types sur réseaux à bas débits (HTML)	L'info est lue, vue et entendue de manière interactive (interaction avec les objets AV), de \oplus en \oplus d'info.



MPEG-4 : Évolution des services

Évolution des modèles de

- Production
- Acheminement
- Consommation



Évolution des services de

- Communication
- Diffusion
- Interactivité



Besoins et Objectifs

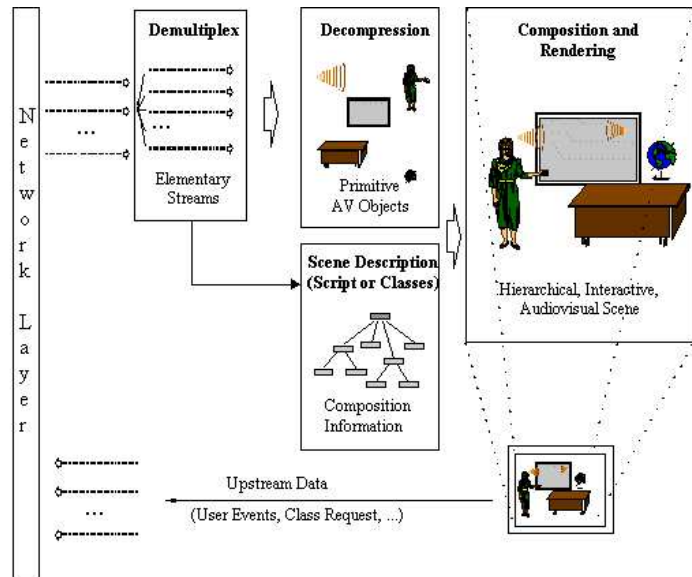
Besoins

- ↪ Technologie commune à **plusieurs types de services** multimédias (interactifs, diffusion, conversationnels)
- ↪ Augmenter l'**interactivité** côté client ou client-serveur
- ↪ Intégrer les **contenus naturels et synthétiques**
- ↪ Couvrir une vaste gamme de **conditions d'accès**
- ↪ Fournir des solutions efficaces pour la **gestion** et la **protection** de la propriété intellectuelle

Objectifs

- ↪ Fondement normatif pour les technologies de l'information multimédia
Outils normalisés pour les besoins des contextes production, acheminement, consommation.
- ↪ Définir un environnement flexible et fiable permettant le développement de services hybrides et leur inter fonctionnement

The diagram illustrates a 3D audio-visual system architecture. A central figure (a woman) interacts with a 3D environment. The system includes a 'multiplexed downstream control/data' block, a 'multiplexed upstream control/data' block, a 'video computer projection plane' showing a 2D view, and a 'hypothetical viewer' at the bottom left. The 3D environment contains 'audiovisual objects' (a globe and a box), a '3D background', and a 'scene coordinate system' with x, y, and z axes. A 'user event' is shown as an arrow from the user to the 3D objects. A 'video computer' at the bottom right processes 'user input' and outputs to a 'display' and 'audio computer'.



→ **Scènes audio-visuelles** faites d'**objets audio-visuels composés** selon une description de scène

- Interaction avec les éléments de la scène AV
- Schéma de codage spécifiques pour chaque objet
- Réutilisation de contenu AV

- Relations spatiales / temporelles entre les objets AV (2D, 3D, description de scène mixte 2D et 3D),
- Comportement et interactivité des objets AV et des scènes,
- Protocoles pour modifier et animer la scène,
- Encodage binaire pour la scène.

MPEG-4 : Principes (suite)

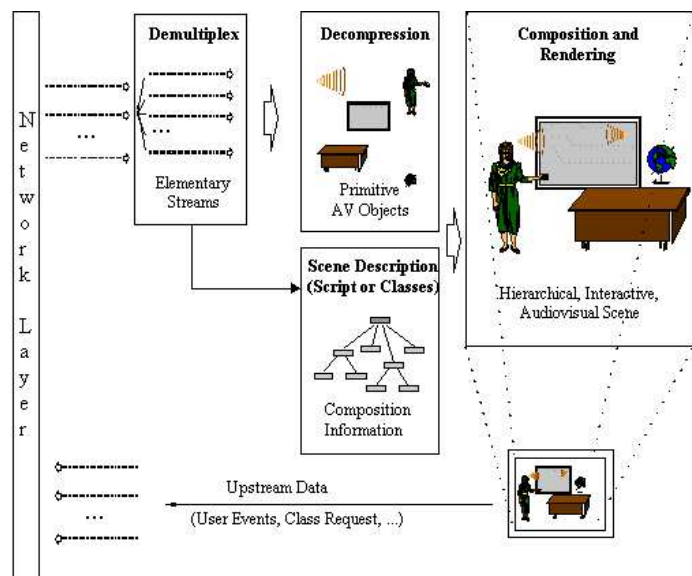
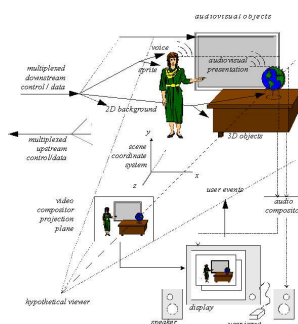
Objets AV de différentes nature

- **Audio** : simples ou multi-canal
- **Vidéo** : de forme quelconque ou rectangulaire
- **Naturelle ou synthétique** : audio/vidéo mais aussi texte et graphique
- **2D ou 3D** : page web → son spatialisé, monde virtuel...
- **Streamed ou téléchargé** : film vidéo ou jingle audio

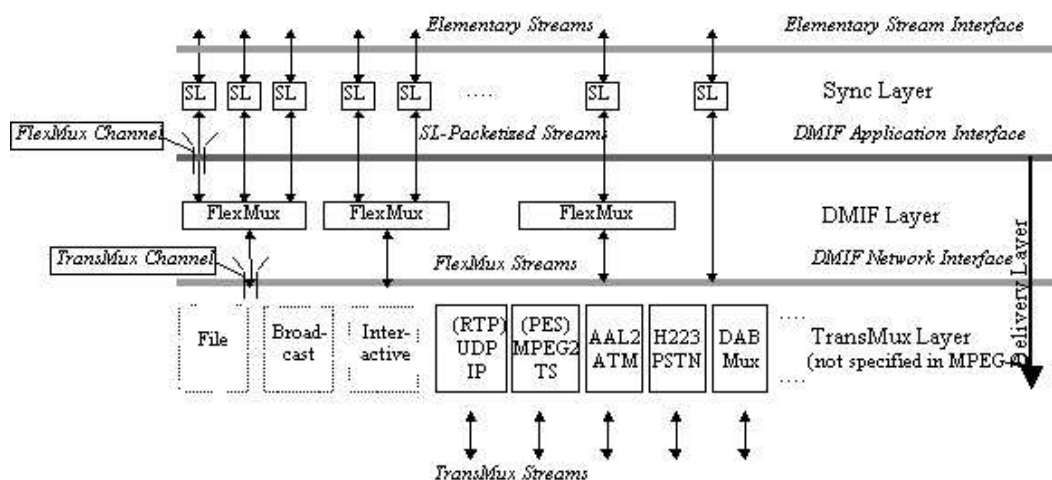
- Ces principes sont indépendants du débit.
- Toutes les informations sont fournies dans un format comprimé.

Couche système définie indépendamment du réseau et des protocoles

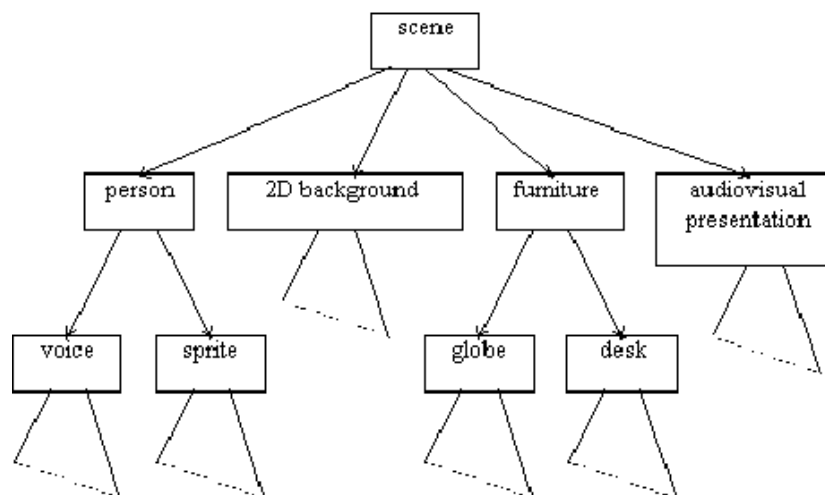
MPEG-4 : Principes



MPEG-4 : Couche System



MPEG-4 : structure de scène



MPEG-4 Système

« La représentation codée de la combinaison des différents flux élémentaires d'information audio-visuelle »

Scène Audio-Visuelle Représentation codée 14496-1	
Information Audio naturelle et Synthétique représentation codée 14496-2	Information Vidéo naturelle et Synthétique représentation codée 14496-3
Synchronisation des informations AV 14496-4	

Architecture générale

/

multiplexage

/

Synchronisation

MPEG-4 Audio Version 1

Codage audio naturel de 2 kbit/s à 64 kbit/s :

- ↪ 2 - 4 kbit/s pour la parole avec 8 kHz de fréquence d'échantillonnage, 4 - 16 kbit/s pour l'audio échantillonné à 8 ou 16 kHz (techniques de codage paramétrique).
- ↪ Codage de la parole à débit moyen entre 6 et 24 kbit/s, 8 et 16 kHz de fréquence d'échantillonnage en méthode Code Excited Linear Predictive (CELP).
- ↪ MPEG-2 AAC : compression générique de l'audio pour les débits les plus élevés. . .
- ↪ Scalabilité (pas de ≈ 1 kbit/s et de ≈ 8 kbit/s).



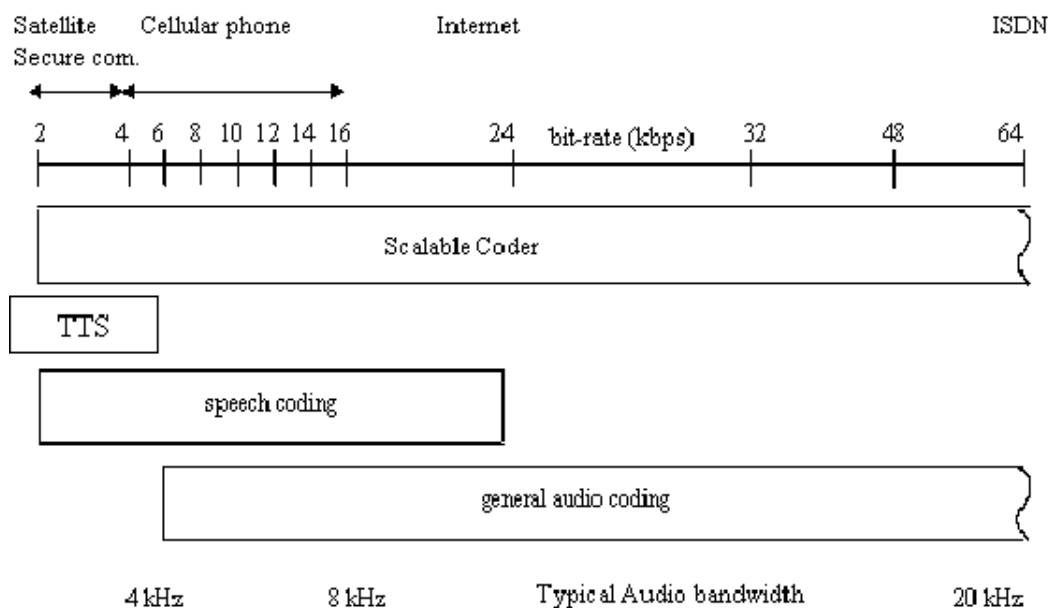
MPEG-4 Audio Version 1 (suite)

Codage de l'audio synthétique :

- ↪ Parole : décodeur *Text-To-Speech* (TTS).
- Génération de parole synthétique à partir d'un texte, ou texte avec des paramètres prosodiques (pitch contour, durée de phonème, etc).
 - Animation faciale contrôlée par l'information phonémologique.
- ↪ Musique synthétique (*Structured Audio Decoder*)
- Un langage de synthèse SAOL (*Structured Audio Orchestra Language*) définit un « orchestre » fait d'« instruments ».
 - Contrôle de la synthèse par téléchargement de « partitions » ou de « scripts » dans le langage SASL (*Structured Audio Score Language*).
Midi peut aussi être utilisé pour le contrôle de l'orchestre.
 - Très bas débit.



MPEG-4 Audio Version 1 (suite)





MPEG-4 Vidéo Version 1 & 2

- ↪ Vidéo naturelle de 10 kbit/s à 10 Mbit/s
 - Objets rectangulaires ou de forme arbitraire,
 - Scalable (divers niveaux de qualité selon une hiérarchie),
 - Balayage entrelacé et progressif ,
 - « *Sprites* » (fonds de scènes) : à envoyer une fois, déformer ensuite.
- ↪ Information visuelle générée par ordinateur
 - Animation faciale,
 - Grilles animées ⊕ textures dynamiques,
 - Texte animé et graphiques.
- ↪ Tous ces objets peuvent être synchronisés entre eux



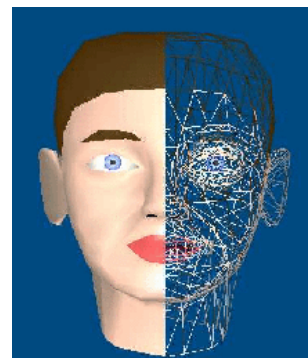
Exemples d'objets vidéo MPEG-4



Objet vidéo de
forme rectangulaire

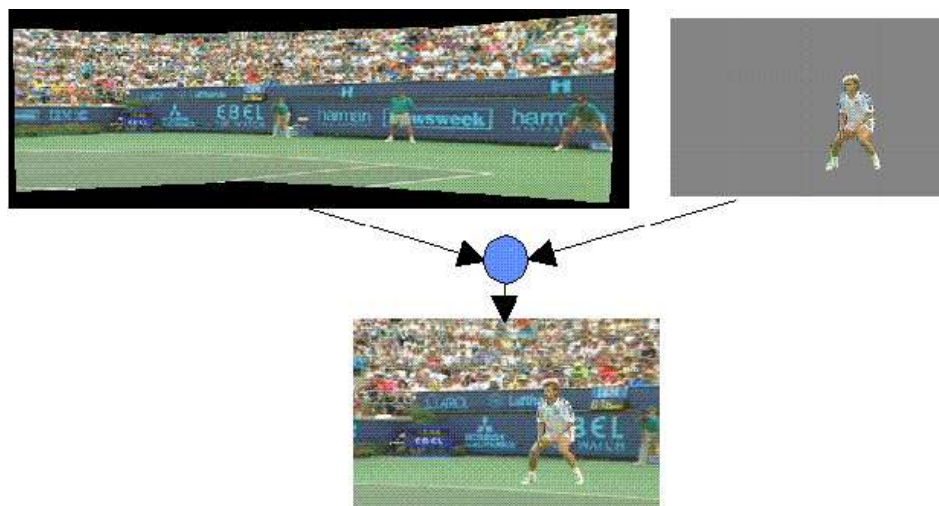


Objet vidéo de
forme quelconque



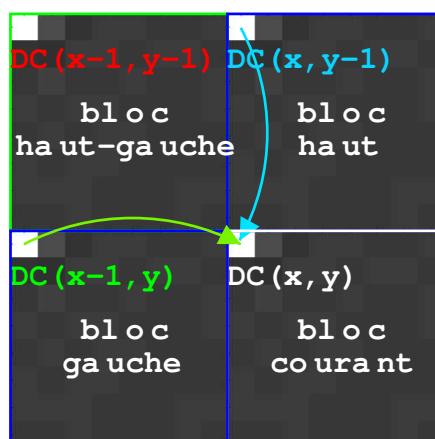
Visage animé

Exemple de « sprite »



Amélioration du codage d'image

■ Prédiction du coefficient DC par DPCM *adaptatif*



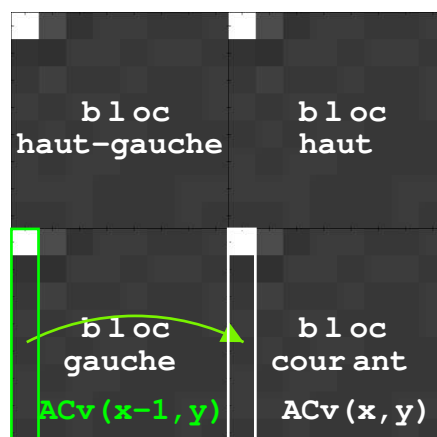
si $|DC(x, y - 1) - DC(x - 1, y - 1)| < |DC(x - 1, y) - DC(x - 1, y - 1)|$
 $DC(x, y)$ prédit par $DC(x - 1, y)$

sinon

$DC(x, y)$ prédit par $DC(x, y - 1)$

Amélioration du codage d'image

- Prédiction de la première ligne ou la première colonne de coefficients AC

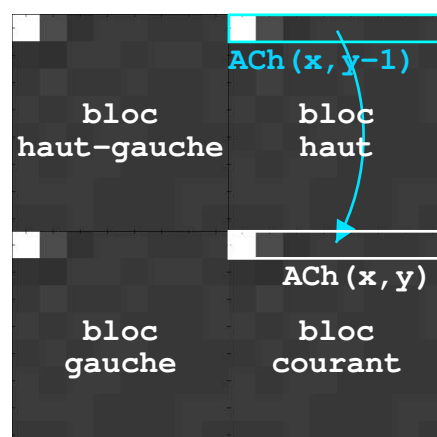


$$\text{si } |DC(x, y - 1) - DC(x - 1, y - 1)| < |DC(x - 1, y) - DC(x - 1, y - 1)|$$

prédiction des coefficients AC verticaux

Amélioration du codage d'image

- Prédiction de la première ligne ou la première colonne de coefficients AC



$$\text{si } |DC(x, y - 1) - DC(x - 1, y - 1)| > |DC(x - 1, y) - DC(x - 1, y - 1)|$$

prédiction des coefficients AC horizontaux

~ 10 % de gain en compression

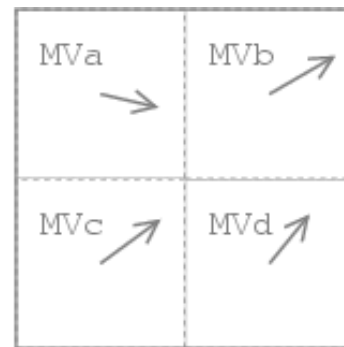
Amélioration du codage du mouvement

Le mouvement des objets peut contenir des zooms ou des rotations

- ↪ Un seul vecteur de mouvement de translation par macrobloc insuffisant
- ↪ Possibilité de spécifier un vecteur de mouvement par bloc (4 par MB)



mode inter

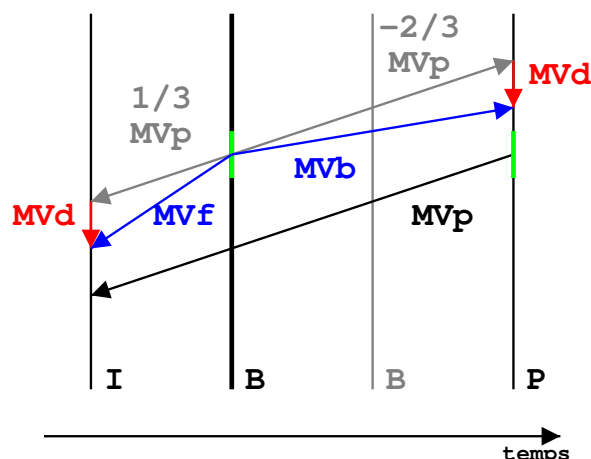


mode inter4v

Amélioration du codage du mouvement

La spécification de 2 vecteurs de mouvement pour les blocs bidirectionnels est coûteuse en débit.

- ↪ Introduction d'un nouveau mode de prédiction *directe*
- ↪ Prédiction par le vecteur dans l'image P précédente



$$MV_f = \frac{1}{3} MV_p + MV_d$$

$$MV_b = -\frac{2}{3} MV_p + MV_d$$

Recherche conjointe du meilleur vecteur MV_d

- ↪ Seul MV_d est transmis

Recherche au quart de pixel

Pour certaines applications où le mouvement est faible (vidéoconférence), une plus grande précision du mouvement permet une meilleure prédiction

↪ Recherche au quart de pixel

- Suréchantillonnage x4
- Interpolation à l'aide d'un filtre RIF :

$$\frac{1}{256}[-8, 24, -48, 160, 160, -48, 24, -8]$$

- Surcoût d'1 bit par composante de vecteur de mouvement
- Coûteux en temps de calcul

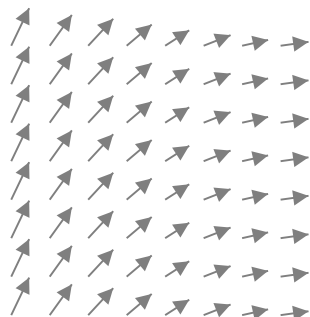
Compensation globale du mouvement

Compenser le mouvement de la caméra de manière globale

↪ modèle de mouvement dense paramétrique

- translationnel (2 paramètres)
- isotrope (4 paramètres)
- affine (6 paramètres)

Sélection du modèle global ou local pour chaque macrobloc



champ de mouvement dense

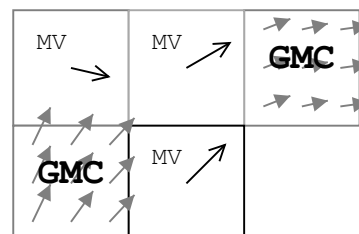


image courante

Scalabilité

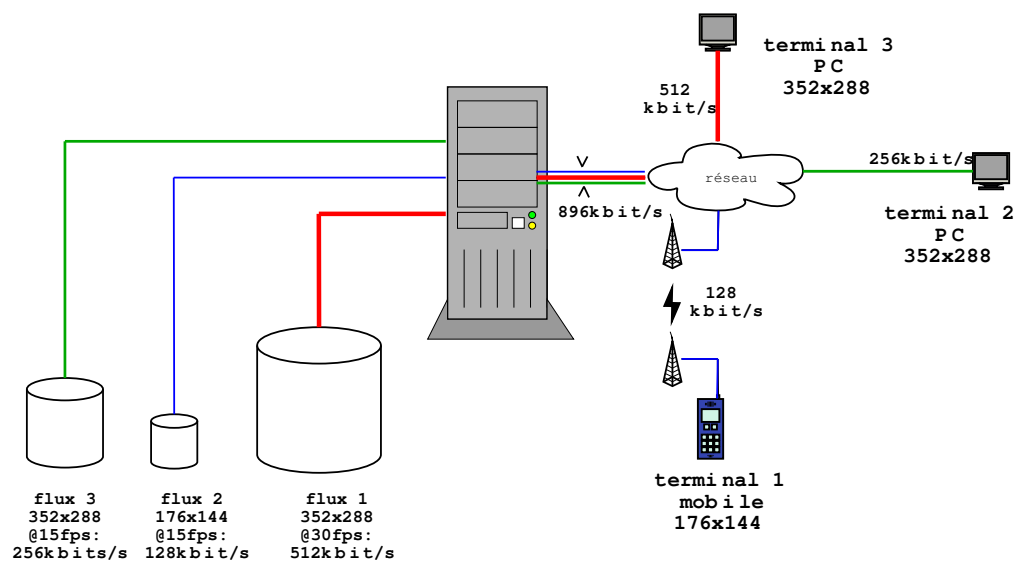
Codage progressif permettant

- d'adapter le débit à la volée **sans réencodage**
- un décodage sur des terminaux différents (PDA, PC, ...)

- ↪ Adaptation du nombre d'images par seconde (temporelle)
- ↪ Adaptation de la résolution (temporelle)
- ↪ Adaptation de la qualité (SNR)

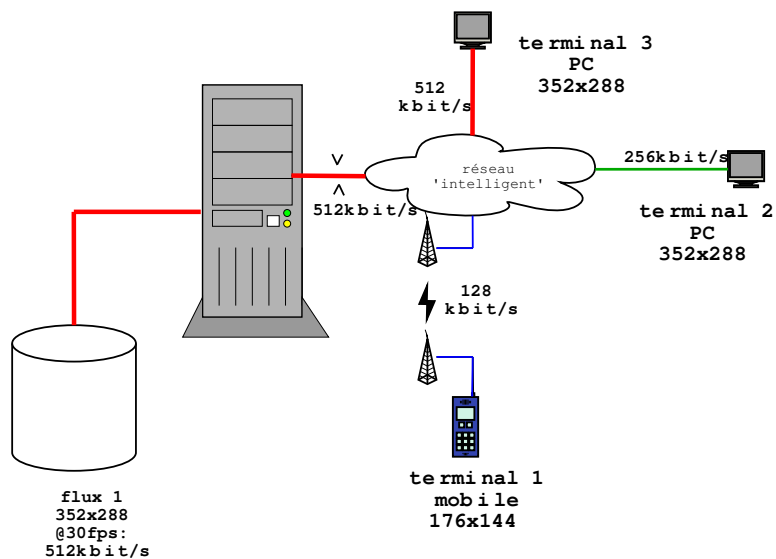
Scalabilité

Cas non-scalable



Scalabilité

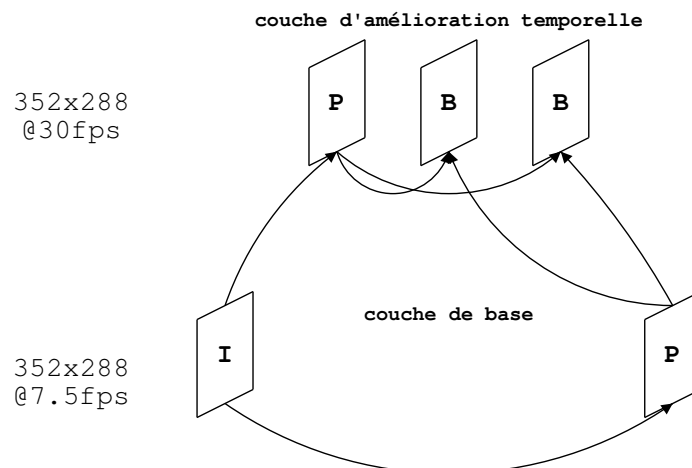
Cas scalable



Scalabilité temporelle

Adaptation du nombre d'images par seconde

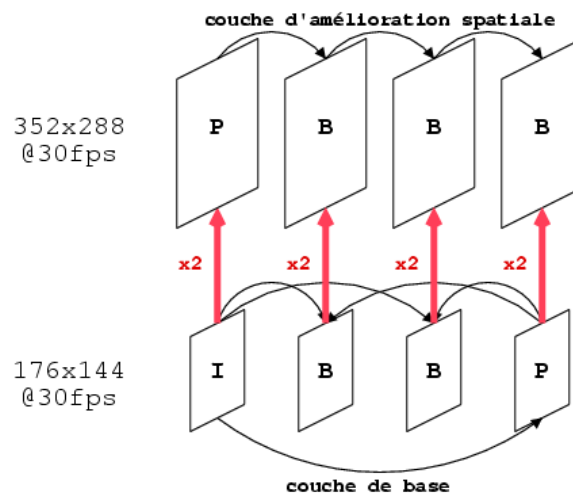
- Les images B peuvent être supprimées directement
- Flux vidéo encodé en *couches* de scalabilité temporelle



Scalabilité spatiale

Adaptation de la résolution

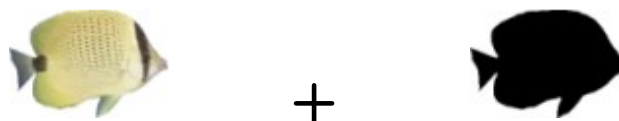
- Flux vidéo encodé en *couches* de scalabilité spatiale
- Interpolation pour obtenir un prédicteur **au même instant** à une résolution supérieure



Objets vidéo

Codage d'objets vidéo de forme arbitraire

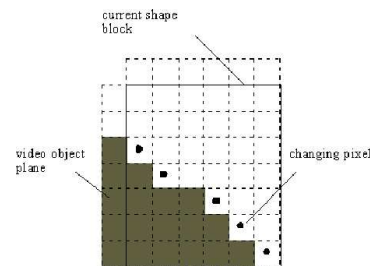
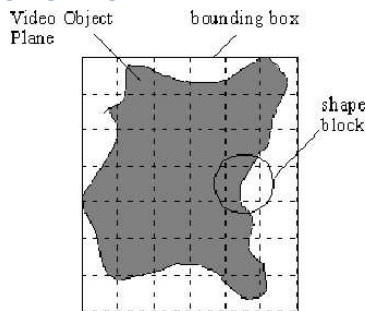
↪ spécification d'un masque de segmentation



- MPEG-4 ne prescrit pas comment segmenter les objets
⇒ Dépend de l'application
- Sachant que :
 - Dans certain cas, la segmentation peut être faite
(*trouver et suivre un visage dans une scène de visiophone*).
 - Dans d'autres cas, la segmentation existe de facto (*chroma key*).

Vidéo de forme quelconque

Codage de forme



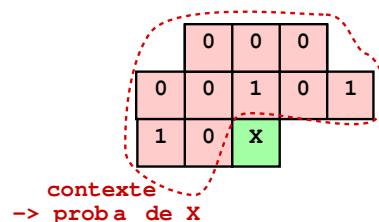
- description d'une *boîte englobante*
- codage du masque des blocs partiellement opaques (*bloc de forme*)

Vidéo de forme quelconque

Codage de forme

Codage arithmétique binaire contextuel

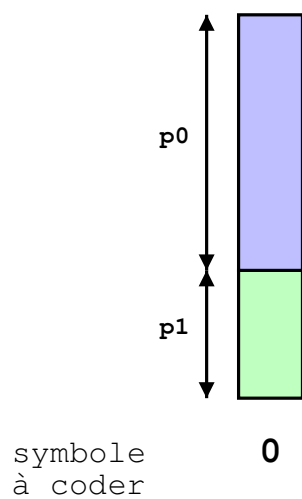
- ↪ **binaire** : masque d'opacité (**1**) ou de transparence (**0**) des pixels
- ↪ **contextuel** : la probabilité qu'un pixel soit opaque est estimée par le masque dans un voisinage causal indexant une table fixe de probabilité



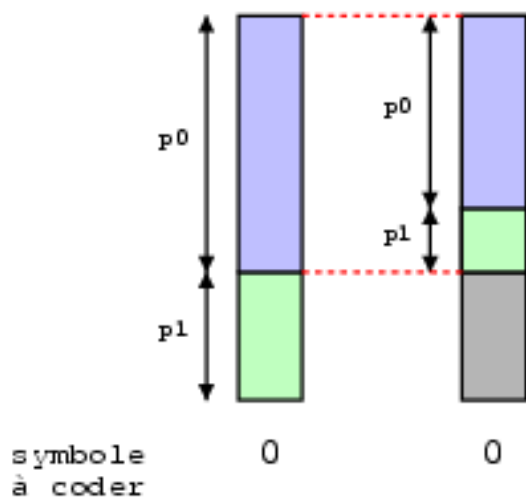
- ↪ **codage arithmétique** : séparation successive d'un intervalle en fonction des probabilités des symboles permettant d'atteindre l'entropie



Codage arithmétique

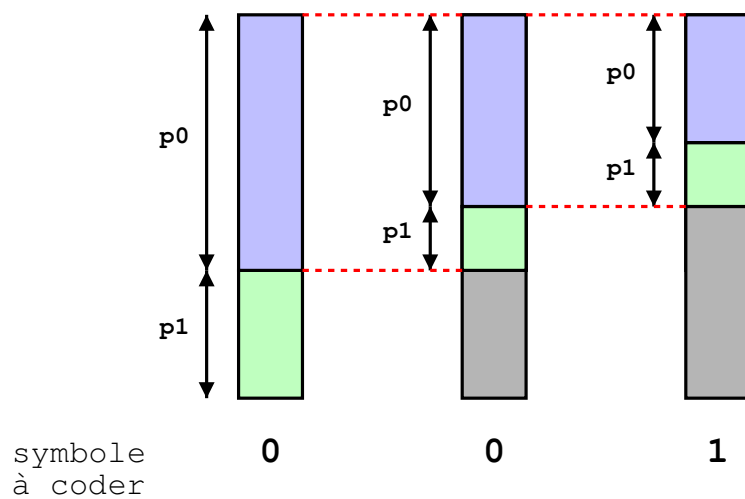


Codage arithmétique

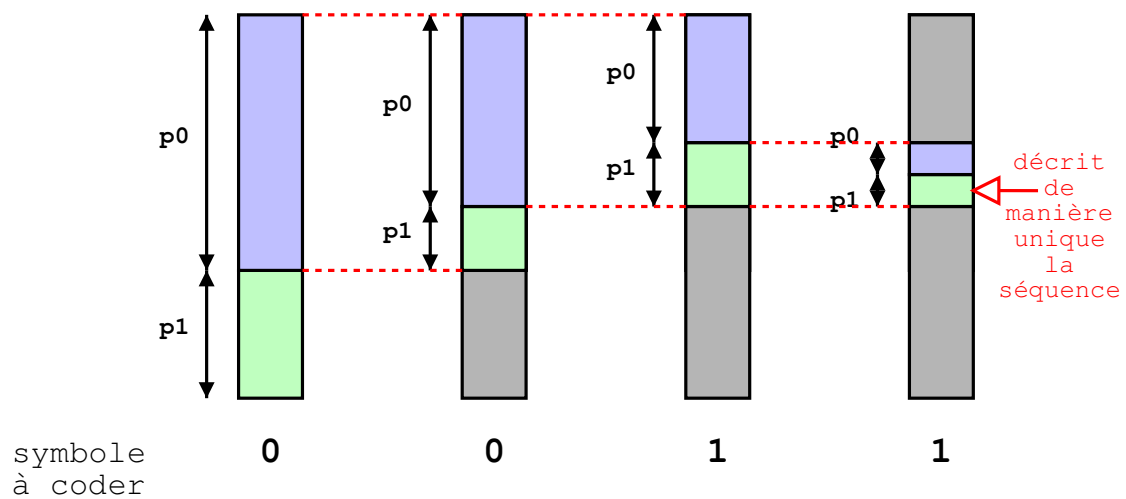




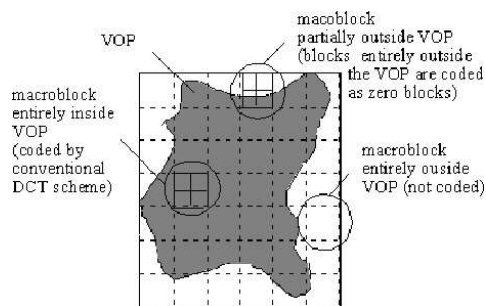
Codage arithmétique



Codage arithmétique

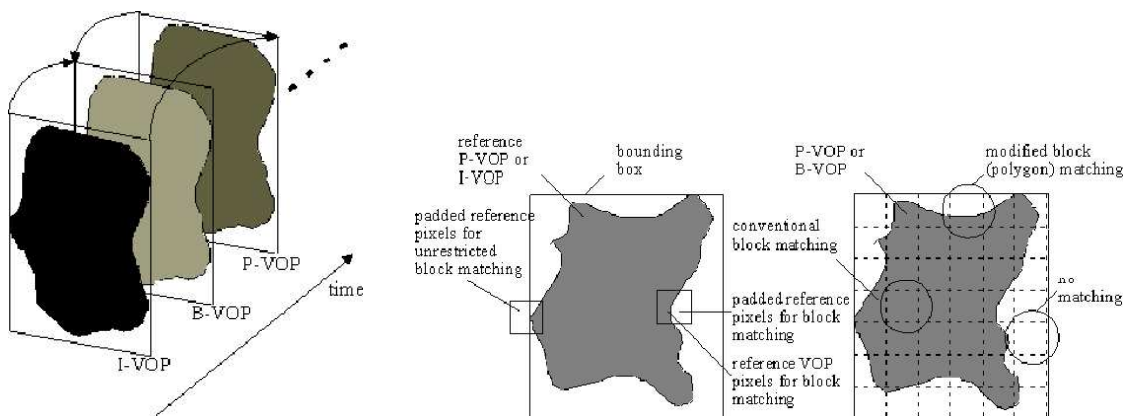


Codage des textures



- blocs totalement transparents non codés
- blocs totalement opaques codés classiquement
- blocs de forme codés après remplissage (*padding*)

Codage de mouvement



- remplissage de l'image de référence en dehors du masque (*padding*)
- blocs matching modifié ne prenant en compte que l'erreur sur pixels à l'intérieur du masque lors de la recherche de mouvement



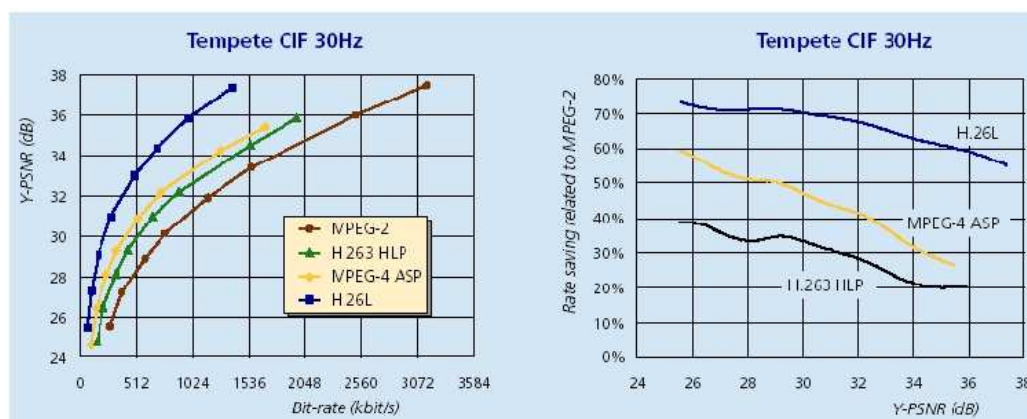
H.264 / MPEG-4 AVC

Dernière norme de codage vidéo :

- Utilisation d'un codeur arithmétique adaptatif contextuel (CABAC) pour le codage entropique (au lieu de RLE-Huffman)
- Blocs de taille variable : 4x4, 8x4, 4x8, ..., 16x16
- Optimisation débit distortion
- Prédiction temporelle à références multiples (plusieurs images dans le passé)



H.264 / MPEG-4 AVC



- + Gain en compression jusqu'à un facteur 2 comparé à MPEG-4 !!
- Pas encore de scalabilité, codage de forme, etc...

MPEG-4 : démultiplexage (principe)

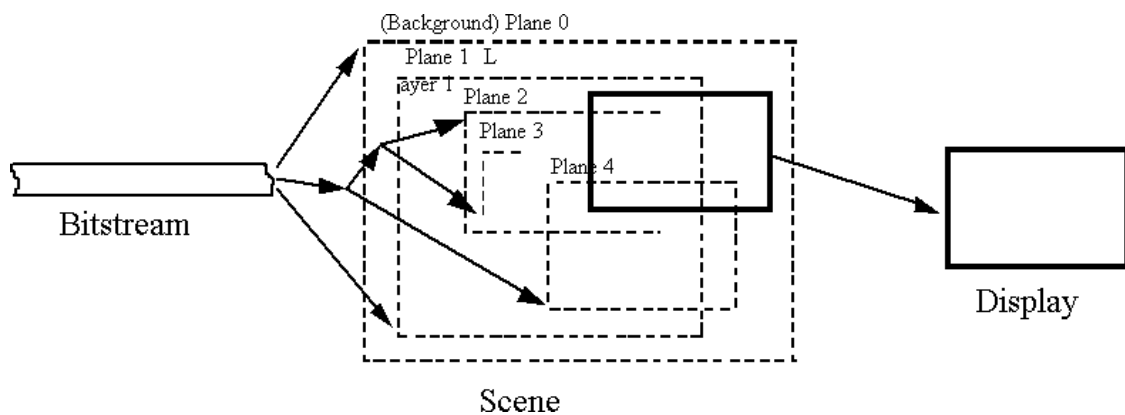
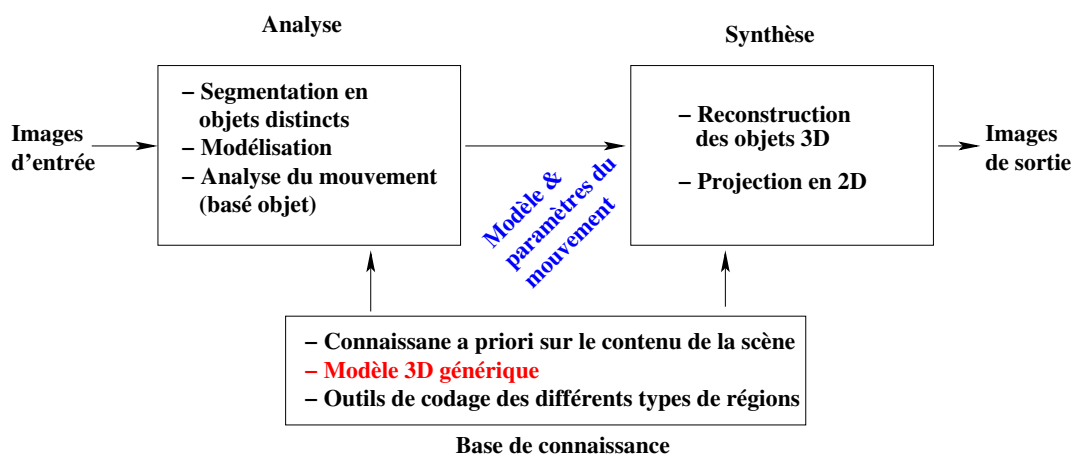
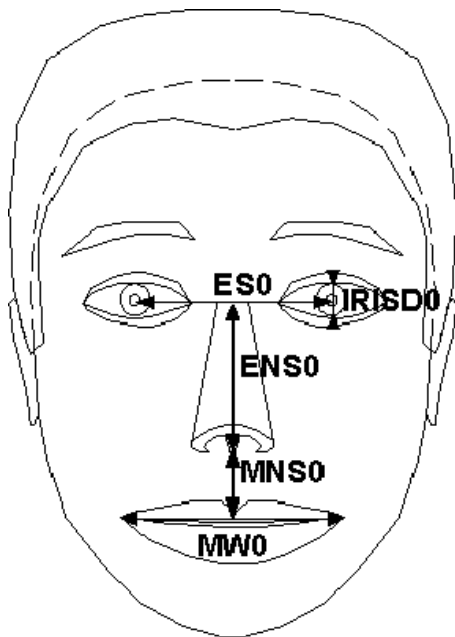


Schéma de codage basé modèle



Le modèle MPEG-4 / SNHC

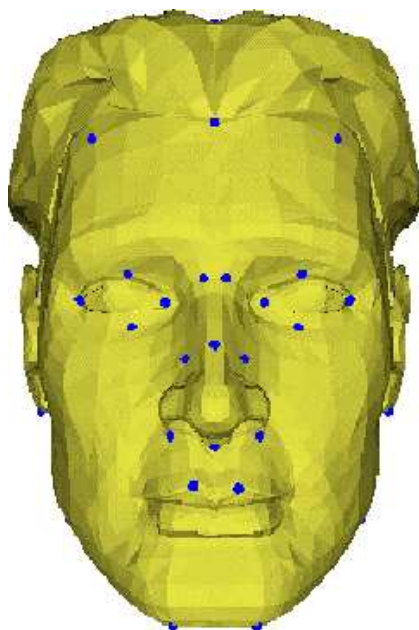


↪ Paramètres de Définition Faciale :
maillage 3D, texture, attributs.

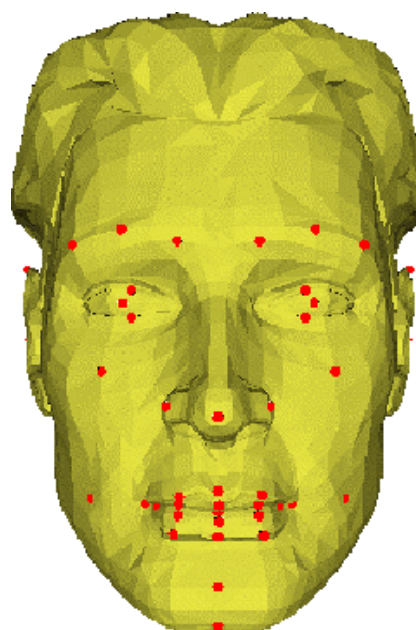
↪ Paramètres d'Animation Faciales :

- 68 paramètres dans 10 classes.
- FAPs exprimés en FAPUs
- Posture, expression,
- Mâchoire, menton lèvres, yeux, sourcils, joues, langue, nez, oreilles et rotation de la tête

Le modèle MPEG-4/SNHC (suite)



Facial Definition Parameters



Facial Animation Parameters

Représentation des objets audiovisuels

- ↪ Une scène se compose de différents objets audiovisuels organisés par hiérarchie. Primitives :
 - Images fixes (un fond de scène)
 - Objets vidéo (une personne qui parle)
 - Objets audio (la voix de la personne)
 - ...
- ↪ MPEG-4 standardise les primitives « naturelles » précédentes
- ↪ MPEG-4 définit une représentation codée d'objets AV tels que :
 - Textes et graphiques
 - Modèles synthétiques de visage et texte associé utilisé pour synthétiser la parole et animer le visage
 - Son synthétique

Tout objet AV est représenté indépendamment de son environnement et de son contexte

MPEG-4 : Messages

- ↪ Le standard des communications multimédia est disponible après plusieurs années de développement
- ↪ Il n'est pas destiné à remplacer les architectures existantes
 - images I proches de la norme d'image JPEG
 - MPEG-4 repose sur les principes de MPEG-1 améliorés et étendus pour le codage de forme, la scalabilité

→ respect des investissements industriels (puces de codage/décodage, etc..)
- ↪ Il dépasse les points forts des architectures existantes à l'aide de nouveaux concepts (ex. Description de scène interactive, APIs Java).
- ↪ Il prolonge les points forts des architectures vers de nouveaux environnements (ex. représentation du contenu indépendant de l'acheminement).



MPEG-4 : Messages

On constate un intérêt mitigé pour les nouvelles fonctionnalités de MPEG-4.

Le standard s'oriente vers de nouvelles approches :

- ↗ Advanced Video Coding (MPEG-4 AVC/H.264) (*blocs*)
- ↗ Scalable Video Coding (*ondelettes*)
 - rupture technologique