



Université du Québec

École de technologie supérieure

BD Vidéo et Audio

Plan

- Rappel des principes de compression vidéo
- Modèles de représentation d'images vidéo
- Normes de compression vidéo

Compression vidéo MPEG

- La compression MPEG (*Moving Picture Experts Group*) permet la compression de la vidéo, elle est réalisée avec perte.
- La compression MPEG reprend une bonne partie des spécifications de la norme JPEG (*Joint Photographic Experts Group*) et lui ajoute des traitements propres à l'image animée.
- Une image de référence entière est conservée, et pour les images successives, on ne conserve que les modifications.
- MPEG-1 restitue une image avec une résolution de 352 pixels et 282 lignes, à raison de 25 images par seconde, ou 352 x 240 à 30 images par seconde.
- MPEG-2 offre un débit plus élevé (plus de 20 Mbit/s). La résolution des images peut atteindre 1920 x 1080 à raison de 60 images par seconde.

Calcul du taux de compression des images vidéo

$$\left[\frac{\text{largeur} * \text{hauteur} * \text{profondeur} * \text{fps}}{\text{Facteur de compression}} \right] = \text{bits/sec}$$

largeur ~ pixels (160, 320, 640, 720, 1280, 1920)

hauteur ~ pixels (120, 240, 480, 485, 720, 1080)

profondeur ~ bits (1, 4, 8, 15, 16, 24)

fps ~ frames par seconde (5, 15, 20, 24, 30)

Facteur de compression (1, 6, 24)

Exemple d'applications

- Applications de divertissement
 - Diffusion d'émissions de télévision
 - Enregistrements
- Applications interpersonnelles
 - Visiophonie et vidéoconférence
- Applications interactives
 - Des fenêtres contenant des clips vidéo

Format de diffusion télévision

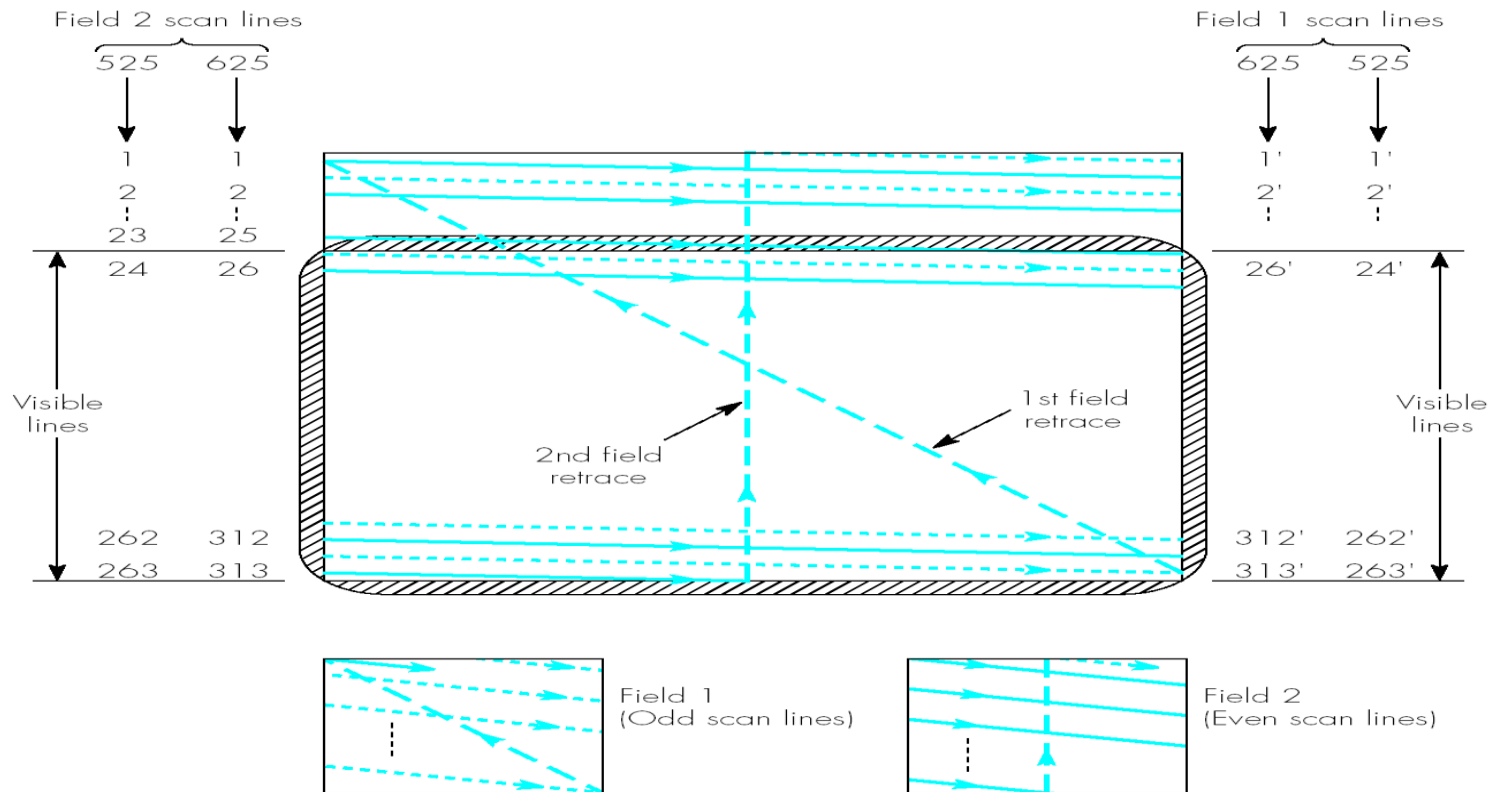
- Résolution spatiale et temporelle

- NTSC: 525 lignes
- PAL/CCIR/SECAM: 625 lignes
- Rafraîchissement de 60 à 50 trames par seconde

- Séquence de balayage

- Le taux de rafraîchissement peut être réduit à 25 trames par seconde
- Transmission des images en mode entrelacé pour réduire la largeur de bande
 - NTSC: 262.5 lignes dont 240 visibles
 - PAL: 312.5 lignes dont 288 visibles

Principe de balayage entrelacé



525-line systems : 262.5 each field, 240 visible
 625-line systems : 312.5 each field, 288 visible

Tirée de Halsall (2000)



Université du Québec

École de technologie supérieure

Département de génie logiciel et des TI

Perception de la couleur

- L'œil est plus sensible aux différences d'intensités (luminosité) que de couleurs.
- Le vert porte le maximum de luminosité; le blanc est le résultat de l'addition de 59% de vert, 30% de rouge et seulement 11% de bleu.
- Transformation du code couleur RGB au code YCC
 - Y représente la luminosité obtenue par $0.30R + 0.59V + 0.11B$
 - Mesure compatible pour la télévision monochrome
 - Le premier C, appelé Cb, est la différence entre la valeur de B et la valeur de Y.
 - Le deuxième C, appelé Cr, est la différence entre la valeur de R et celle de Y.
- Le code YCbCr (proche du système HSB - Hue, Saturation, Brightness) est utilisé pour les images de télévision pour lesquelles l'exigence de fidélité est inférieure à celle d'une image fixe.

Transmission des images couleurs

- La largeur de bande est la même pour la diffusion monochrome et couleur
 - Sous échantillonnage des signaux de couleurs pour éviter la dégradation de la diffusion monochrome
- Diffusion PAL
 - $Y = 0.299 R + 0.587 G + 0.114 B$
 - $U = 0.493 (B - Y)$
 - $V = 0.877 (R - Y)$
- Diffusion NTSC
 - $Y = 0.299 R + 0.587 G + 0.114 B$
 - $I = 0.74 (R - Y) - 0.27 (B - Y)$
 - $Q = 0.48 (R - Y) + 0.41 (B - Y)$

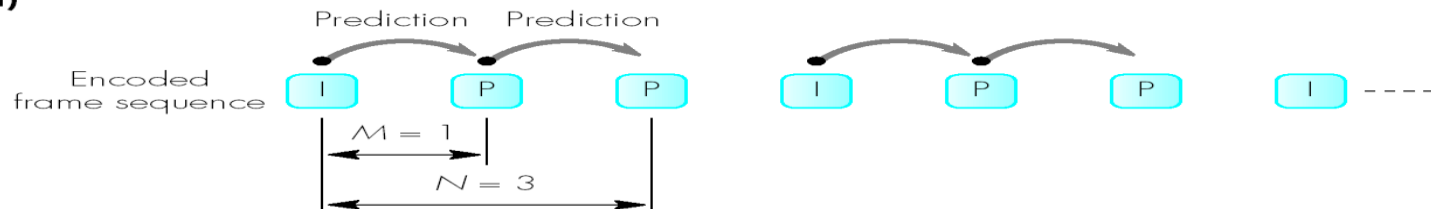
Interpolation d'images

- Principe: à partir d'une image du passé et une image du futur, reconstruire les images intermédiaires par interpolation
- Les blocs d'images intermédiaires peuvent fournir des prédictions bidirectionnelles (avant et arrière)
- Trois grands types d'images pour établir un compromis entre le besoin d'accès direct et l'efficacité de compression:
 - Les images à codage intra (images I): codées sans référence aux autres images. Elles fournissent des points d'accès à la séquence mais leur taux de compression est modéré (10:1 à 20:1)
 - Les images codées par prédiction (images P): codage plus efficace (20:1 à 30:1) utilisant une prédiction compensée en mouvement, d'après une image antérieure, intra ou prédictive, et sont utilisées comme référence pour une prédiction future
 - Les images codées par prédiction bidirectionnelles (images B): taux de compression le plus élevé (30:1 à 50:1) mais nécessitent une image antérieure et future de référence. Elles ne peuvent pas être considérées comme référence de prédiction

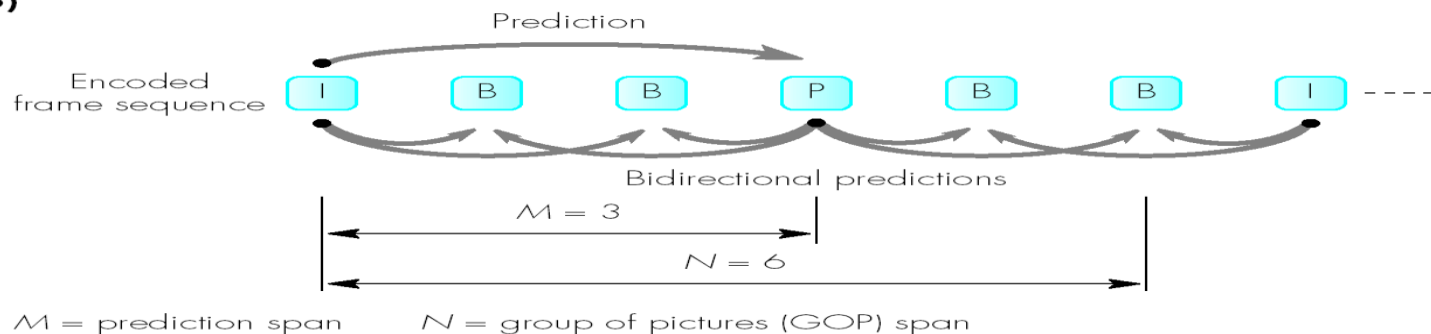


Exemples de séquences d'images interpolées

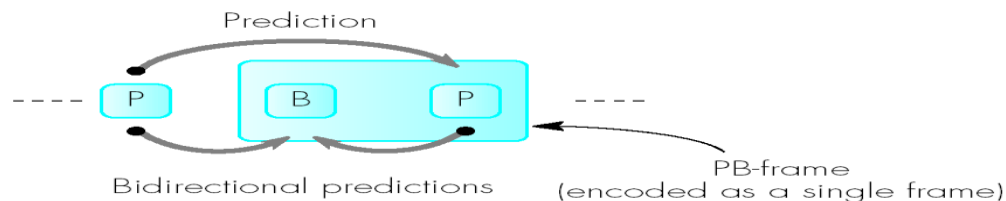
(a)



(b)

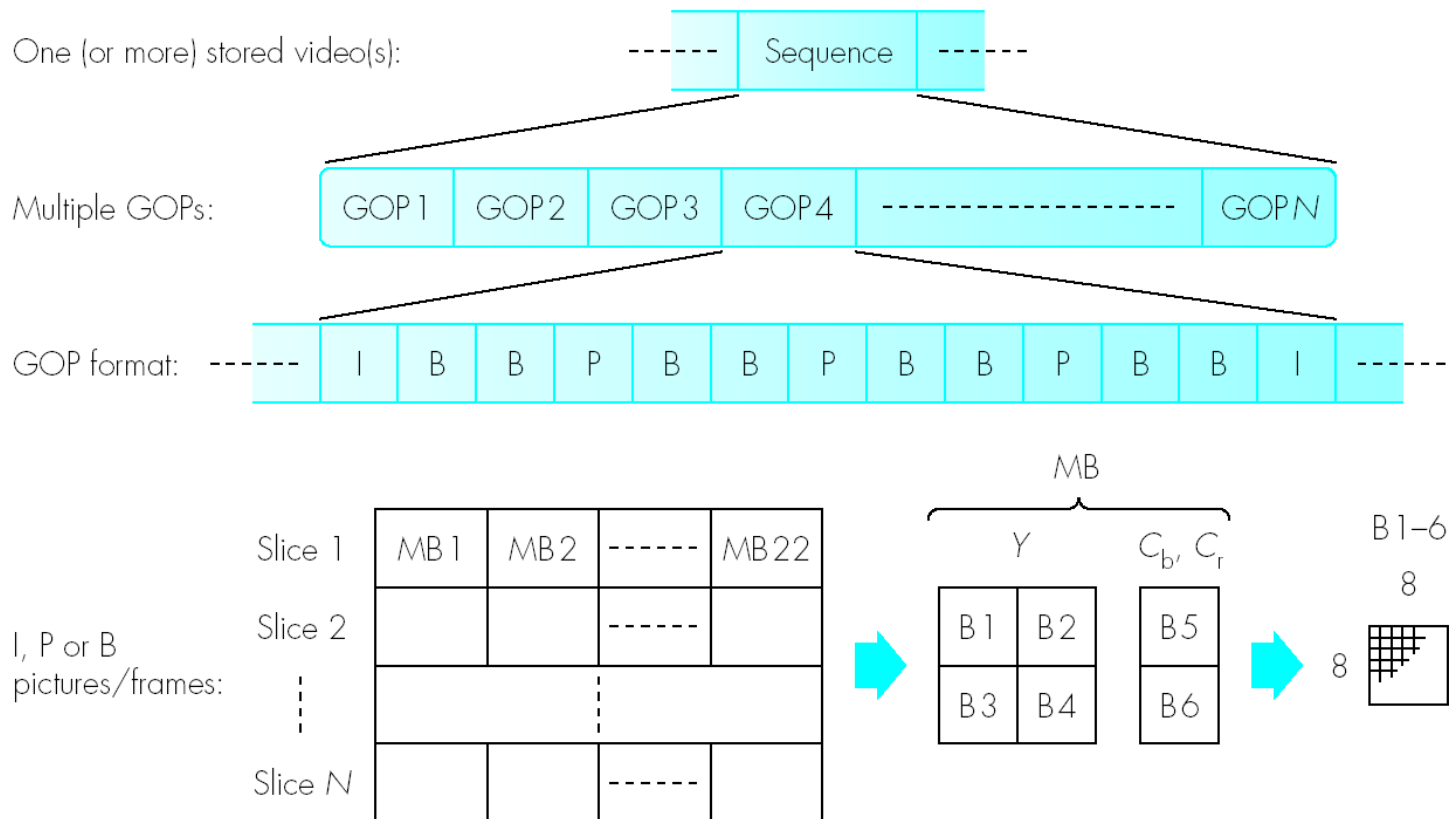


(c)



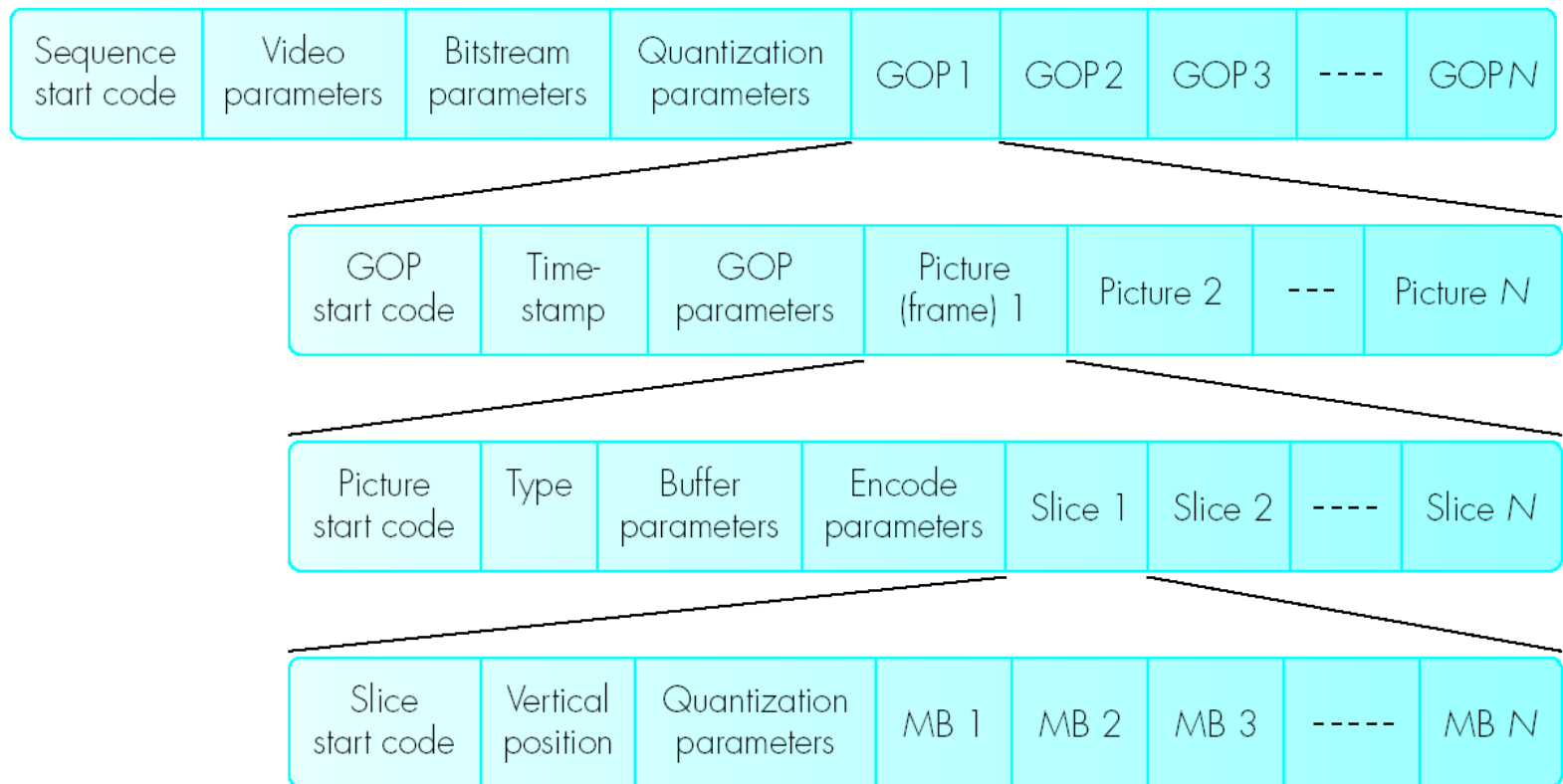
Norme MPEG1 (suite)

- Composition du train binaire



Norme MPEG1 (suite)

- Format du train binaire



Norme MPEG2

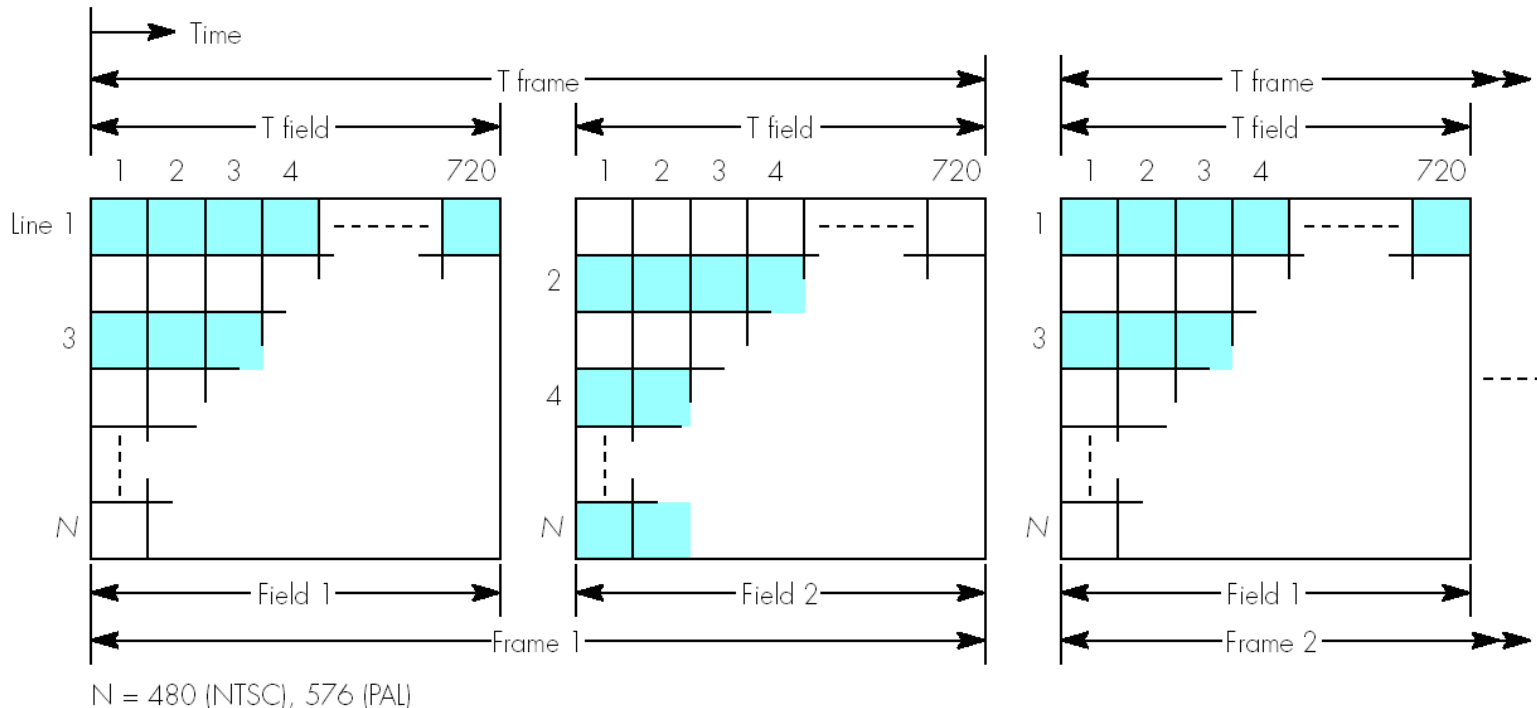
- La spécification MPEG2 est générique
 - Plusieurs profils et niveaux adaptés à des applications particulières
 - Fournit des méthodes en laissant le soin à l'utilisateur de les paramétrer
- Niveaux et profils dans MPEG2
 - 4 niveaux: low, main, high 1440 et high
 - 5 profiles: simple, main, spatial resolution, quantization accuracy and high
- Niveau main et profile main de MPEG2 (MP @ML)
 - Application de distribution de télévision numérique
 - Balayage entrelacé
 - Format 4:2:0

Norme MPEG2 (suite)

- MPEG2 repose sur une structure d'algorithme de compression hybride à base de transformées en cosinus discrète DCT (Discrete Cosine Transform) et de prédiction temporelle compensée en mouvement
- MPEG2 permet de nombreux modes d'encodage supplémentaires
 - Prédiction inter-image ou une prédiction intra-trame pour les signaux 4:2:0, 4:2:2 et 4:4:4 entrelacés
 - Transformation DCT intra-image (après mélange des deux trames) ou intra-trame
 - Possibilités de redéfinir les matrices de quantification
- MPEG2 offre un mode d'extensibilité spatiale supplémentaire (*data partitioning*)
 - Transmettre sur la couche basse les coefficients DCT de basse fréquence
 - Les coefficients DCT de fréquences supérieures sont transmis sur les couches d'améliorations

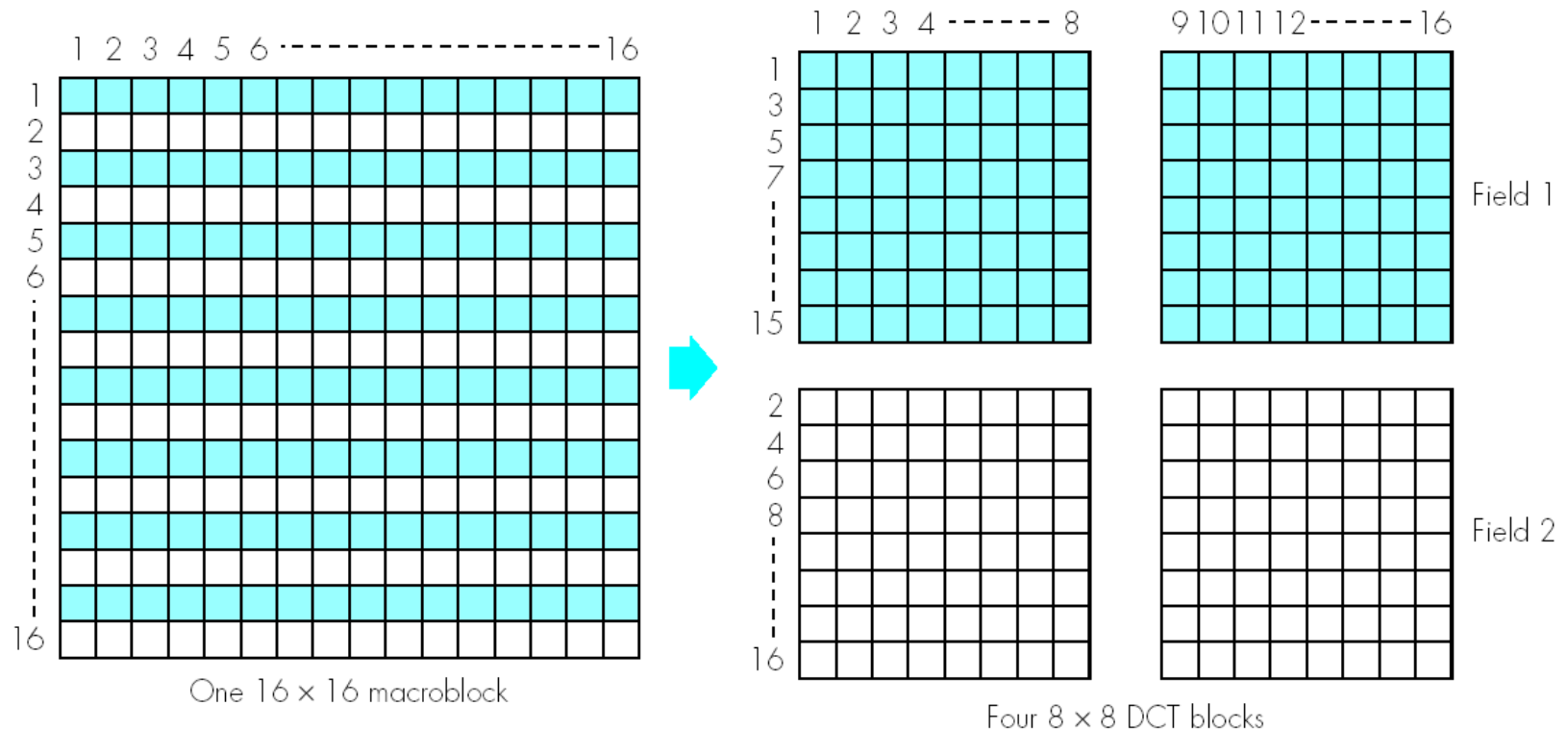
Norme MPEG2 (suite)

- Composition des blocs DCT pour le codage des images I



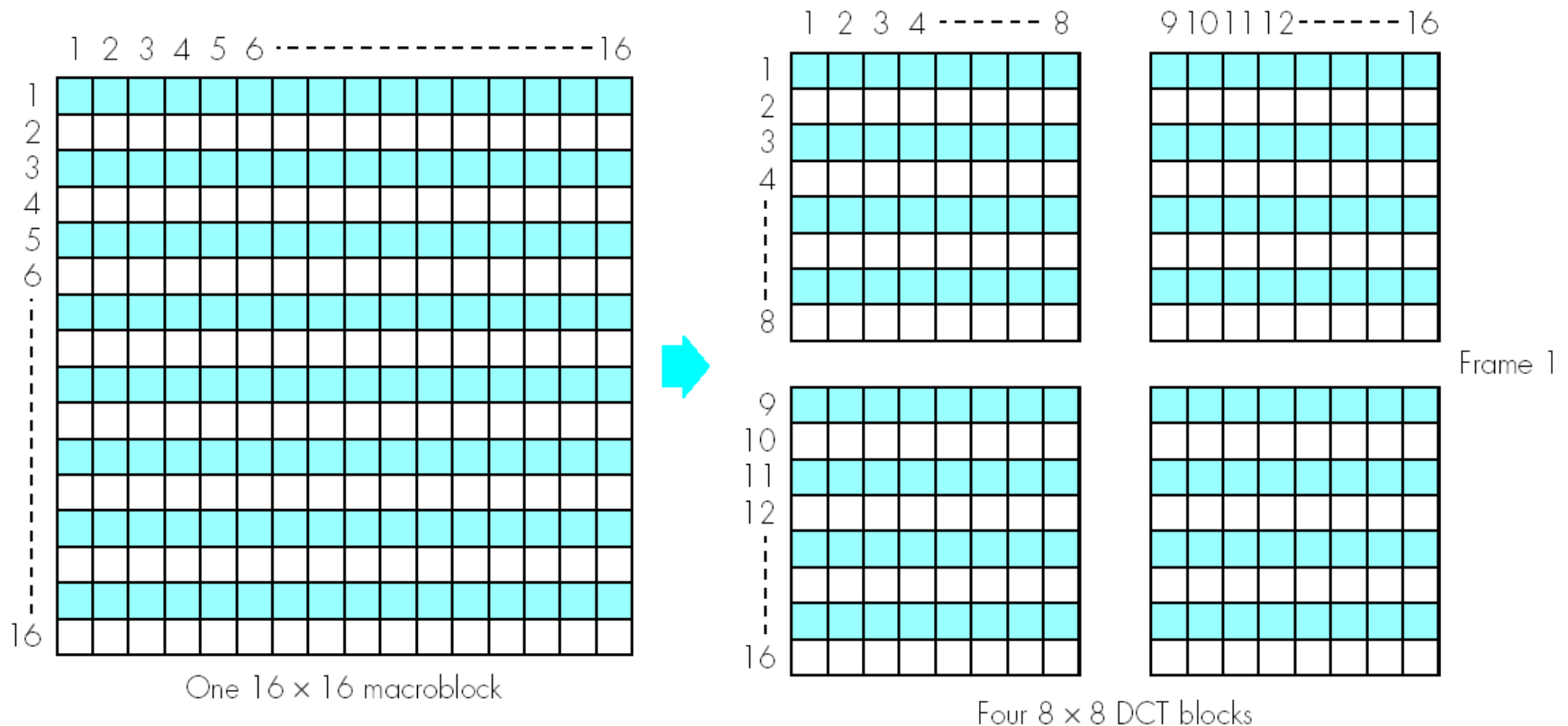
Norme MPEG2 (suite)

- Composition des blocs DCT en mode field



Norme MPEG2 (suite)

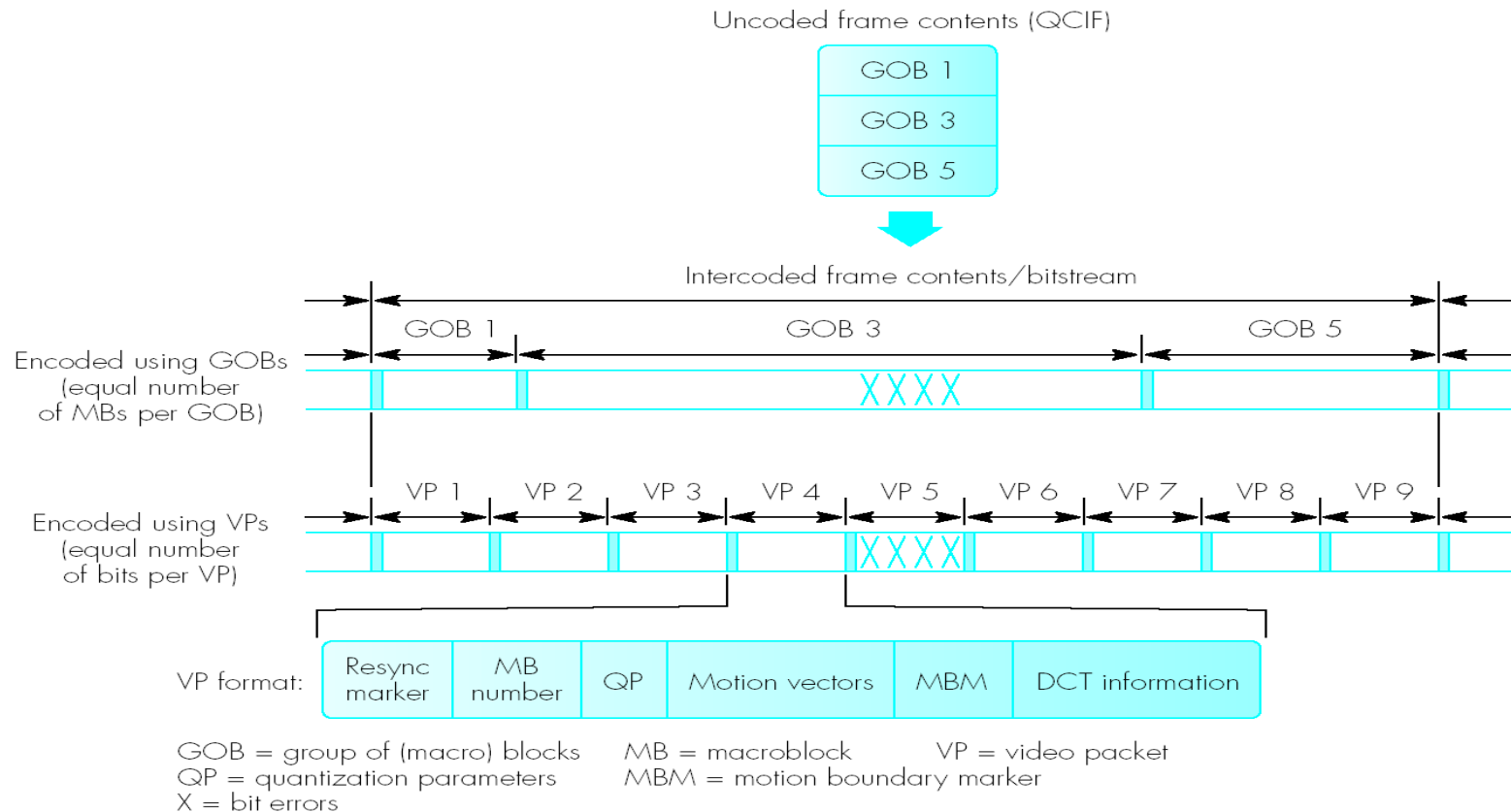
- Composition des blocs DCT en mode frame



Norme MPEG4

- La norme MPEG4 permet:
 - La prise en compte d'une grande variété de sources audio et vidéo naturelles et synthétiques
 - Modèles 3D synthétiques
 - Séquences vidéo naturelles
 - Parole synthétisée
 - Morceau de musique compressé
 - La prise en compte de la transmission des informations multimédias sur des supports hétérogènes en terme de débit
 - Réseaux mobiles: dizaines de kilobits/s
 - Réseaux câblés: plusieurs mégabits/s
 - Interaction entre l'utilisateur et le contenu des images (sous forme d'objets)
- MPEG4 repose sur un schéma de compression hybride, basé DCT et compensation de mouvement

Norme MPEG4 (suite)

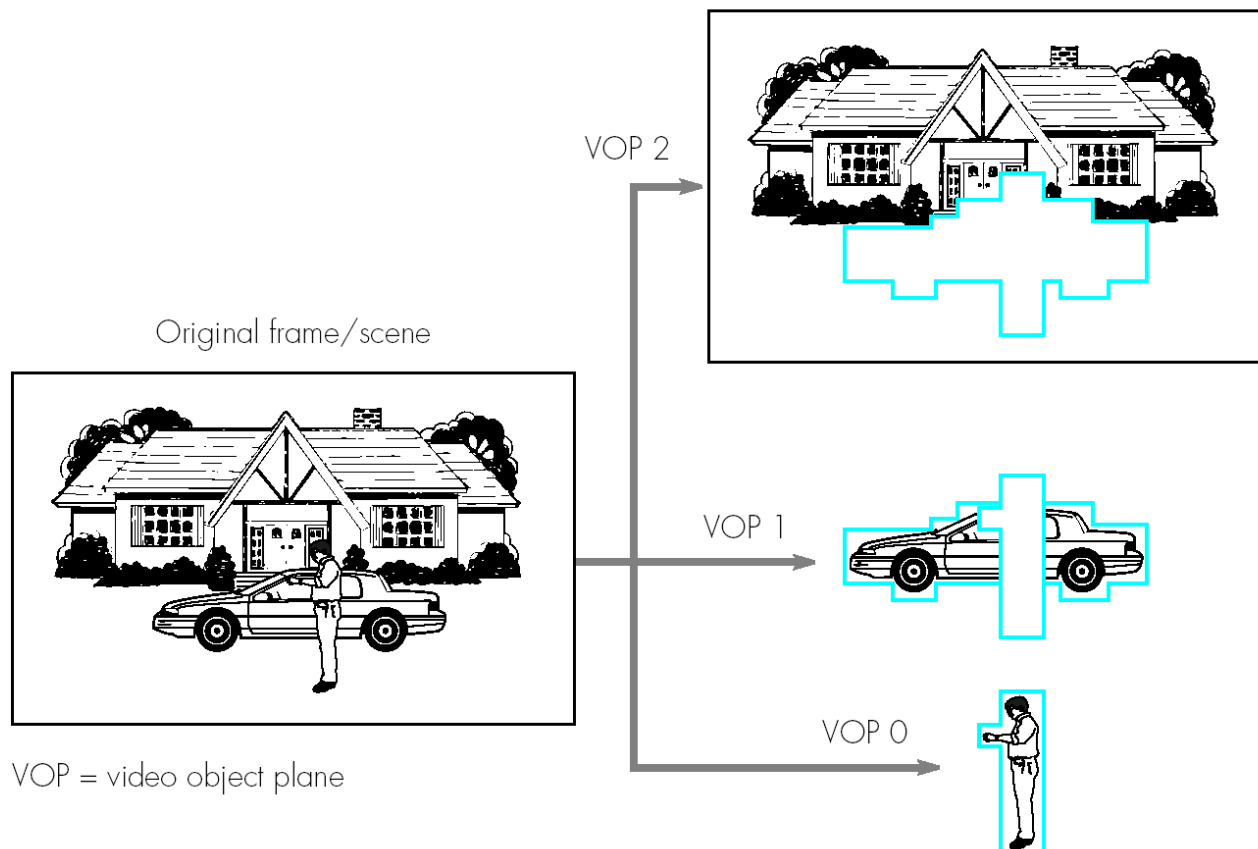


Norme MPEG4 (suite)

- MPEG4 définit une représentation compacte de l'audio, la vidéo, le texte et les graphiques sous forme d'objets audiovisuels
- Format de description de scènes BIFS (BInary description For Scene)
 - Format binaire compressé d'un langage textuel de description de scène inspiré de VRML
 - BIFS permet de décrire
 - Le positionnement spatio-temporel des différents objets
 - Les interactions possibles avec les objets ou leur animation
- MPEG4 représente les images comme une superposition d'objets vidéo plans VOP (Video Object Plane)

Norme MPEG4 (suite)

- Composition de scènes



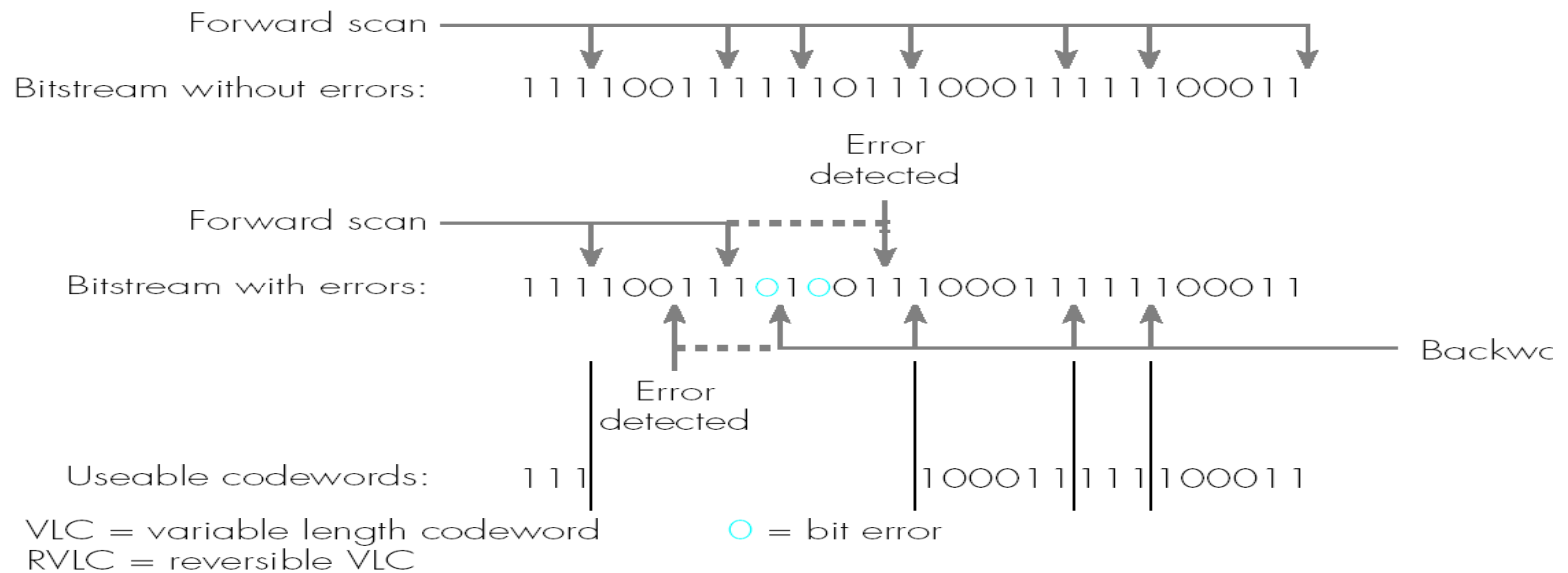
Norme MPEG4 (suite)

- MPEG4 améliore la robustesse aux erreurs
 - Les images sont découpées en paquets indépendants délimités par des marqueurs spécifiques
 - Possibilité de resynchronisation à l'intérieur même d'un VOP
 - Confiner les erreurs à l'intérieur d'un paquet
- Partition de données
 - Réorganisation interne des données des macroblocs
 - Un réarrangement de données est effectué par type: forme, mouvement et texture
- Codage à longueur variable réversible RVLC (Reversible Variable Length Coding)
 - Code symétrique
 - La réception d'un code non symétrique permet de détecter une erreur
 - Il permet aussi de corriger d'éventuelles erreurs lors du décodage

Norme MPEG4 (suite)

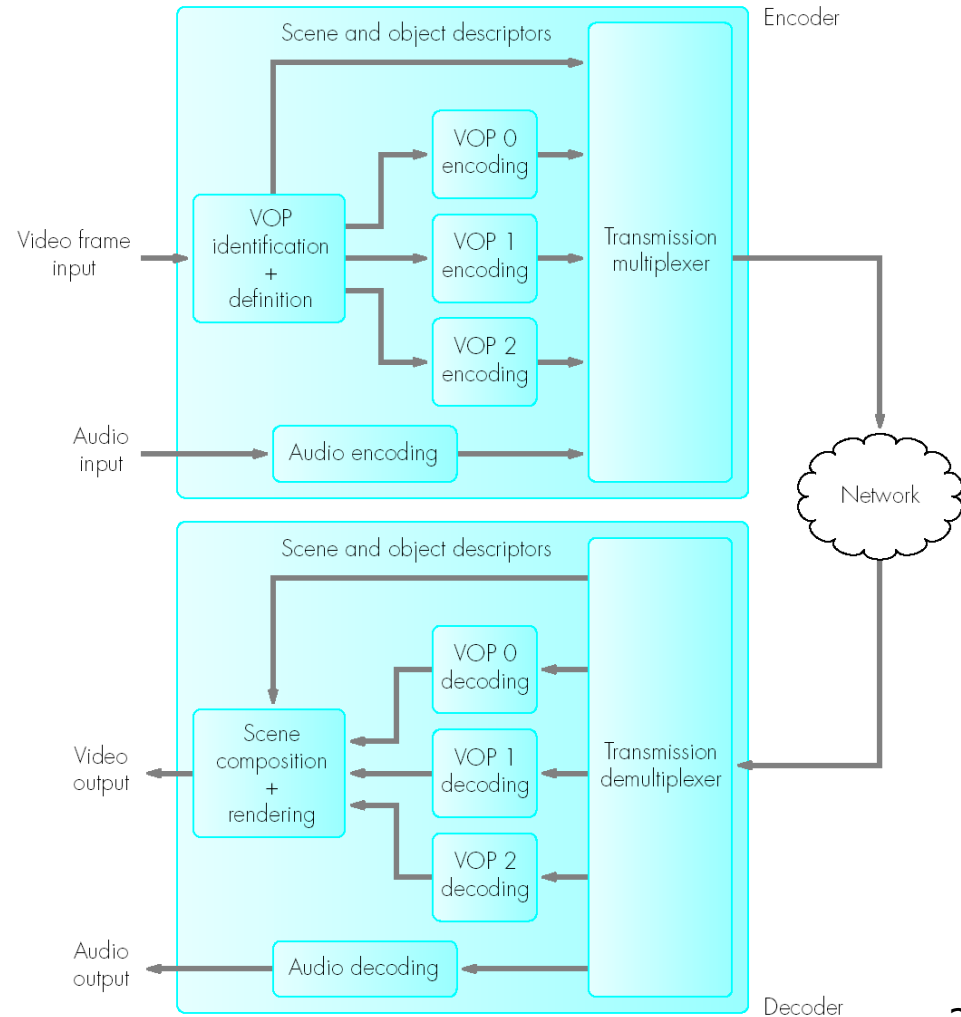
VLC	RVLC
1	111
01	1011
001	10011
0001	100011

Maximum codeword length = 6 bits



Norme MPEG4 (suite)

- Compression audio et vidéo dans MPEG4



Norme MPEG4 (suite)

- Le mode objet dans MPEG4

- Permet de coder des objets vidéo préalablement extraits par segmentation de la séquence originale
- Un VOP est le plus petit rectangle englobant l'objet considéré et contenant un nombre entier de macroblocs
- Codage de la texture et de la forme d'un VOP
- Postproduction vidéo, réalité augmentée, la réalité virtuelle, les jeux vidéo, les présentations multimédias...

- Codage de la texture

- Transformation DCT suivie de quantification scalaire et codage entropique
- Une technique de codage spécifique est définie pour les blocs frontières (technique de padding utilisée dans la version 1)
- Technique dite de Shape Adaptive DCT (SA-DCT) introduite dans la version 2

Norme MPEG4 (suite)

- Codage de la forme

- La composante de forme est codée macrobloc par macrobloc
- Le bloc contenant l'information de forme est appelé BAB (Binary Alpha Block)
- Plusieurs modes de codage
 - Codage par recopie du BAB
 - Codage par compensation de mouvement
 - Codage arithmétique intra
 - Codage arithmétique inter
 - Codage arithmétique inter avec compensation de mouvement

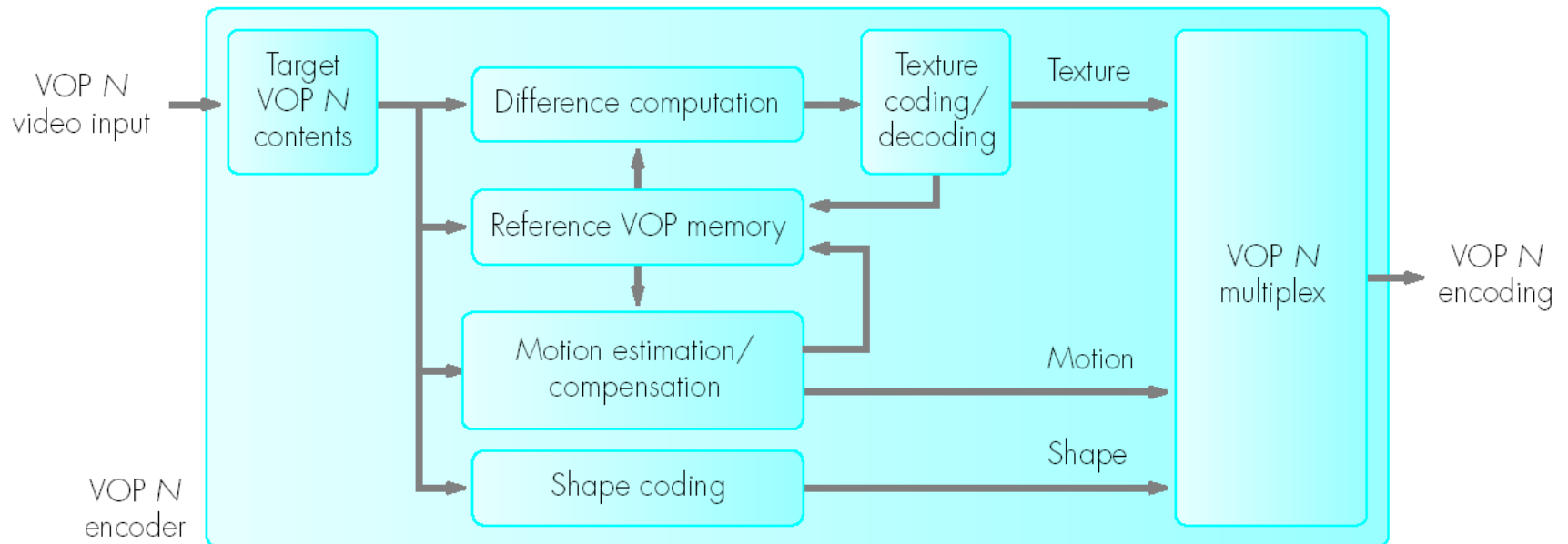
- Le mode sprite dans MPEG4

- Image panoramique (ou mosaïque) représentant une partie statique d'une séquence vidéo naturelle
- Deux modes principaux de codage: mode basique et le mode low-latency sprite qui permet une transmission progressive du sprite



Norme MPEG4 (suite)

- Codage des VOP



Plan de la présentation

- **Pourquoi extraire les traits d'une vidéo (12.1)**
- Analyse de vidéo et segmentation (12.2)
- Métadonnées d'audio et de vidéo (12.6, 12.7)
- Classification de vidéo
- Indexation de vidéo (12.7.1)
- Requêtes vidéo (12.9)
- *interMedia* pour vidéo et audio

Pourquoi extraire les traits d'un vidéo 12.1

- Repérage audio et vidéo
 - Traits basés sur les images
 - Traits basés sur les mouvements
- Détection d'objets et suivi du mouvement
- Reconnaissance de la parole / transcription textuelle
- Identification vocale
- Repérage de mots
- Classification d'audio

Plan de la présentation

- Pourquoi extraire les traits d'un vidéo (12.1)
- **Analyse de vidéo et segmentation (12.2)**
- Métadonnées d'audio et de vidéo (12.6, 12.7)
- Classification vidéo
- Indexation de vidéo (12.7.1)
- Requêtes vidéo (12.9)
- *interMedia* pour vidéo et audio

Segmentation manuelle

- Trouver un logiciel de segmentation
ex: http://www.boilsoft.net/download/avi_mpg_splitter.exe
- Faire les coupures en identifiant les frontières manuellement
- Enregistrer les temps de départs et d'arrêts

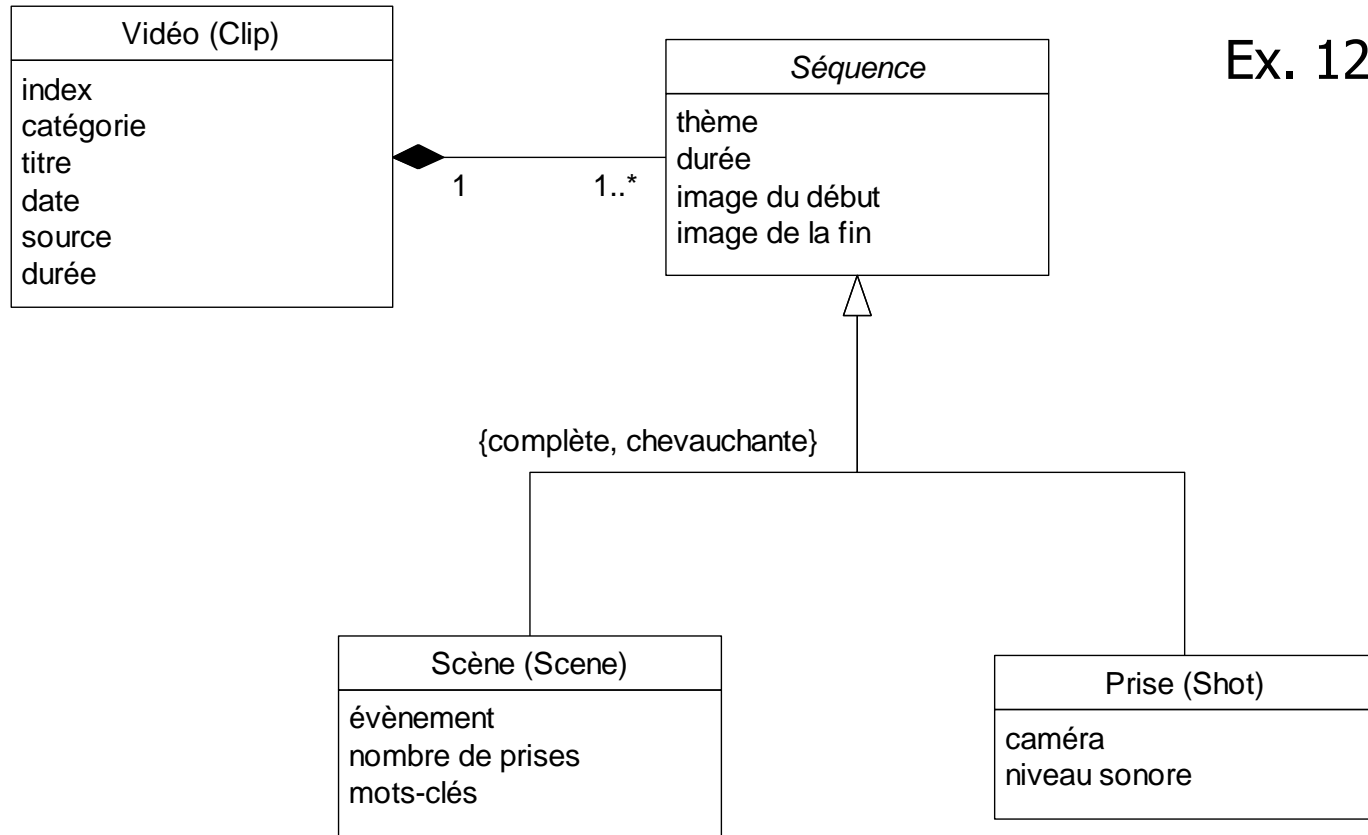
Analyse et segmentation automatique

- Techniques CBR d'images utilisées pour les vidéos
 - Segmenter le vidéo en objets plus petits
- Créer une hiérarchie d'images (frame), de scènes (scene) ou de prises (shot)
 - Images unitaires donnent un volume trop élevé
 - Bas niveaux syntaxiques
 1. la scène
 2. la prise
- Technique de segmentation la plus répandue
 - Détection automatique des transitions de prises

Terminologie vidéo

- **Vidéo** (clip) : succession d'images, regroupées en scènes ou en prises
- **Image** (frame) : une seule image
 - inefficace comme unité organisationnelle en BD Vidéo
- **Segment** ou **Scène** (scene) : regroupement logique des prises d'un évènement, d'un endroit
 - unité organisationnelle de base en BD Vidéo
- **Prise** (shot) : succession d'images capturées en continue par une seule caméra
 - unité fondamentale de travail en production de film

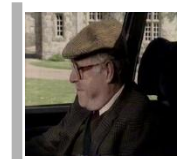
Organisation hiérarchique d'une vidéo



Exemple avec cette terminologie

de Harald Kosh

- image (frame)
- prise (shot)
- 2 prises
- scène (scene)
- audio



La détection de prise

- 2 types de transition de prise
 - coupure
 - graduelle : couverture, fondu
- 2 approches de technique de détection
 - détection sur une vidéo non compressé
 - détection sur vidéo compressé

Exemples de transition de prises



Coupure
(Cut)



Couverture
graduelle
(Wipe)



Fondu
(Dissolve)

Exemples de transition - Dissolve

- Fondu enchaîné : mélange de deux séquences vidéo où la première est graduellement estompée (fade-out) et la seconde apparaît graduellement (fade-in)



+



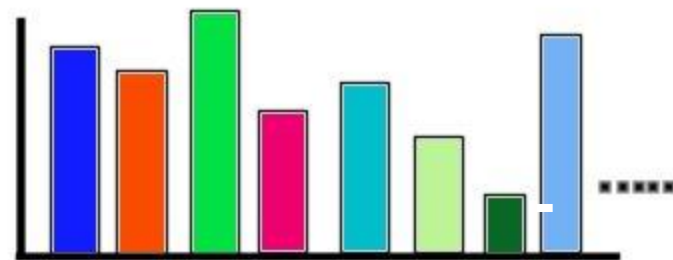
Méthode de détection de transition de prise

1. Par histogramme de couleur (non compressé)

- Calculer les pourcentages de couleur par image
- Comparer ces pourcentages entre images adjacentes
- Si différentielle moyenne $>$ seuil
 - changement de prise
 - sinon, retenir la différentielle



Histogramme de couleur



Méthode de détection de transition de prise

2. Par détection de la forme (non compressé)

- Algorithme de détection des contours appliqué aux images
- Comparer les vecteurs entre images adjacentes
- Si différentielle moyenne $>$ seuil
 - changement de prise
 - sinon, retenir la différentielle



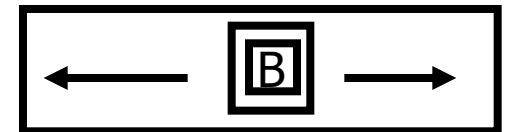
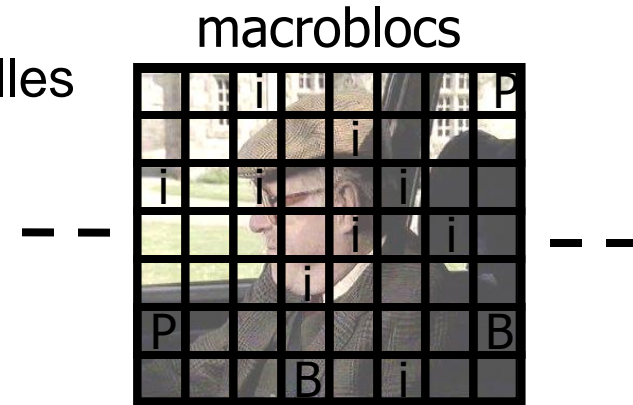
Méthode de détection de transition de prise

3. Par macrobloc (compressé)

- MPEG: format compressé encodé par différentielles
- Images segmentées en macroblocs

3 types de macrobloc

1. type I (Intra-coded): chaque macrobloc est encodé indépendamment des autres
 2. type P (Predicted) : encode le vecteur de mouvement et la différentielle par rapport à l'image précédente
 3. type B (Bidirectional) : encode le vecteur de mouvement et la différentielle par rapport à l'image précédente ou suivante
- Détection de changement de prise selon le nombre de macroblocs qui changent



Cut detection with Motion Vector and Macro block statistics

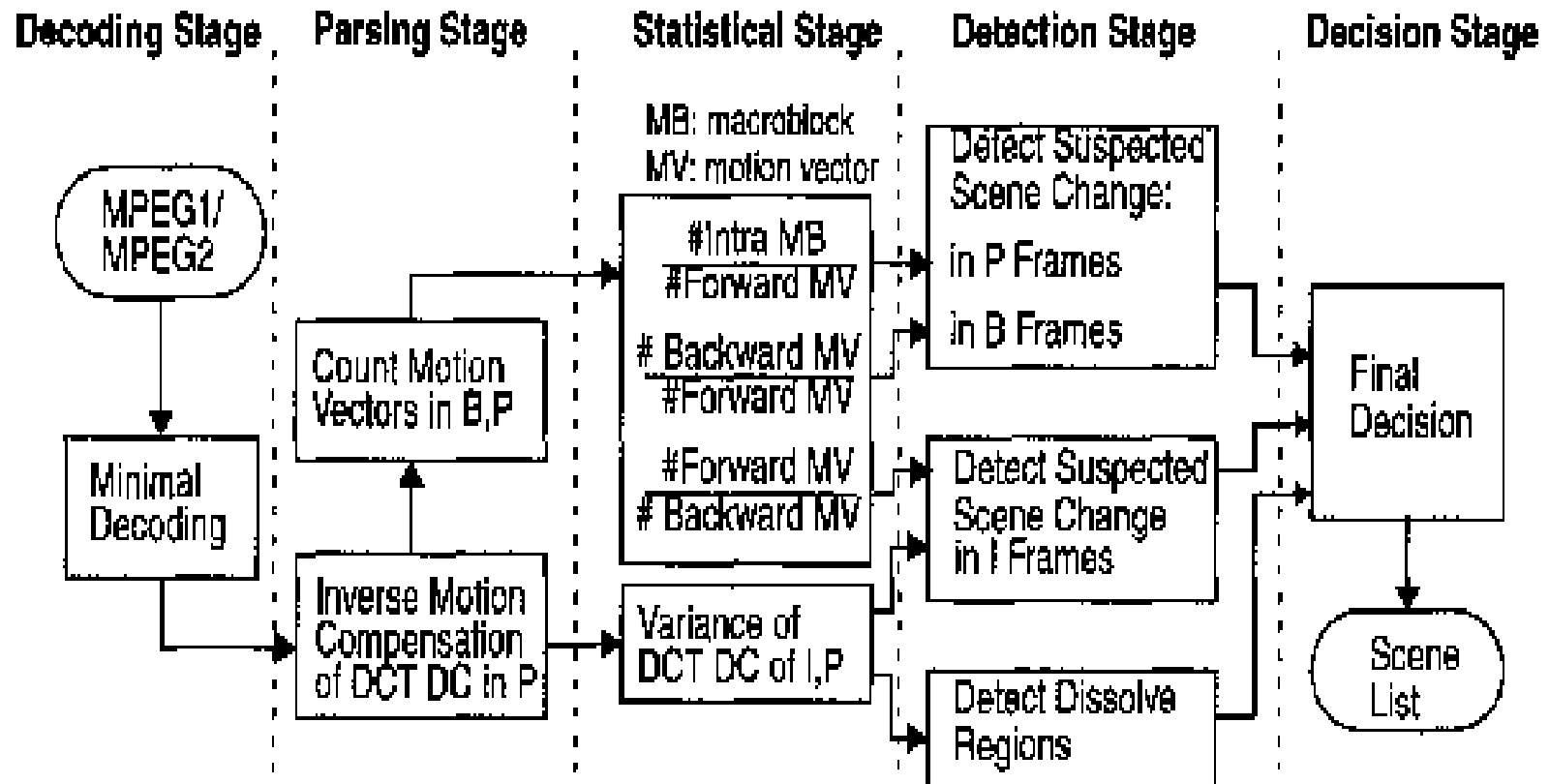
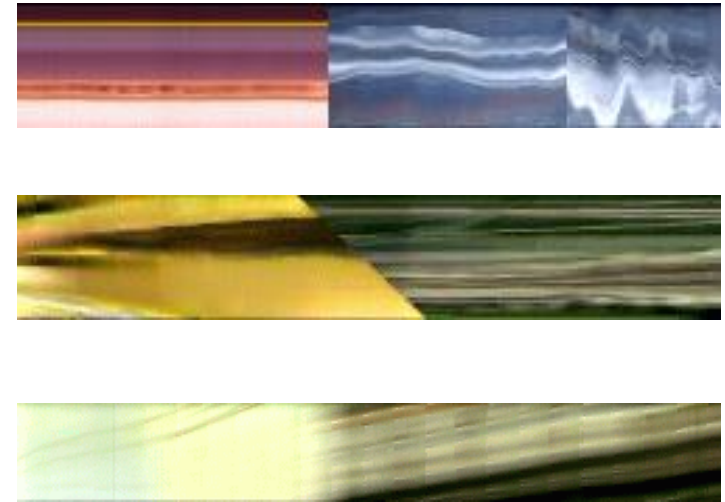
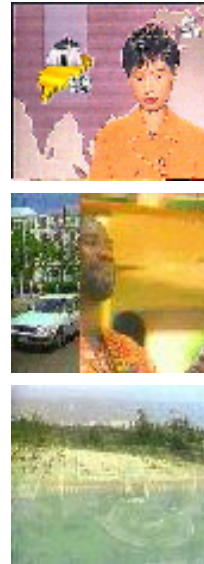
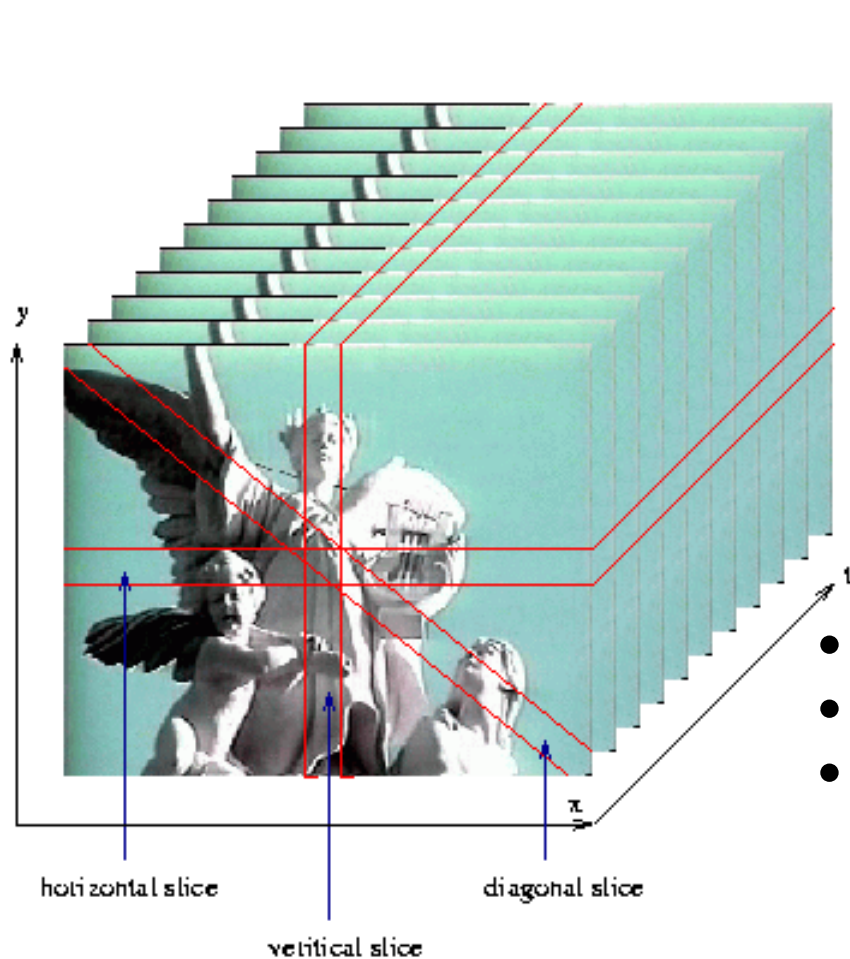


FIGURE 2. The Compressed Domain Scene Cut Detection Algorithm

Méthode de détection de transition de prise

4. Avec le modèle de découpage spatio-temporel



- Identification d'images-clés
- Détection mouvement d'objets ou caméra
- Distinction objet vs caméra par analyse des macroblocs: caméra si mouvement consistant sur majorité de macroblocs



Évaluation des méthodes

de : Paul Browne, Centre for Digital Video Processing, Dublin City University

Mesures d'évaluation

- Rappel :

$$\frac{\text{Nombre de bonnes prises trouvées}}{\text{Nombre total de prises}}$$

- Précision :

$$\frac{\text{Nombre de bonnes prises trouvées}}{\text{Nombre de prises trouvées (bonnes + mauvaises)}}$$

→ Équilibrer Rappel-Précision

Évaluation des méthodes

de : Paul Browne, Centre for Digital Video Processing, Dublin City University

Précision moyenne sur 8 heures de vidéo

- Histogramme 90.4
- Forme 90.0
- Macrobloc 87.4

Rappel moyen sur 8 heures de vidéo

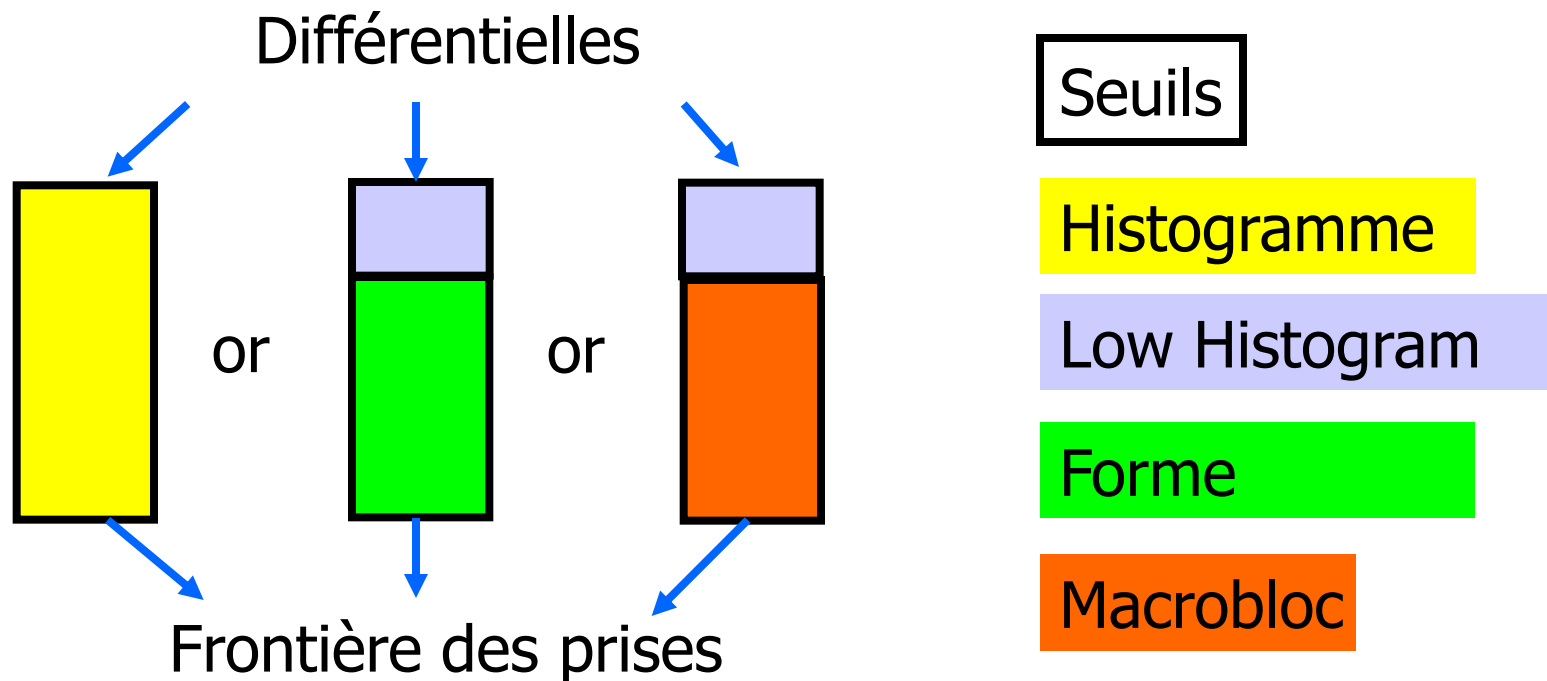
- Histogramme 78.9
- Forme 70.2
- Macrobloc 75.3

Émissions avec les plus bas niveaux de rappel

- Home & Away (Australian soap)
- Cooking Program

Combiner les approches

- Détection des frontières par différentes méthodes
- Combinaison des décisions par un 'OU' logique

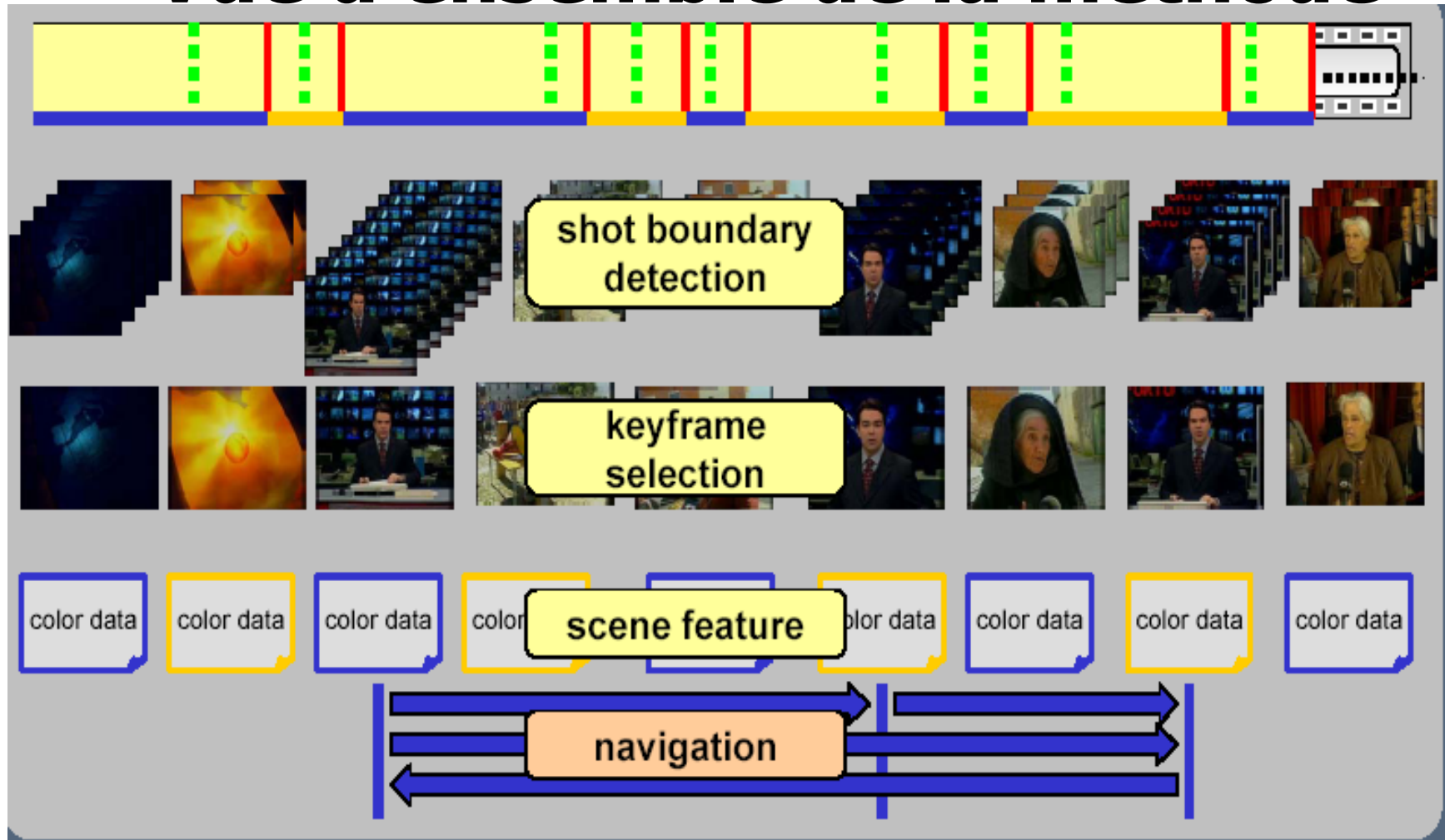


Nouvelles approches

- Segmentation temporelle de vidéo avec détection de frontière de prises en MPEG-7

[source: *MPEG-7 Descriptors for Temporal Video Segmentation*, Michael Höynck, Institute for Communications Engineering Aachen, University of Technology]

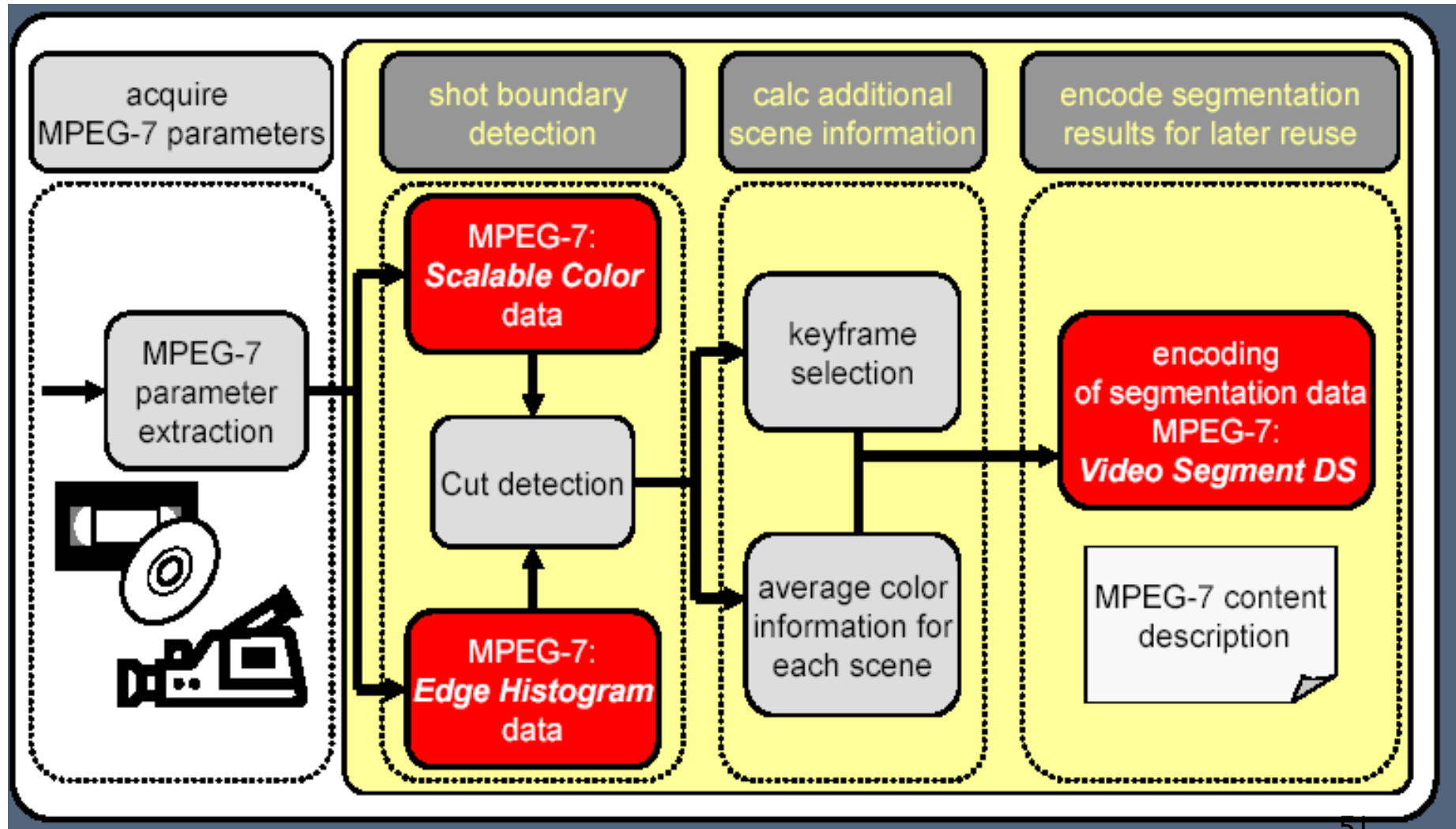
Vue d'ensemble de la méthode



Base de la méthode

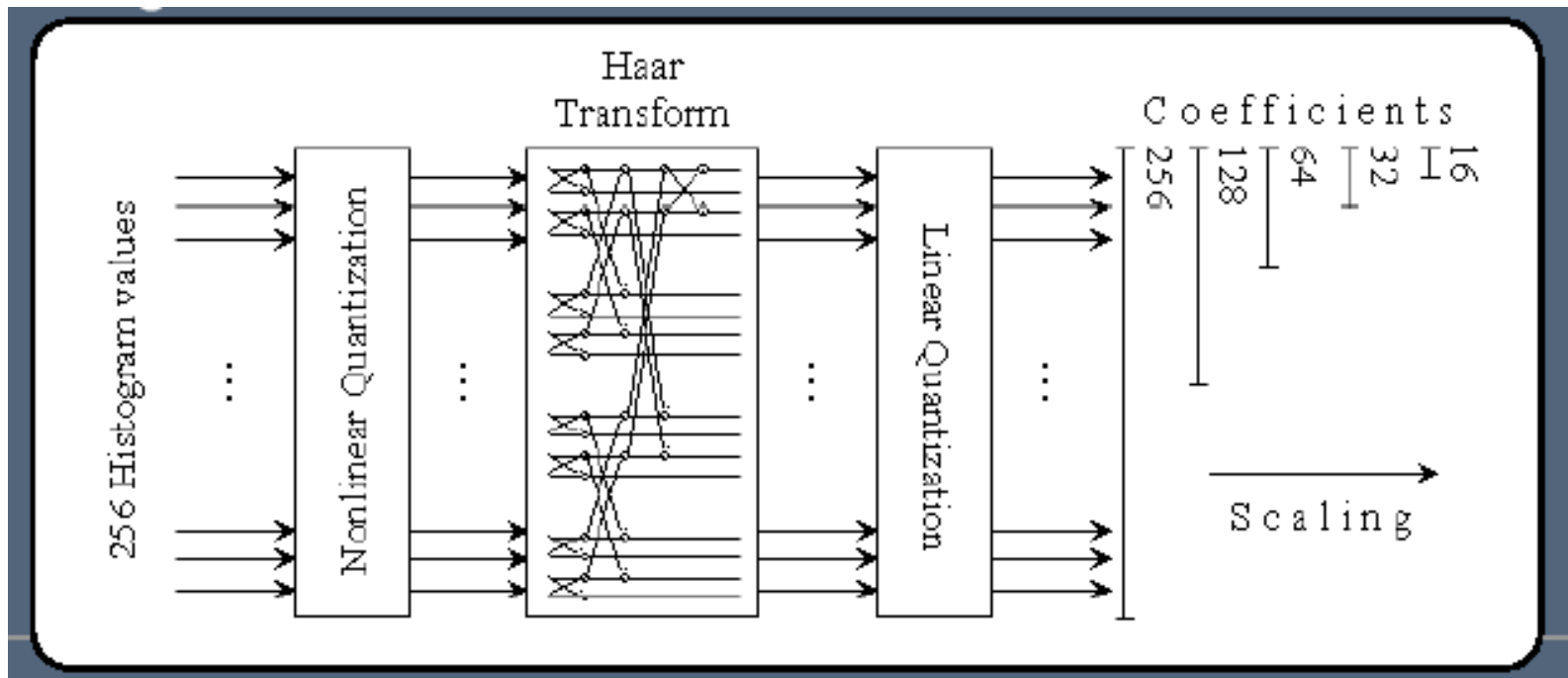
- MPEG-7 normalise la description du contenu
 - Scalable Color Descriptor (SCD) : histogramme de couleurs basé sur l'espace HSV, encodée par la transformée de Haar
 - Edge Histogram Descriptor (EHD) : histogramme de formes
- MPEG-7 est flexible (interopérable)
 - description des traits utilisés pour segmentation
 - algorithmes d'extraction des traits
 - méthodes de repérage

Vue d'ensemble



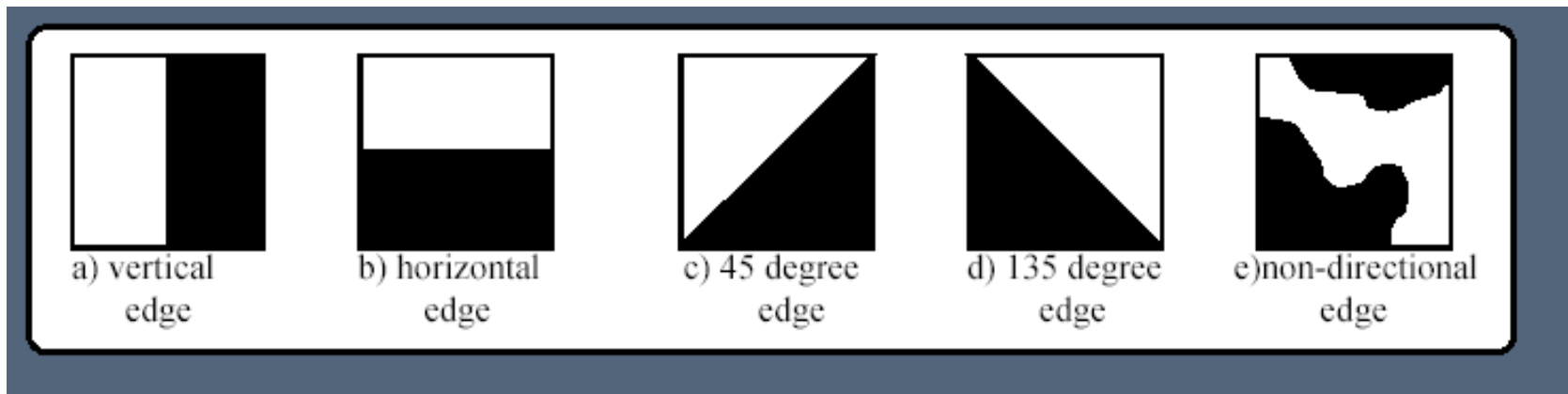
SCD – Scalable Color Descriptor

- Transformée de Haar avec 256-bin d'histogramme dans l'espace de couleur HSV
- Représentations SCD enregistrées sous différentes résolutions [16 à 256 coefficients par histogramme]



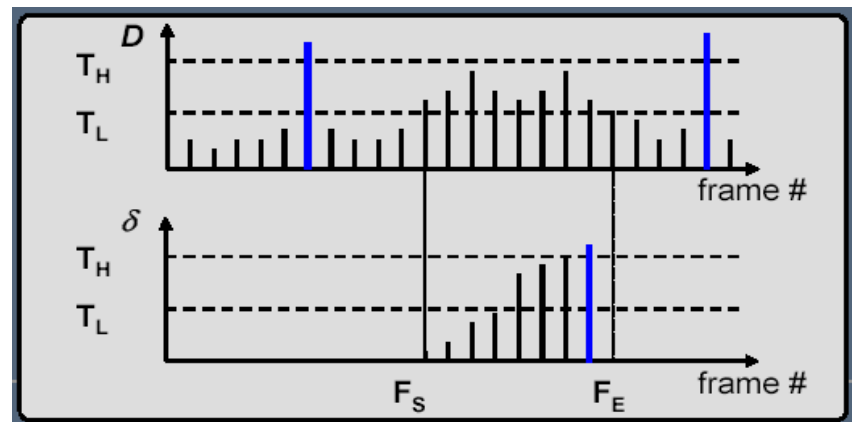
EHD – Edge Histogram Descriptor

- Spécifie la distribution spatiale de 5 types de formes dans 16 régions d'images
- Dérive aussi la forme générale



Méthode de détection de prise

- Segmentation par les histogrammes de couleurs et de formes
- Calcul des différentielles par une mesure de distance (e.g., L1-norm)
- Pour la couleur : twin comparison method
 - Si $\text{diff} < T_L$: omettre
 - Si $T_L < \text{diff} < T_H$: cumule les différentielles
 - Si $T_H < \text{Cumul}(\text{diff}) \Rightarrow$ changement graduel détecté
 - Si $T_H < \text{diff} \Rightarrow$ frontière de prises



Exemple de détection



abrupt changes

(direct cut)

gradual transitions

(e.g. wipe and dissolve or fade)



#1259



#1260 **)



#1365



#1368



#1372



#1375 **)



Évaluation de la performance de la détection de prises

Résultats de performance globale

- 97% rappel et 80% précision
- Définition de l'ensemble de test (naturel, synthétique, genres)
 - détermination sur échantillon (1170 scènes)
 - performance évaluée par les mesures de rappel et de précision

Segmentation vidéo et reconnaissance de forme

- Structure sémantique d'une vidéo difficile à analyser
- Nécessite la connaissance spatio-temporelle des interactions entre objets de même séquence
- Reconnaissance des figures et des personnes
- Reconnaissance d'objets à travers plusieurs images
- Segmentation spatio-temporelle combinée à l'estimation des mouvements de caméra
 - Permet d'identifier des objets spécifiques et d'en faire le décompte (ex. nombre de personnes ou de voitures dans une application de surveillance)



Utilisation d'heuristiques (*lignes directrices*) pour déterminer les frontières

- Les prises d'une séquence avec un contenu similaire en couleurs partagent souvent le même arrière-plan
 - Couleurs changent drastiquement à la fin d'une séquence
 - Changement d'angle de la caméra n'a pas d'effet sur les couleurs de l'arrière-plan
- Contenu audio varie considérablement d'une séquence à l'autre; coïncidence d'une coupure audio avec une coupure vidéo suggère un changement de séquence
- Prises consécutives regroupées en une séquence si identification d'un même dialogue

Plan de la présentation

- Pourquoi extraire les traits d'un vidéo (12.1)
- Analyse de vidéo et segmentation (12.2)
- **Métadonnées d'audio et de vidéo (12.6, 12.7)**
- Classification vidéo
- Indexation de vidéo (12.7.1)
- Requêtes vidéo (12.9)
- *interMedia* pour vidéo et audio

MPEG-1, MPEG-2

- Focus sur compression des images
- MPEG-1
 - standard de base pour CD vidéo et MP3
- MPEG-2
 - standard de base pour télévision numérique et DVD
 - supporte vidéo de plus haute qualité (nouveaux codec)
 - supporte l'encodage prédictif (par différentielles) pour audio/vidéo imbriqués
 - modes extensibles pour dégradation graduelle (temps réel)

MPEG-4

- Même support que MPEG-2 + support pour
 - traitement distribué
 - reconnaissance d'objet (segmentation avant-plan / arrière-plan)
 - canaux mobiles
 - transmissions interactives (vidéoconférences)
- En cours d'élaboration
 - Représentation des contenus
 - Interfaces pour la gestion des droits numérisés

MPEG-4

- Scènes audiovisuelles organisées en hiérarchie d'objets
- Objets de base
 - Images fixes (arrière-plan)
 - Objets vidéo (personne qui parle, sans l'arrière-plan)
 - Objets audio (discours de la personne qui parle)

MPEG -7

- Support additionnel pour l'indexation et le repérage de contenu audio et visuel (CBR)
- Utilisé avec MPEG-4 (complément)
- Couvre
 - Format d'encodage des descripteurs de contenu audio et vidéo (Systems)
 - Langage de définition des descriptions (DDL)

MPEG -7

- **Couvre (suite)**

- Outils pour description de contenus visuels (Visual)
- Outils pour description de contenus audio (Audio)
- Outils pour description de traits génériques (Multimedia Description Schemes)
- Implémentation des logiciels de support pour les différentes parties MPEG-7 (Reference Software)
- Procédure de validation de la conformité à l'implémentation de MPEG-7 (Conformance)
- Information sur l'extraction et l'utilisation de certaines descriptions de contenus (Extraction and use of descriptions)



Plan de la présentation

- Pourquoi extraire les traits d'un vidéo (12.1)
- Analyse de vidéo et segmentation (12.2)
- Métadonnées d'audio et de vidéo (12.6, 12.7)
- **Classification vidéo**
- Indexation de vidéo (12.7.1)
- Requêtes vidéo (12.9)
- *interMedia* pour vidéo et audio

Capture vidéo (video record)

- Segmentation utilisée pour identifier des séquences vidéo spécifiques
 - Séquence des images d'un objet spécifique
 - Débute par l'image où l'objet apparaît et se termine par l'image où il disparaît
- Format
 - identifiant unique
 - annotation textuelle (manuelle)
 - des vecteurs de mouvement de l'objet (motion vector)
 - ensemble de primitives de mouvement (e.g. rotation vers la droite, nord)
 - Tuples spatio-temporels (x, y, t)

Traitement d'images en mouvement

- Caractéristiques temporelles d'une vidéo
- Traiter les relations de
 - temps
 - durée
 - synchronisation
- Doit habituellement décompresser l'objet vidéo

Synchronisation

- **Synchronisation des lèvres** – synchronisation des mouvements de lèvre de la bande vidéo avec la bande sonore pour rendre le réalisme
- **Synchronisation des estampilles** – consistance des estampilles temporelles sur chaque bande d'une vidéo
- **Synchronisation du transport** – ordonnancement temporel des paquets transmis par voie de télécommunication (temps réel)

Synchronisation en vidéoconférence

- Aspect temps réel et bidirectionnel
- **Synchronisation intra-flux** : encapsulation des paquets et des estampilles (protocole RTP – Real-Time Transport Protocol)
- **Synchronisation inter-flux** : synchronisation des flux des différents sites; utilisation d'une horloge globalement synchronisée (protocole NTP – Network Time Protocol)
- **Synchronisation inter-media** : synchronisation des lèvres (protocoles NTP ou H221)

Plan de la présentation

- Pourquoi extraire les traits d'un vidéo (12.1)
- Analyse de vidéo et segmentation (12.2)
- Métadonnées d'audio et de vidéo (12.6, 12.7)
- Classification vidéo
- **Indexation de vidéo (12.7.1)**
- Requêtes vidéo (12.9)
- *interMedia* pour vidéo et audio

Indexation automatique de vidéo segmentées

- Approche par images
 - Déterminer l'image-clée de chaque séquence (scène ou prise)
 - Utiliser les techniques d'indexage d'images pour indexer les images-clées
- Requêtes CBR à partir de cet index

Indexation automatique de vidéo segmentées

- Approche par traits
 - Extraction automatique de courtes séquences du vidéo qui représentent les traits importants des longues séquences
 - Plus informatif que les images fixes
- Utile pour requêtes de type QBE
- Utilisé pour le 'video skimming'

Plan de la présentation

- Pourquoi extraire les traits d'un vidéo (12.1)
- Analyse de vidéo et segmentation (12.2)
- Métadonnées d'audio et de vidéo (12.6, 12.7)
- Classification vidéo
- Indexation de vidéo (12.7.1)
- **Requêtes vidéo (12.9)**
- *interMedia* pour vidéo et audio

Requêtes vidéo

- **Niveau 1: traits primitifs** (couleurs, formes, textures, localisation spatiale et mouvement)
- **Niveau 2: traits logiques** (objets ou relations nommés, e.g. avion au décollage)
- **Niveau 3: attributs abstraits** (associé à la compréhension, e.g. trouver une vidéo sur les désastres naturels)
- L'organisation des traits du vidéo et de ses métadonnées est importante pour les requêtes
- Extraction d'un ensemble de vidéos ou d'extraits qui répondent à la requête; l'utilisateur **doit regarder les clips un à un**

Abstractions vidéo (p.415)

- Méthodes d'accélération de consultation de vidéo (un à un):
 - Consultation syntaxique: On présente des images miniatures « thumbnails » des prises
 - Consultation sémantique: On présente des graphiques de transition de prises

Plan de la présentation

- Pourquoi extraire les traits d'un vidéo (12.1)
- Analyse de vidéo et segmentation (12.2)
- Métadonnées d'audio et de vidéo (12.6, 12.7)
- Classification vidéo
- Indexation de vidéo (12.7.1)
- Requêtes vidéo (12.9)
- ***interMedia* pour vidéo et audio**

Application vidéo avec Oracle

- Fait avec Oracle *interMedia* ORDvideo et ORDaudio

*inter*Media ORDvideo

- *inter*Media vous offre déjà des méthodes communes à ORDaudio, ORDimage et ORDvideo.
- Travaille au niveau du Blob et du BFile
- Peut extraire les traits du vidéo dans un fichier XML
- Types d'objets ORDvideo
- Constructeurs ORDvideo
- Méthodes additionnelles de ORDvideo

Type objet **ORDvideo**

- Supporte enregistrement et gestion d'objets vidéo

```
CREATE OR REPLACE TYPE ORDVideo AS OBJECT
```

```
( -- ATTRIBUTES
```

```
description VARCHAR2(4000),
```

```
source ORDSource,
```

```
format VARCHAR2(31),
```

```
mimeType VARCHAR2(4000),
```

```
comments CLOB,
```

```
-- VIDEO RELATED ATTRIBUTES
```

```
width INTEGER, height INTEGER, frameResolution INTEGER, frameRate
```

```
INTEGER, videoDuration INTEGER, numberOfFrames INTEGER,
```

```
compressionType VARCHAR2(4000), numberOfColors INTEGER, bitRate
```

```
INTEGER,
```



Méthodes ORDvideo

- Initialiser une instance vidéo

`init(srcType, srcLocation, srcName)`

```
BEGIN INSERT INTO pm.online_media (product_id,  
product_video) VALUES (2030,  
ORDSYS.ORDVideo.init('FILE', 'FILE_DIR','speakers.rm'));  
COMMIT; END; /
```

- `processVideoCommand()` : pour passer une commande et des paramètres à un plug-in

ORDvideo méthodes (en plus)

checkProperties	getFrameResolution	setDescription
getAllAttributes	getFrameSize	setFormat
getAttribute	getNumberOfColors	setFrameRate
getBitRate	getNumberOfFrames	setFrameResolution
getCompressionType	getVideoDuration	setFrameSize
getContentInLob	import	setKnownAttributes
getContentLength	importFrom	setNumberOfColors
getDescription	processVideoCommand	setNumberOfFrames
getFormat	setBitRate	setProperties
getFrameRate	setCompressionType	setVideoDuration

Compression ORDvideo

- Apple Quicktime 3.0
- Microsoft AVI
- Realnetworks Real Video format
- MPEG-1
- MPEG-2
- MPEG-4