

Les fondamentaux de l'IHM et du multimédia

L'image animée
La vidéo numérique
Les nouvelles normes MPEG



UE NSY116
Multimédia et Interaction
Homme-Machine
2011-2012

Alexandre Topol

Plan du cours

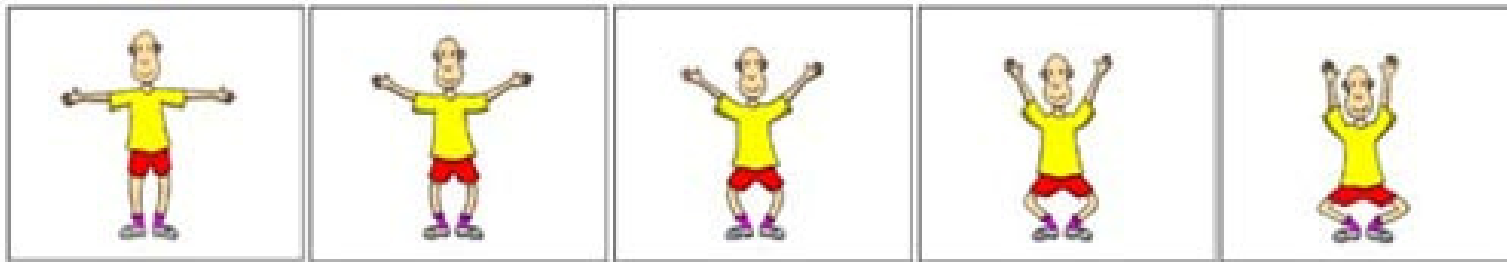
- Film analogique vs vidéo numérique
- Standards et formats
- Codecs et compression
- Capture
- Montage
- MPEG-4 – Architecture pour la diffusion multimédia
- MPEG-7 – Indexation de contenu
- MPEG-21 – Infrastructure multimédia

- Qu'est-ce que la vidéo ?
 - Une vidéo, comme un film, est une animation
 - Une animation est une série d'images se succédant à une fréquence suffisante pour produire par persistance rétinienne un effet d'images animées
 - Dans le cas de la vidéo, l'animation est stockée par des moyens électroniques ou informatiques
 - Dans le cas d'un film, le support est la pellicule analogique



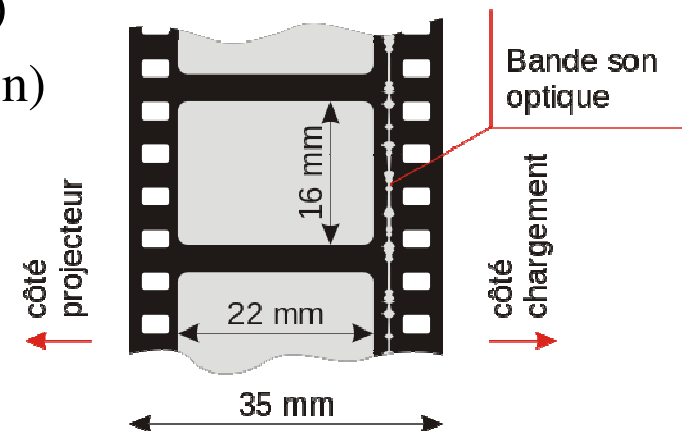
le cnam Film vs Vidéo

- Exemple du GIF animé



le cnam Analogique et Numérique

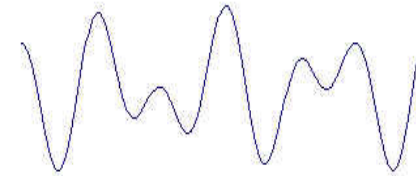
- Une vidéo ou un film contient de manière codée la suite des images ainsi que le son correspondant à ces images
- Ces deux types d'information sont traitées en même temps mais elles sont bien distinctes dans le codage
- En codage analogique, chacune des informations (image, son) se présente sous forme d'un :
 - signal électrique (conducteur électrique),
 - magnétique (bande magnétique, cassette)
 - ou électromagnétique (ondes de télévision)



le cnam Analogique et Numérique

- En numérique, chacune des informations est d'abord codée sous forme numérique (suite de 0 et de 1). Ces 0 et ces 1 peuvent ensuite être transmis ou mémorisés par tous les procédés utilisés en analogique
- On peut transmettre par des moyens analogiques (électrique, magnétique, électro-magnétique) des codes numériques
- CD et DVD = stockage numérique

le cnam Analogique Vs Numérique



- L'analogique
 - Définition :
 - Grandeur physique variant de façon continue (*i.e.* dérivable)
 - Avantage :
 - Grande facilité à traiter les données en temps réel par des circuits électroniques
 - Inconvénients :
 - Moins de possibilités de montage
 - Dégradation de l'image à chaque recopie
 - Qualité limitée
 - Sensibilité au bruit et parasites

le cnam Analogique Vs Numérique

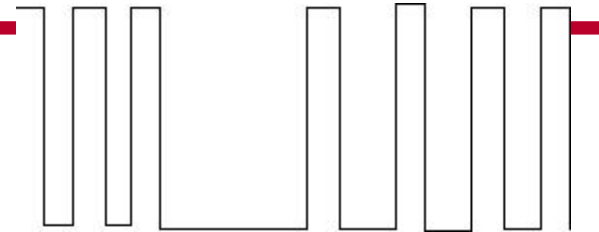
- Le numérique

- Avantages :

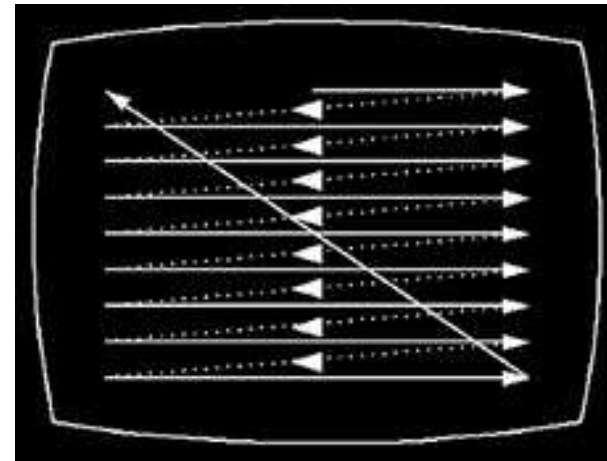
- Nombreuses possibilités de traitements, facilité de montage (montage non linéaire)
 - Recopie des données sans pertes
 - Qualité contrôlée et paramétrable
 - Insensibilité au bruit et parasites

- Inconvénients :

- Traiter les données en temps réel nécessite du matériel pointu et coûteux (les bancs de montage en analogique sont encore plus coûteux)
 - Il faut maîtriser l'outil informatique. Encore des progrès à faire dans la qualité des logiciels



- **Standard** analogique de diffusion couleur : manière de coder l'ajout de couleur aux images N&B
 - NTCS
 - Apparu en 1953
 - Résolution de 640x480
 - Images/sec : 29,97 fps
 - Lignes : 525
 - Pb : générateur de diaphotie
 - PAL/SECAM
 - 1963/1959
 - 768x576
 - 25fps
 - 625
- Différence entre nb lignes et lignes utiles (49 lignes en 50Hz, 45 lignes en 60Hz) : un subterfuge pour laisser le temps au faisceau d'électrons de remonter sur les écrans CRT

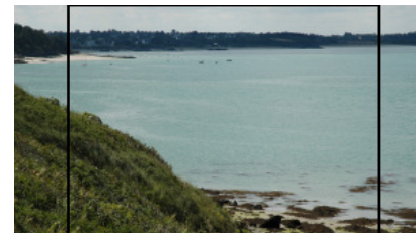
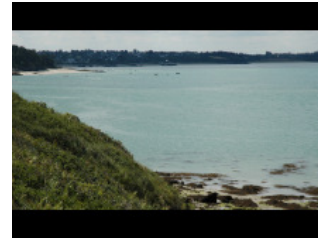


Standards

le cnam

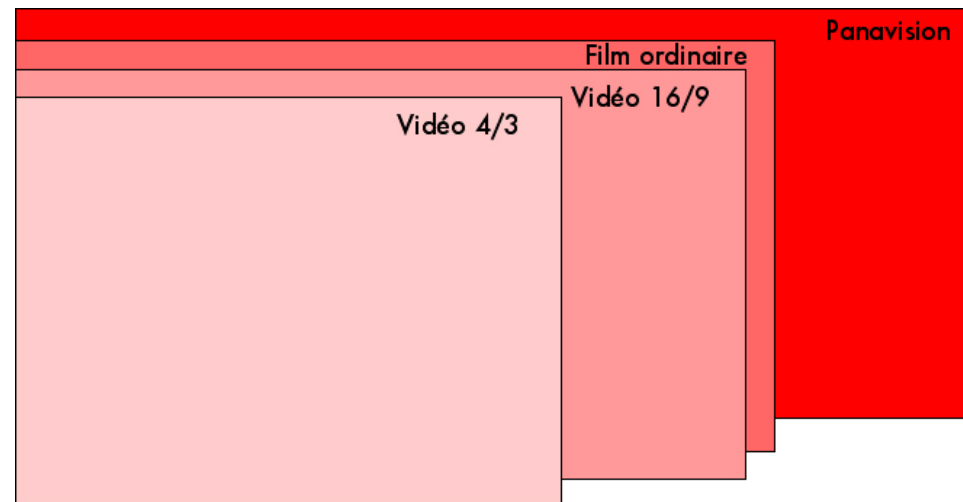
- Avec ces standards analogiques, le débit de la vidéo est constant
- Leur mode de balayage est dit entrelacé : trames impaires d'abord puis trames paires ensuite
 - En NTSC : le tube cathodique est balayé 60 fois par secondes donc deux trames de 240 lignes à 60Hz = 480i@60Hz en NTSC
 - En PAL/SECAM : le tube cathodique est balayé 50 fois par secondes donc deux trames de 288 lignes à 50Hz = 576i@50Hz en PAL

- 16/9 de plus en plus utilisé pour la télévision, proche du format cinéma, plus agréable car l'angle de vision du regard humain est plus grand en largeur qu'en hauteur
- Affichage d'une image 16/9 sur un écran 4/3 (*Letterbox*)
- Affichage d'une image 4/3 sur un écran 16/9 (*Rideaux*)
- Recadrage



Modes d'affichage

- Télévision traditionnelle : $4/3$ soit un rapport de 1,33
- Vidéo large : $16/9$ soit un rapport de 1,78
- Film ordinaire : rapport 1,85
- Film Panavision : rapport 2,35



- Luminance : Plus une couleur est claire plus sa luminance est élevée
- Le contraste : C'est l'écart de luminance entre deux parties d'une image. Le contraste est fort entre un noir et un blanc



- La saturation : une couleur est saturée lorsqu'elle est pure. Ainsi un rose est un rouge non saturé (il ne contient pas que du rouge mais aussi du blanc)

le cnam Codage numérique des images

- RVB
- YUV :
 - Y est la luminance ($Y = 0,3 R + 0,59 V + 0,11 B$). Les coefficients (pourcentages) ont été déterminés en fonction de la sensibilité de l'œil humain pour les trois couleurs primaires
 - U la différence de couleur exprimée à partir du bleu ($\text{DiffBleu} = B - Y = 0,89 B - 0,3 R - 0,59 V$)
 - V la différence de couleur exprimée à partir du rouge ($\text{DiffRouge} = R - Y = 0,7 R - 0,59 V - 0,11 B$)
- Le passage du RVB au YUV est sans perte
- Le système visuel humain est moins sensible aux nuances de couleurs qu'aux différences d'intensités lumineuses. D'où l'intérêt de ce codage au moment du compactage

le cnam **Les formats**

- Encore un terme employé pour beaucoup de choses
- Format = format de stockage
- Analogiques :
 - Grand public : VHS, 8mm, S-VHS, Hi8
 - Pro : Betacam, Betacam SP
- Numériques :
 - Historique et pro : D1 à partir de 1989
 - DV :
 - deux pistes audio à 48 khz d'échantillonnage
 - Qualité équivalente au Betacam, au DVPro ou DVCam
 - Seule différence au niveau du matériel
 - DV Pro et DVCAM :
 - Vitesse de défilement + élevé que pour le DV

- Format = « résolution » de la vidéo ou de l'écran
- Résolutions des ordinateurs :
 - VGA 640 x 480 (4:3)
 - SVGA 800 x 600 (4:3)
 - XGA 1024 x 768 (4:3)
 - WXGA 1280 x 768 (15:9)
 - SXGA 1280 x 1024 (5:4)
 - SXGA+ 1400 x 1050 (4:3)
 - WSXGA 1680 x 1050 (16:10)
 - UXGA 1600 x 1200 (4:3)
 - WUXGA 1900 x 1200 (1.58:1)
 - QXGA 2048 x 1536 (4:3)

- Résolutions des TV analogiques :
 - PAL 720 x 576
 - PAL VHS 320 x 576 (approx.)
 - NTSC 640 x 482
 - NTSC VHS 320 x 482 (approx.)
- Résolutions TV numériques :
 - NTSC (preferred format) 648 x 486
 - PAL 720 x 486
 - HDTV 1920 x 1080
- Résolutions cinéma :
 - Academy standard 2048 x 1536

Codecs

le cnam

- Les codecs (COmpactage-DECompactage)
- Ce sont les algorithmes permettant de coder la succession des bits pour qu'ils aient une signification
- Ce n'est pas le moyen de transmettre ou de stocker de la vidéo !!
- Pour lire une vidéo il faudra donc :
 - Un matériel compatible :
 - Avec le support physique utilisé
 - Avec le codec utilisé pour encoder la vidéo
 - Avec la norme utilisée pour le format de la vidéo
 - Ex : un DVD qui stocke une vidéo en PAL encodée en DIVX

- Exemple simple :
 - le format DV propose des images dont la taille est de 720×576
 - Si l'on raisonne sur une seule image de cette taille codée en millions de couleurs, il faut 3 octets par pixel, donc 1 244 160 octets par image
 - A 25 images par secondes on arrive au nombre impressionnant de 31 104 000 octets/seconde
 - soit en valeurs arrondies 31 Mo par seconde ou encore 1,8 Go par minute
- En réalité le débit du DV est environ de 3,2 Mo par secondes en qualité maximale (compression minimale) soit environ 10 fois moins que ce calcul théorique

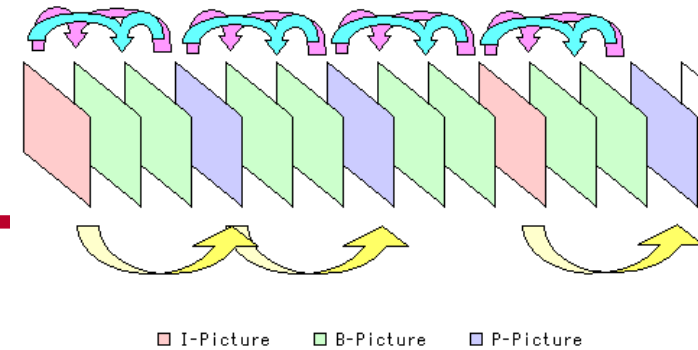
Codecs

le cnam

- Deux catégories de compression :
 - Spatiale : algo identique à ceux utilisés avec les images fixes
 - Temporelle : on ne stocke que ce qui a changé d'une image à l'autre → cohérence temporelle
- En pratique les deux techniques sont employées conjointement
- Fonctionnement de base :
 - On conserve des images clés auxquelles est appliquée une compression spatiale
 - Les images situées entre deux images clés sont allégées en ne stockant que les différences par rapport à ces images clés

- Quelques codecs :
 - M-JPEG (Motion JPEG) utilise la cohérence spatiale pour la compression. L'accès à une image précise est facile. Il est utilisé dans les studios de montage numérique (débit 1Mo/s environ)
 - MPEG (Moving Pictures Experts Group) utilise la cohérence temporelle pour la compression.
 - MPEG1 débit de 180 Ko/s qualité proche du VHS (352x240 à 30 images par seconde en NTSC et 352x288 à 25 images par seconde en PAL/SECAM)
 - MPEG2 débit de 5 Mo/s qualité prévue pour la télévision à haute définition. Jusqu'à 5 canaux audios. C'est le format typique des DVD.
 - MPEG4 version vectorisée permettant de l'interactivité avec les objets composants le flux vidéo. Application au multimédia et au WEB
 - mais aussi MPEG7, MPEG21,...

- Quelques codecs :
 - Le DivX se caractérise par un très fort taux de compression permettant de stocker plus d'une heure de vidéo sur un cédérom de 650Mo (ce qui lui vaut son succès)
 - Qualité bien inférieur au MPEG-2 mais moins gros en taille
 - XviD, VP3, 3ivX sont différentes implémentations du codec
 - MKV n'est pas un codec mais un format de fichier permettant d'encapsuler des types d'informations vidéos produites par différents codecs
 - Comme AVI
 - Comme Quicktime



- Codage MPEG
 - algorithme de compactage spatial de type jpeg
 - compactage temporel qui cherche à ne coder que ce qui change d'une image à l'autre
 - Pour cela on définit différents type d'images :
 - les images **I** qui ne présentent qu'un compactage interne (compactage spatial en jpeg). Ce sont les images qui occupent le plus de place en mémoire.
 - Les images **P** calculées à partir de l'images I ou P la plus proche en amont. P ne mémorise que les différences avec cette image de référence (compactage temporel)
 - Les images **B** sont obtenues par interpolation de deux images calculées. Ces deux images codent les différences avec une image I ou P situées de part et d'autre de B. Ce sont les images qui occupent le moins de place en mémoire (compactage temporel)

le cnam **Capture**

- Le préalable au traitement de la vidéo
- 2 cas :
 - Les données de départ sont analogiques (bande provenant d'une prise de vue Betacam et HI8 par exemple) :
 - ➔ il faut les numériser (carte d'acquisition vidéo ou carte de capture vidéo) pour produire les fichiers sur le disque de travail
 - Les données sont numériques dès le départ (bande provenant d'une caméra DV par exemple) :
 - ➔ il faut les recopier sur le disque de travail

- Théorème de Shannon

la fréquence d'échantillonnage doit être au moins égale au double de la fréquence du signal analogique. Si l'on se situe sous cette limite théorique, il y a perte d'information dans le signal

- Exemples dans l'audio

- pour $F < 20$ kHz (son Hi-Fi), $F_e = 44,1$ kHz
- pour $F < 3400$ Hz (voix humaine en téléphonie), $F = 8$ kHz

- Exemples pour la vidéo

- le format DV (720 points/ligne) convient bien pour la capture VHS
Pal (352 points/ligne) sur une caméra avec une entrée analogique

- Matériels :
 - Disque dur :
 - AV
 - Vitesse de rotation (> 7200 tours/minute)
 - Vitesse de transfert
 - La norme IEEE 1394 (FireWire) permet sans problème la copie des infos sur le disque (400 Mbits/s)
 - Idem USB2 (480 Mbits/s)

- Un fichier vidéo reprend certaines des caractéristiques des ressources photo et audio :
 - Les dimensions : largeur et hauteur toujours exprimées en pixels.
 - La profondeur de couleur : elle indique la quantité de mémoire utilisée pour coder la couleur de chaque pixel : usuellement 24 bits
 - Le nombre d'images par seconde : NTSC 30 fps, PAL et SECAM 25 fps, le cinéma traditionnel 24 fps
 - Taux d'échantillonnage : ce paramètre détermine le nombre d'échantillons sonores réalisés par seconde
 - Amplitude : elle indique la quantité de mémoire utilisée pour coder un échantillon : usuellement de 12 ou 16 bits
 - Mono/stéréo/5.1 : Précise le nombre de canaux sonores présents

- Terminologie
 - Acquisition (capture) : processus consistant à transférer la vidéo d'un caméscope vers un support exploitable par un ordinateur. C'est l'étape préalable au **montage**
 - Montage : opération consistant à enchaîner dans un ordre déterminé des **plans** ou des images fixes
 - Plan (rush) : suite d'images consécutives extraites d'une vidéo, accompagnée éventuellement de son
 - Séquence : Suite de plans assemblés afin de véhiculer un sens

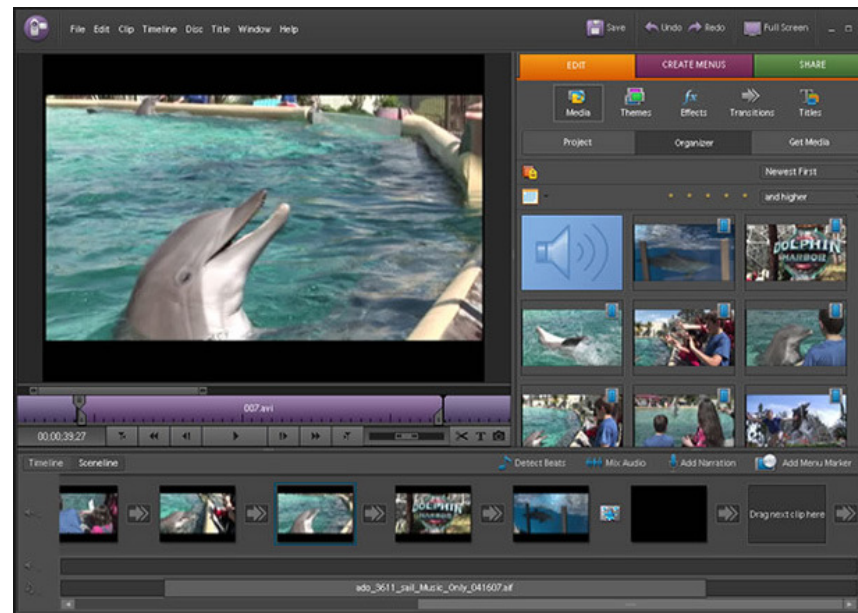
le cnam **Montage**

- Le but est de sélectionner des images enregistrées pour les assembler en une suite cohérente
- C'est l'une des opérations finales dans la réalisation d'un film ou d'un documentaire
- Les bancs de montage « linéaires » équipés au minimum d'un magnétoscope lecteur et d'un autre enregistreur
 - En option un mélangeur vidéo,
une console de mixage,
une console multi-effets
 - Laissent la place au montage logiciel



le cnam Montage

- Les bancs de montage « non-linéaires » ou « virtuels » sous la forme d'un logiciel
 - Seul besoin matériel, la prise pour acquérir la vidéo (DV ou HDV pour les particuliers, BETACAM ou HDCAM pour les pros)

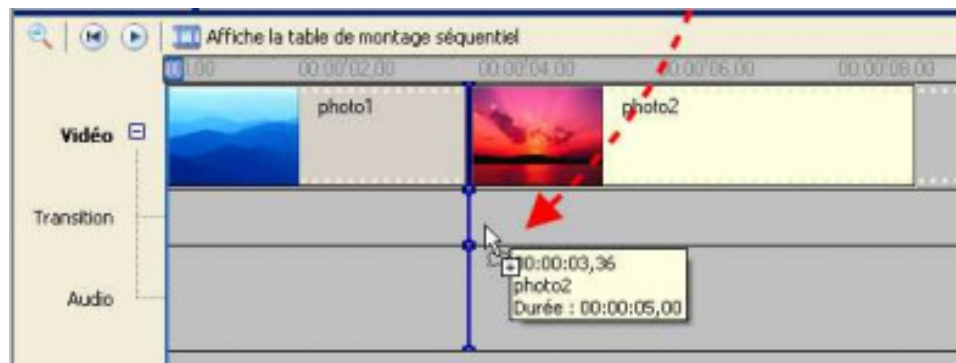


le cnam **Opérations de montage**

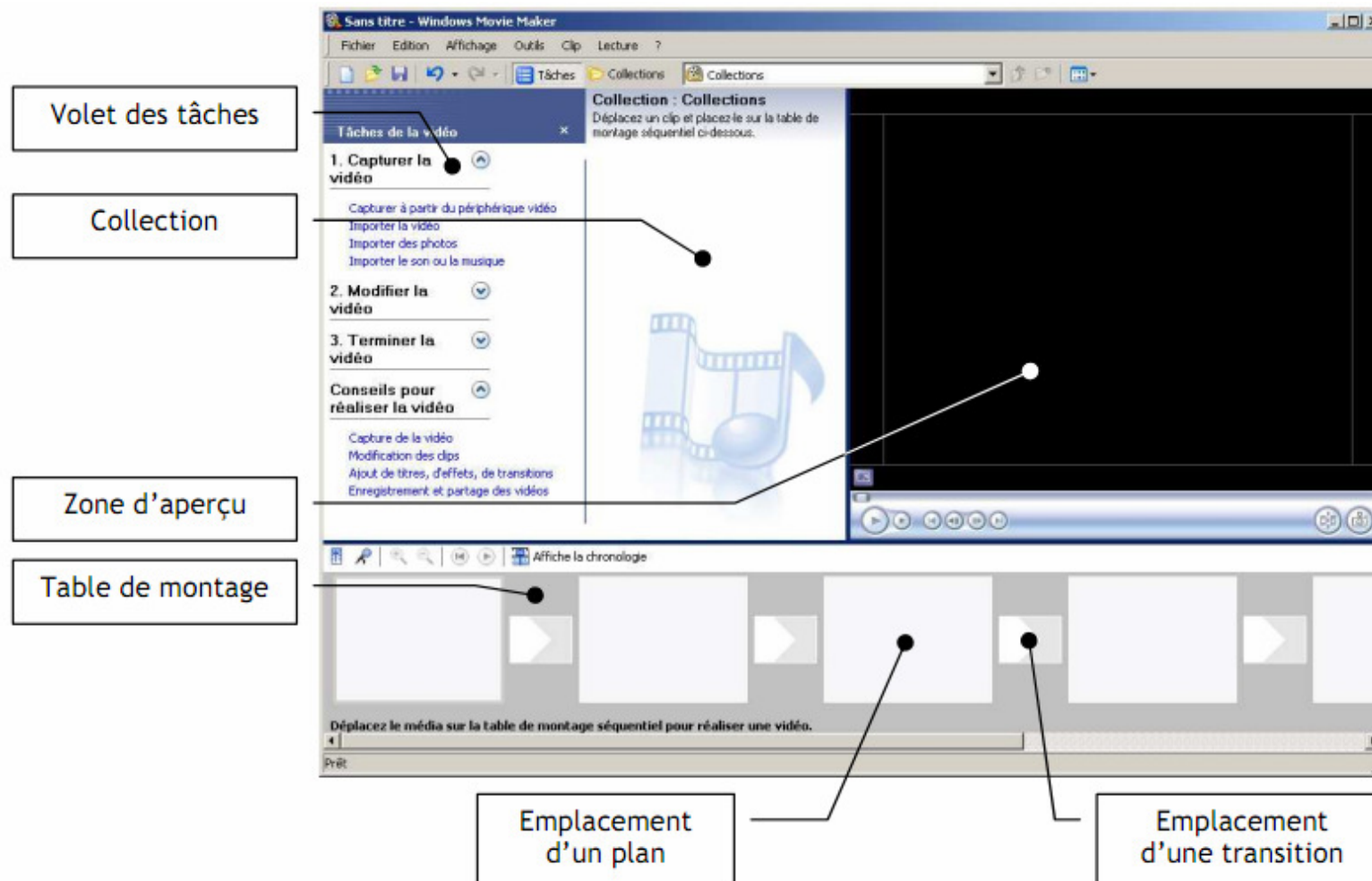
- sélection des séquences
- ordonnancement
- Insertion
- sonorisation
- ajout de transitions
- traitement des séquences vidéos (amélioration des images)
- titrage (en plein écran ou en incrustation)
- effets spéciaux par traitement des images (filtres, animations, etc.)
- production pour la diffusion en différents formats à partir du même original

le cnam Séquence d'opérations classique

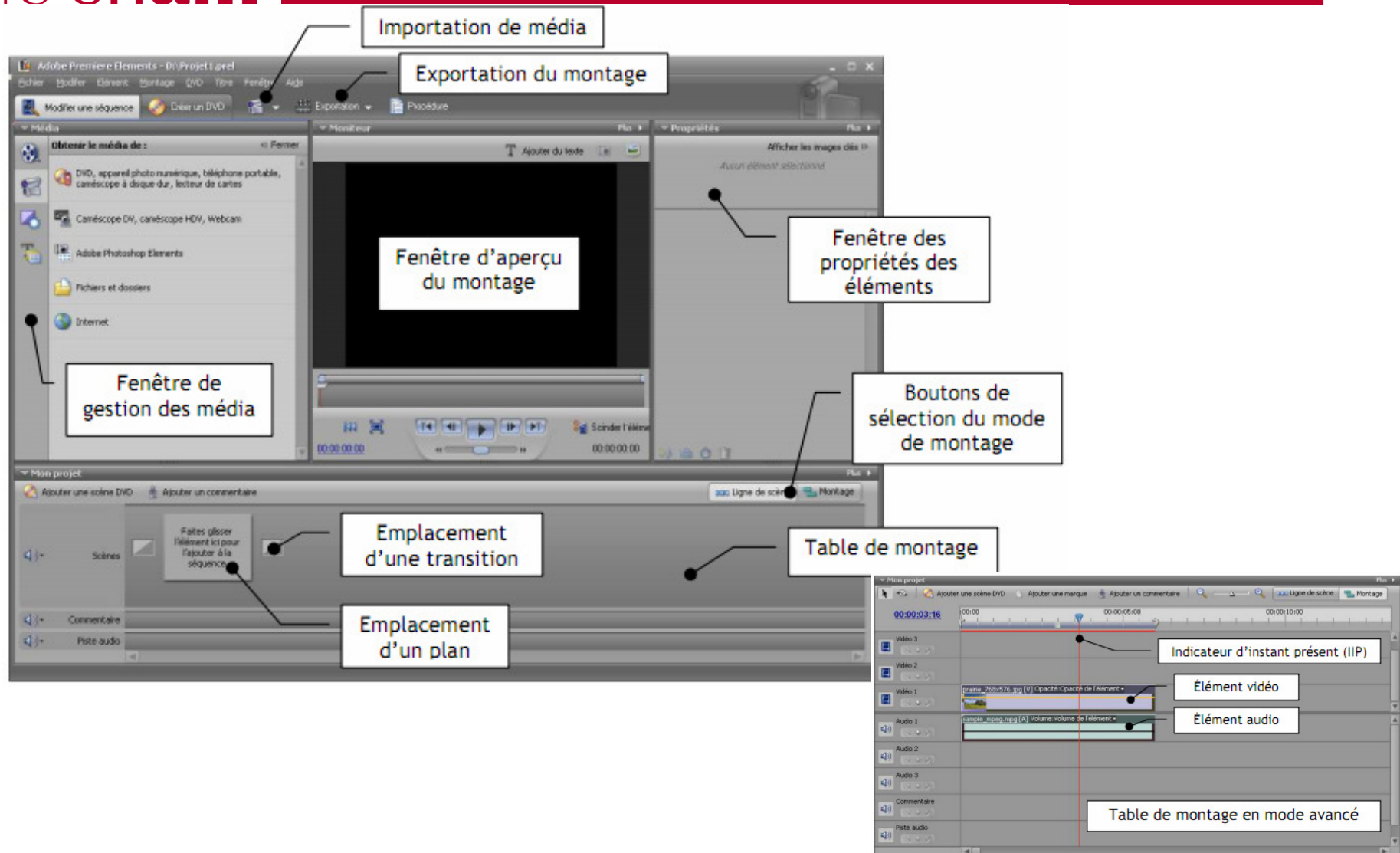
- Créer un projet
- Importer les clips vidéos
- Gérer l'enchaînement des clips sélectionnés (i.e. gérer leur positionnement sur la *timeline*)
- Déterminer les transitions entre les clips
- Ajout de textes ou images fixes
- Sonoriser : ajout de son et/ou de musique



Interface de Windows Movie Maker



Interface d'Adobe Premiere



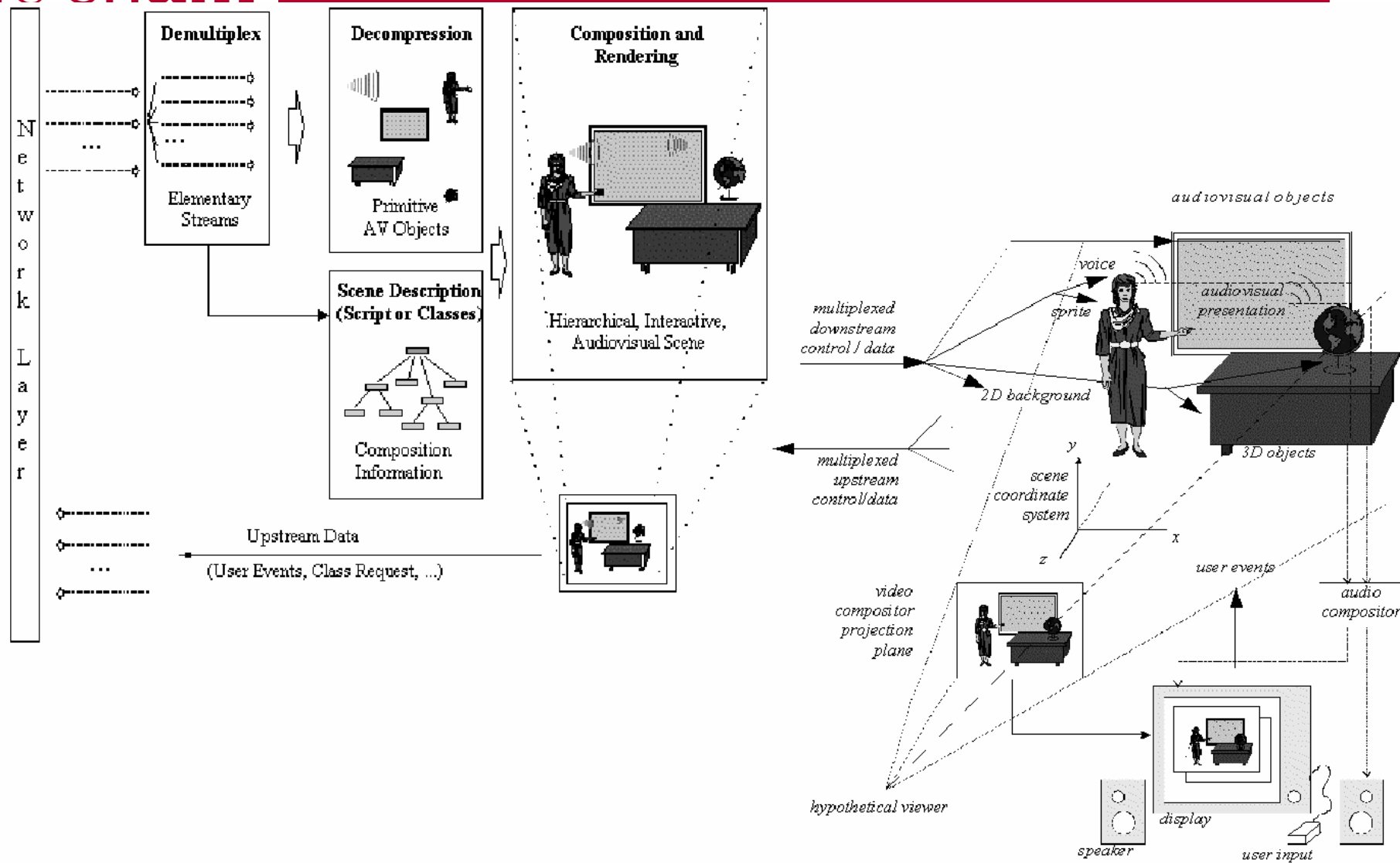
- Vue d'ensemble
 - Présentation générale
 - Fonctionnalités
 - DMIF & Système
 - Les objets audio-visuels

MPEG-4

- Beaucoup plus générale que les précédentes normes
 - Standard défini il y a une dizaine d'années
- Standardisation en vue de supporter les éléments suivants :
 - codage
 - Représenter les contenus audios, vidéos et audio-vidéos
 - composition
 - Décrire comment ces contenus sont assemblés pour former une scène
 - multiplexage
 - multiplexer and synchroniser les informations associées aux médias
 - interaction
 - Intégrer avec l'utilisateur

MPEG-4

le cnam

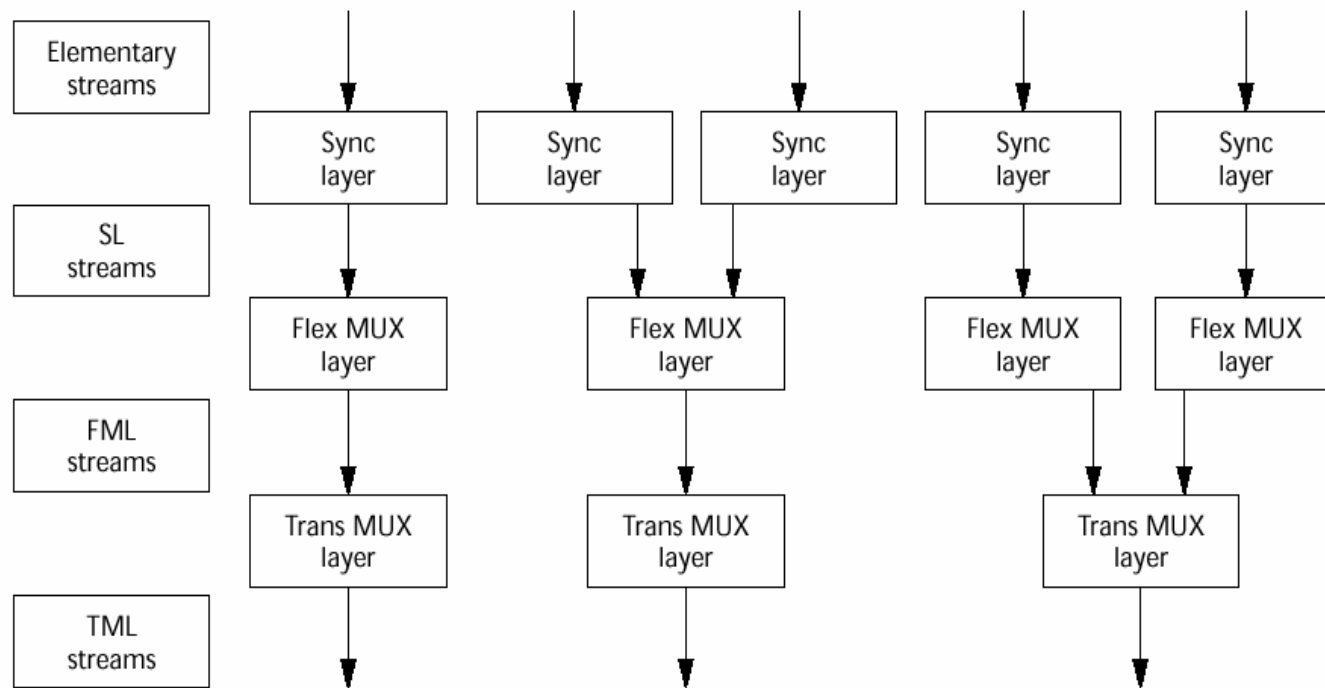


- **Fonctionnalités**

- Représentations codées des médias
 - Image fixe, vidéo, audio, texte, graphiques, synthèse vocale, animation faciale, ...
 - Techniques de compression/décompression séparées
- Composition
 - Les médias sont des nœuds d'un graphe de scène
 - Les nœuds intermédiaires sont des transformations 2D ou 3D
 - Application de flux de données aux médias
 - son, texture, paramètres d'animation, ...
- Modification interactive de la scène par l'utilisateur
 - En local et à distance

- **Fonctionnalités**
 - *Streaming* des données
 - Chaque objet peut être associé avec plusieurs flux de données
 - Flux élémentaires :
 - Identificateur d'objet
 - Descripteur (ressources de décodage, précision, QoS, ...)
 - Unités d'accès (les paquets)
 - Synchronisation
 - Timestamps des unités d'accès (UA)
 - Couche synchronisation
 - » Identification des UAs
 - » Détermination des temps propres au flux
 - » Synchronisation avec les autres

- **Fonctionnalités**
 - Transmission des flux de données
 - Multiplexage sur 2 niveaux :
 - FlexMux
 - Groupement bas niveau de flux élémentaires
 - Réduction du nombre de flux, combinaison des flux avec même QoS, ...
 - TransMux et DMIF
 - Association flux QoS et service de transport
 - » RTP/UDP/TCP, MPEG-2 TS, AAL5/ATM, ...
 - *Delivery Multimedia Integration Framework* (DMIF)
 - » Abstraction des services de transport
 - » Protocole de gestion de session (similaire à ftp)
 - » Gestion des QoS, gestion FlexMux
 - » Interface d'application



- **Fonctionnalités**
 - Interaction de l'utilisateur
 - Navigation dans la scène (point de vue/position d'écoute)
 - Préhension des objets
 - Déclenchement d'événements associés aux objets
 - Sélection de la langue
 - Restrictions spécifiées par l'auteur de la scène
 - Gestion des identifications des propriétés intellectuelles
 - Identificateurs uniques (ISAN, ISRC, ...)
 - Paires clé/valeur (ex. auteur: Alexandre Topol)
 - Interface standardisée pour accéder aux modules de contrôle

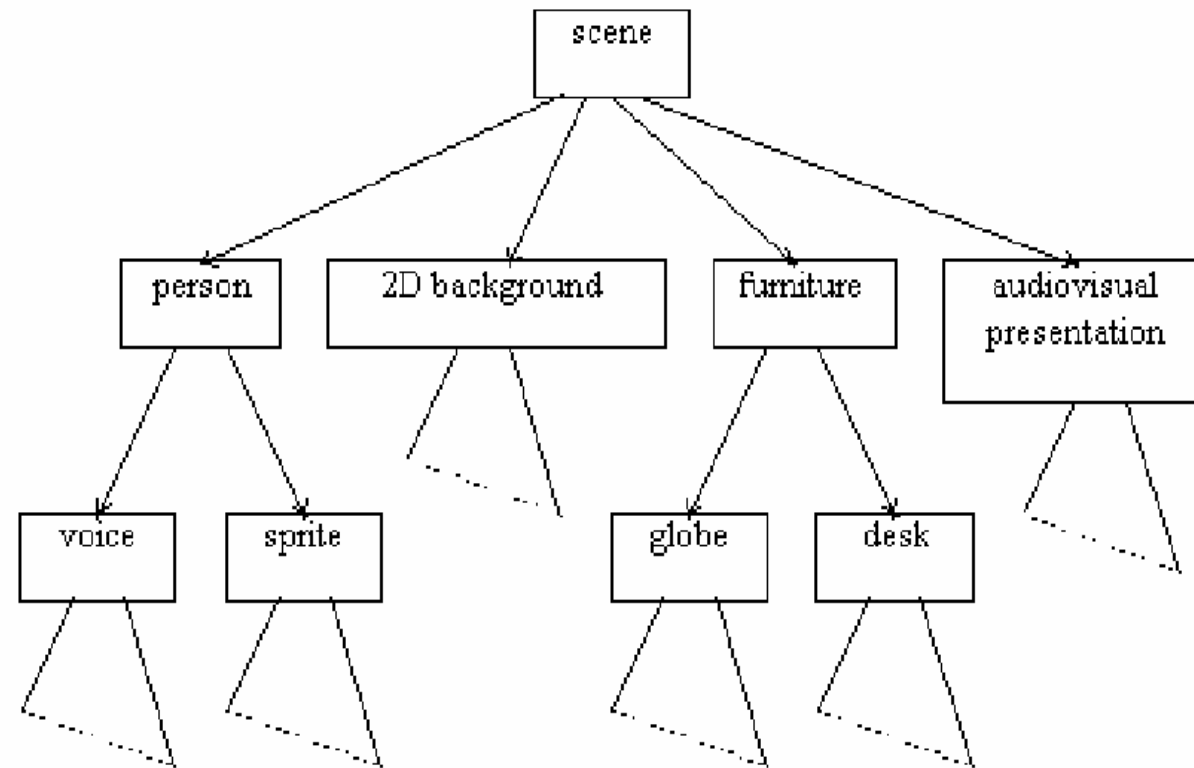
- **DMIF**
 - Similaire à ftp
 - Établissement de connexion au site
 - Requêtes sur les flux avec des paramètres de QoS
 - Renvoie les pointeurs vers les flux d'interfaces TransMux
 - API: DMIF Application Interface (DAI)
 - Différentes sources
 - Réseaux interactifs (internet), broadcast, disque
 - Implémentées grâce à des filtres DMIF (*plugins*)
 - Identification grâce à des URLs
 - Fonctionnement
 - Établissement d'une session locale
 - Etablissement d'une session réseau
 - Sélection à distance d'une application
 - *Streaming*

- **Modèles du décodeur système**
 - Description des décodeurs
 - Couche synchro
 - CRC, time stamp, ...
 - Descripteurs d'objet
 - Nombres et propriétés des flux élémentaires
 - Association entre le média et ses flux élémentaires
 - Transports dans des flux séparés
 - Gestion des buffers
 - Taille paramétrable
 - préchargement

le cnam MPEG-4 : DMIF & système

- **Syntaxe pour la description**
 - Syntaxe C++-like pour la description des flux
- ***Binary Format for Scenes (BIFS)***
 - Encodage binaire de scènes VRML
 - Spécifie :
 - Groupement et instanciation d'objets
 - Positionnement des objets dans le temps et l'espace
 - Accès aux attributs
- **Interaction utilisateur**
 - Interaction du côté client
 - Modification d'attributs et événements VRML (mouse click)
 - Interaction côté serveur par un tunnel retour

le cnam MPEG-4 : DMIF & système



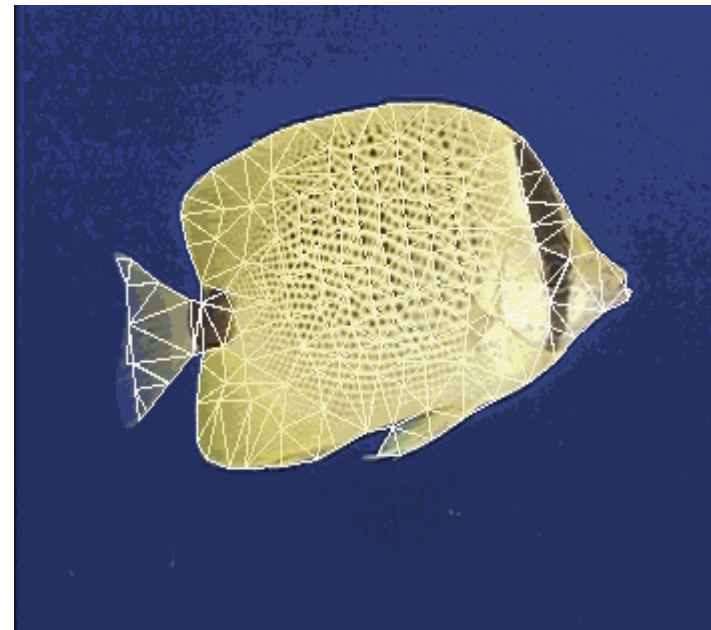
Scene description

le cnam MPEG-4 : DMIF & système

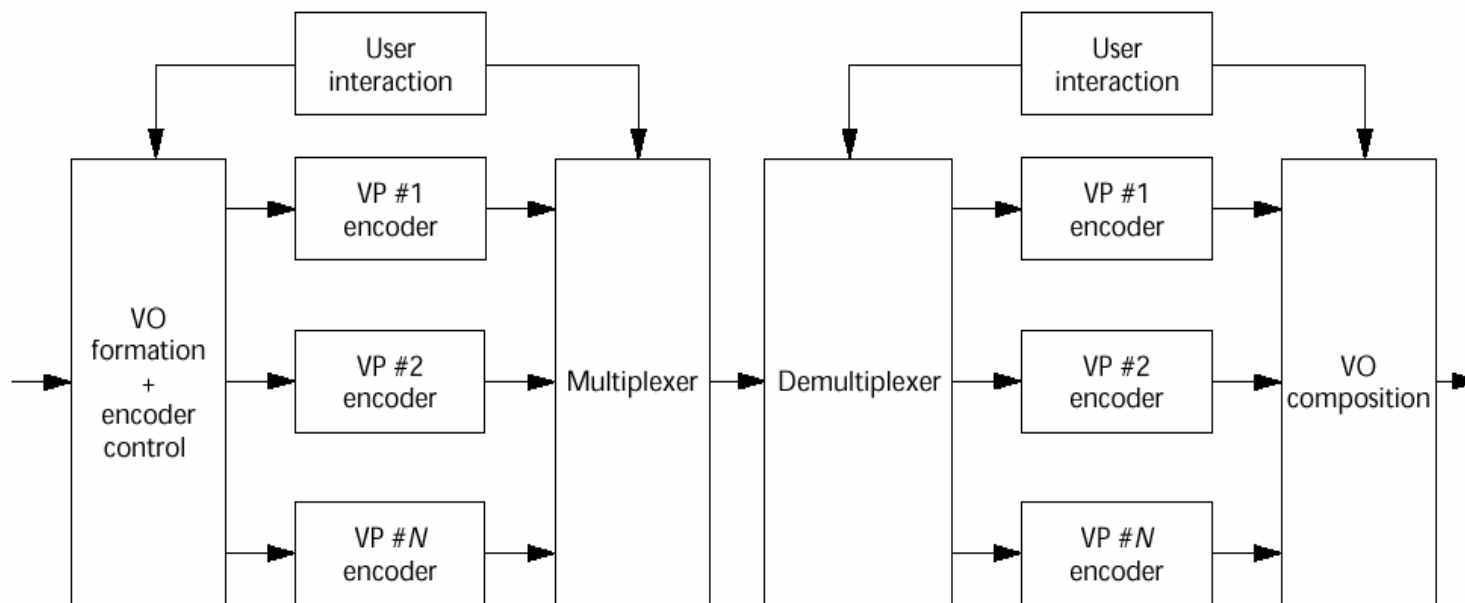
- **Format du fichier mp4**
 - Basé sur QuickTime d'Apple
 - Approche orientée objet grâce aux atomes (tag, length)
 - La plupart des atomes décrivent une hiérarchie de métadonnées
 - Format permettant le *streaming*

le cnam MPEG-4 : objets audio-visuels

- Vidéo
 - Textures, images, vidéos
 - Maillage 2D, animations des maillages
 - Informations au niveau des pixels et masques
 - Objets synthétiques
 - Description paramétriques du visage et du corps
 - Maillages 2D et 3D texturés
 - Techniques MPEG-2 incluses

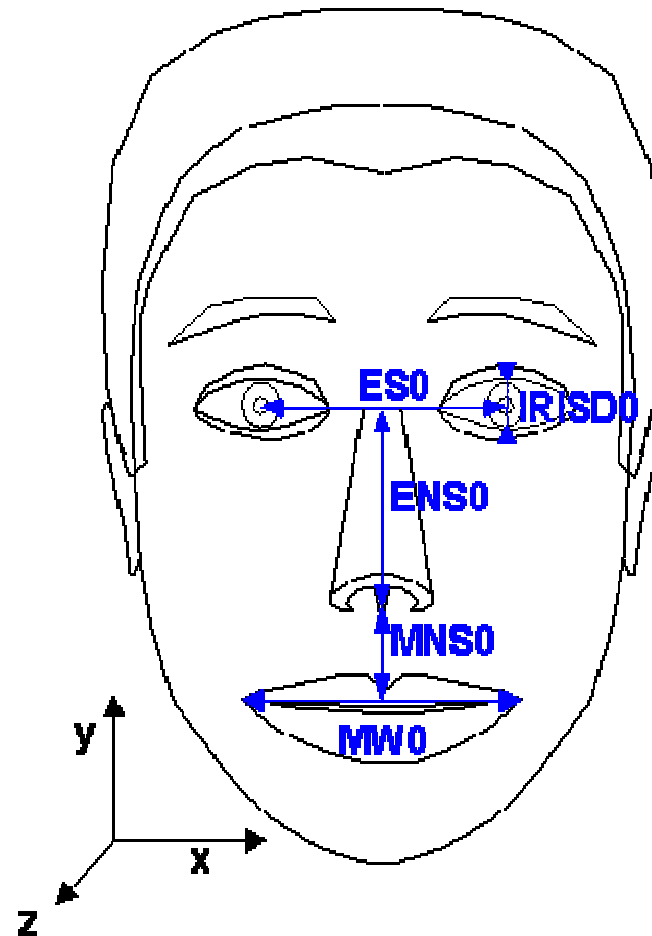


le cnam MPEG-4 : objets audio-visuels



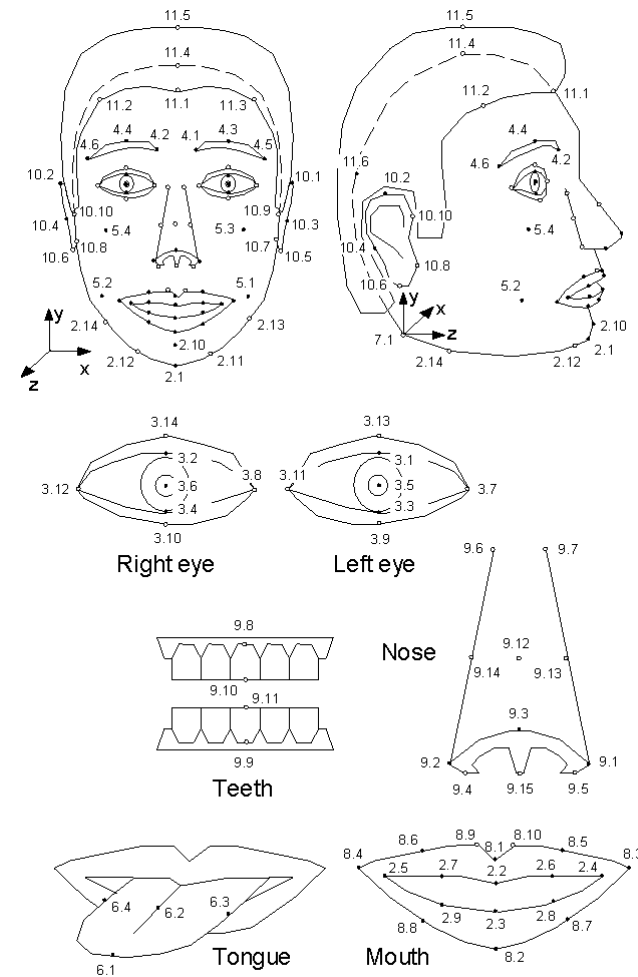
le cnam MPEG-4 : objets audio-visuels

- Exemple d'objets visuels
- Définition faciale
 - Ensemble de mesures caractéristiques



le cnam MPEG-4 : objets audio-visuels

- Exemple d'objets visuels
- Animation faciale
 - Ensemble de points caractéristiques
 - Animation de ces points



• Feature points affected by FAPs
• Other feature points

le cnam MPEG-4 : objets audio-visuels

- Exemple d'objets visuels
- Montage temps réel
 - Les lettres sont imprimées sur l'objet vidéo et suivent le mouvement de manière synchronisée



le cnam MPEG-4 : objets audio-visuels

- **Audio**
 - son naturel (parole et musique)
 - Son synthétique
 - Texte, description instrumentale, paramètres de codage
 - Le codage inclus des effets
 - spatialisation, réverbération, ...
 - Techniques employées
 - Text-to-Speech (TTS) << 6 Kbit
 - Speech coding (HVXC) 2-4 Kbit
 - Speech coding (CELP) 4-24 Kbit
 - General coding (TwinVQ and AAC) > 6Kbit
 - Scalable coding

MPEG-4

le cnam

- A retenir :
 - Peut-être vu comme un sur-ensemble de VRML, Mpeg1 et Mpeg2
 - Architecture très générale, de nombreux concepts
 - Peut reconnaître et mélanger de nombreux différents objets compressés avec différents algorithmes
 - Prend tout en charge (audio, video, 2d, 3d, figures, mesh...)
 - C'est un premier essai d'intégration de la 2D avec de la 3D
 - DivX est une application logicielle qui utilise le standard MPEG-4 pour compresser une vidéo numérique

MPEG-7

le cnam

- Constat
 - De plus en plus d'informations multimédias disponibles
 - De plus en plus d'utilisateurs veulent trouver des informations multimédias spécifiques
 - Difficultés de trouver ces informations
 - Sur le Web
 - Sur les nombreux programmes TV
 - Dans sa propre base de données
 - Les systèmes pour rechercher sur le web existent mais :
 - Surtout pour les informations de textures
 - Difficile d'utilisation

MPEG-7

le **cnam**

- Interface de description des contenus multimédias
- Vers le web sémantique
- Scénarios envisageables :
 - Compréhension d'images (surveillance)
 - Vision intelligente
 - Caméras intelligentes
 - Recherche d'informations
 - Filtrage d'informations

le cnam **Indexation du contenu**

- De manière automatisée
 - Nom de fichiers + métadonnées
 - Analyse image fixe ou animés
 - Colorimétrie, luminosité, contours
 - Analyse sonore
 - Vision par ordinateur
 - Donner un sens aux objets
- De manière pupitrées
 - Saisir les informations de contenu

le cnam Exemple d'ajout de contenu

- Description XML de l'image



Unstructured news image

le cnam Exemple d'ajout de contenu

- Description XML de l'image



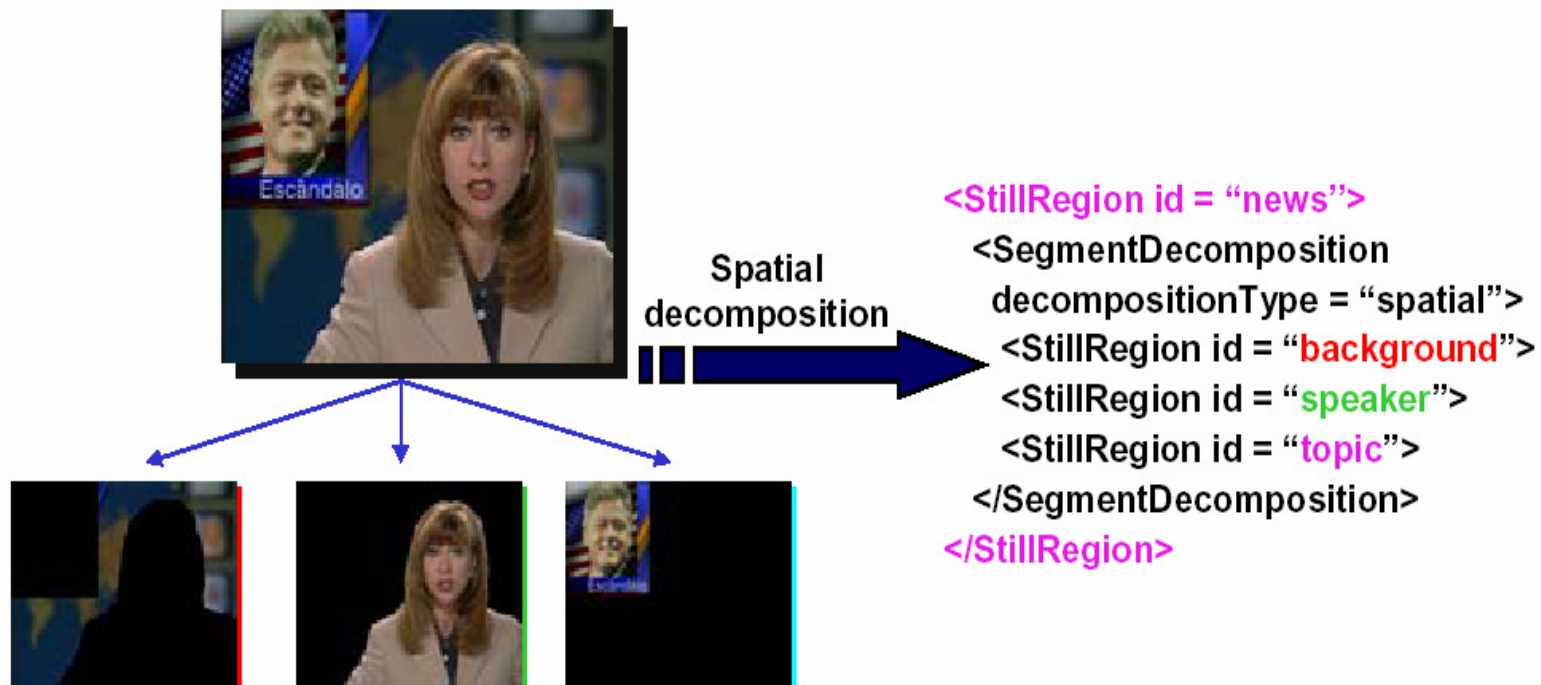
Title

>

```
<StillRegion id = "news">  
</StillRegion>
```

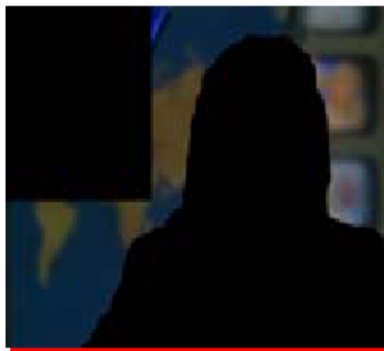
le cnam Exemple d'ajout de contenu

- Description XML de l'image



le cnam Exemple d'ajout de contenu

- Description XML de l'image



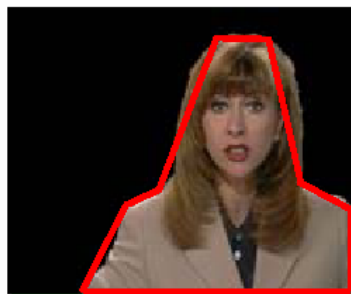
Background
feature



```
<StillRegion id = "news">  
  <SegmentDecomposition  
    decompositionType = "spatial">  
    <StillRegion id = "background">  
      <DominantColor> 10 10 250 </DominantColor>  
    <StillRegion id = "speaker">  
    <StillRegion id = "topic">  
    </SegmentDecomposition>  
  </StillRegion>
```

le cnam Exemple d'ajout de contenu

- Description XML de l'image



More
features
➔

```
<StillRegion id = "news">
  <SegmentDecomposition
    decompositionType = "spatial">
    <StillRegion id = "background">
    <StillRegion id = "speaker">
      <TextAnnotation>
        <FreeTextAnnotation> Journalist Judite Sousa
      </FreeTextAnnotation>
    </TextAnnotation>
    <SpatialMask>
      <Poly>
        <CoordsI> 80 288 100 200 ... 352 288 </CoordsI>
      </Poly>
    </SpatialMask>
    <StillRegion id = "topic">
  </SegmentDecomposition>
</StillRegion>
```

le cnam Exemple d'ajout de contenu

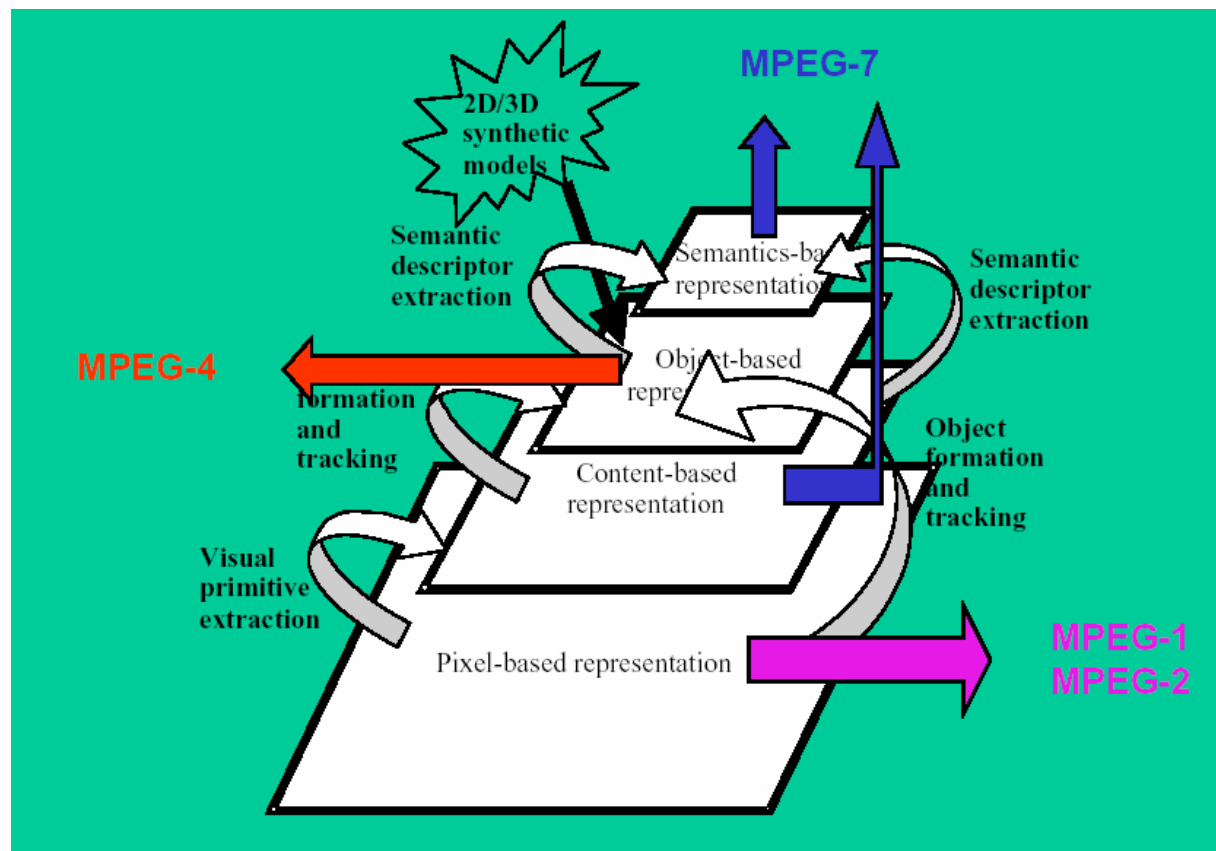
- Description XML de l'image



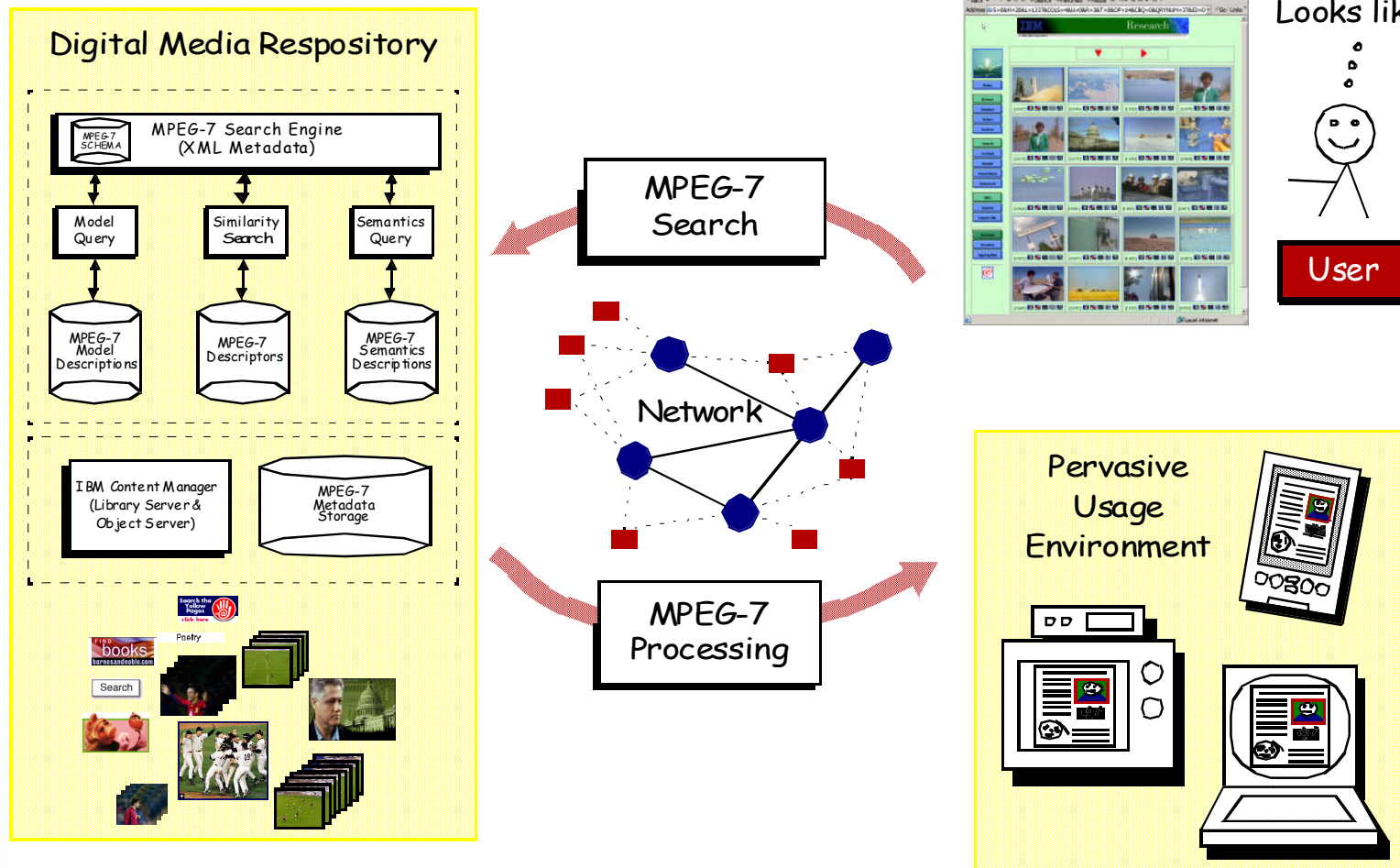
```
<StillRegion id = "news">
  <SegmentDecomposition decompositionType = "spatial">
    <StillRegion id = "background">
      <DominantColor> 10 10 250 </DominantColor>
    </StillRegion>
    <StillRegion id = "speaker">
      <TextAnnotation>
        <FreeTextAnnotation> Journalist Judite Sousa </FreeTextAnnotation>
      </TextAnnotation>
      <SpatialMask>
        <Poly>
          <CoordsI> 5 25 10 20 15 15 10 10 5 15 </CoordsI>
        </Poly>
      </SpatialMask>
    </StillRegion>
    <StillRegion id = "topic">
      <TextAnnotation>
        <FreeTextAnnotation> Clinton's affair</FreeTextAnnotation>
      </TextAnnotation>
    </StillRegion>
  </SegmentDecomposition>
</StillRegion>
```

MPEG7

- Structuration pyramidale de l'information



MPEG7

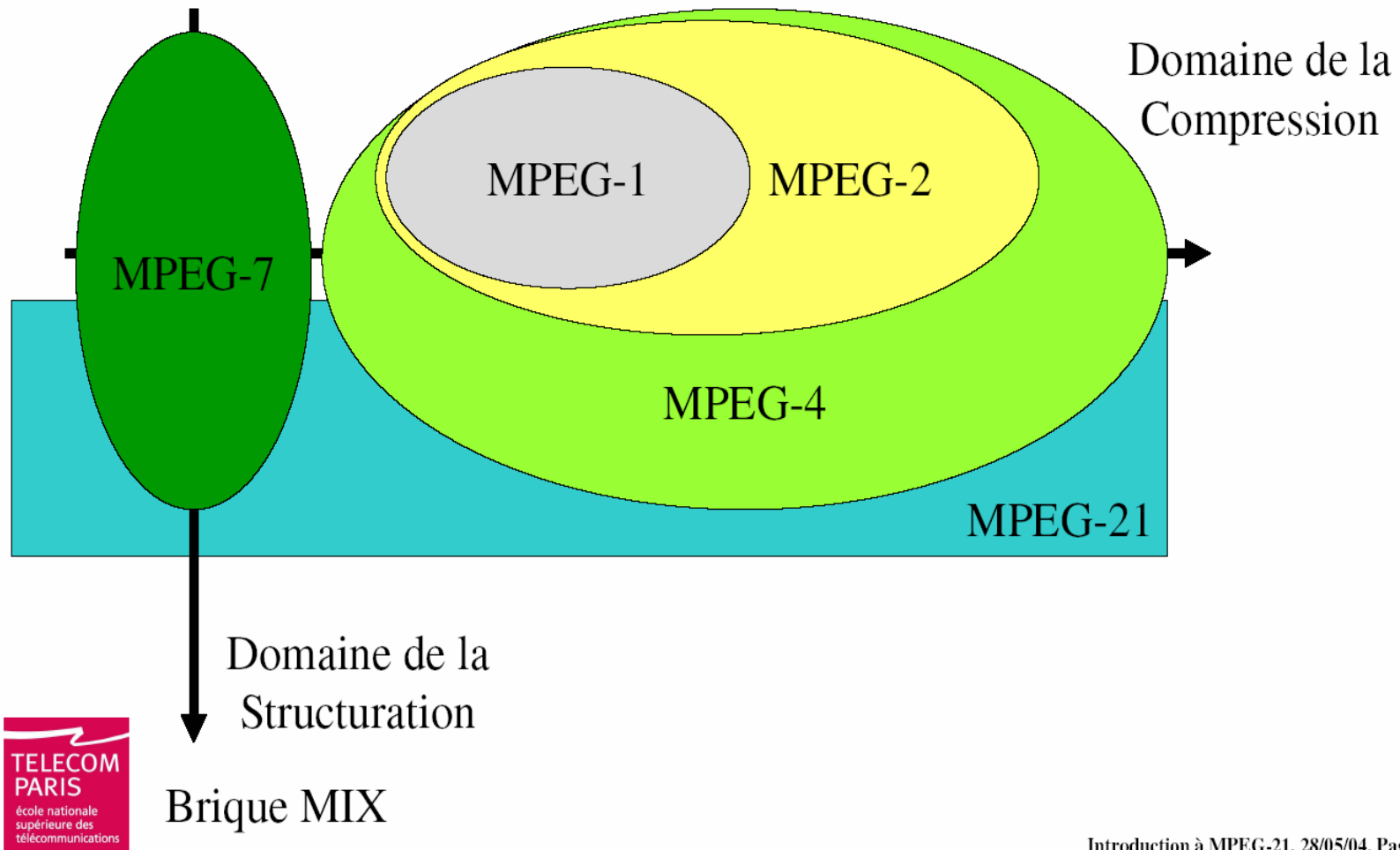


MPEG-21

le cnam

- Constat
 - De nombreuses infrastructures existent pour le transport et la consommation de contenus multimédias
 - Ces infrastructures diffèrent d'une communauté à l'autre
- Motivation
 - MPEG-21 donne un cadre pour décrire ces différents éléments pour qu'ils fonctionnent ensemble
 - *Multimedia framework*

Le « monde » MPEG



le cnam Webographie

- Le cours de Philippe Martin (dont je me suis grandement inspiré)

<http://www.sciences.univ-nantes.fr/info/perso/permanents/pmartin>

- Comment ça marche : les codecs

<http://www.commentcamarche.net/video/compvid.php3>