

Modélisation et indexation de documents audiovisuels

Yannick Prié
LIRIS – Université Lyon 1
DEA DISIC / 2002-2003

Argument

- Documents audiovisuels
 - documents multimédia éventuellement séquentiels,
 - en tout cas non textuels
- Conséquence
 - toute instrumentation passe obligatoirement par une annotation,
 - le plus souvent manuelle

2

Une image...

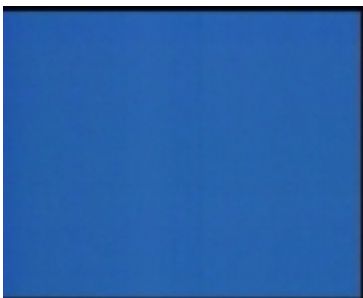


3

Du son

4

Des images animées et du son



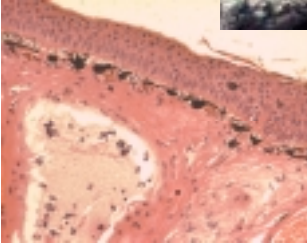
5

Plan


- Image, son, **audiovisuel**
- ... numérique
- ... systèmes d'information
- ... normalisation
- ... sujets chauds
- ... systèmes locaux

6


Images



Epithélium



Harrisburg



Images : poignées de mains



Jewish Agency
Chairman
Sallai Meridor
with Russian
President
Vladimir Putin
(Photo: Yaacov
Saar, Israel
Government
Press Office)

8

Images : poignées de mains

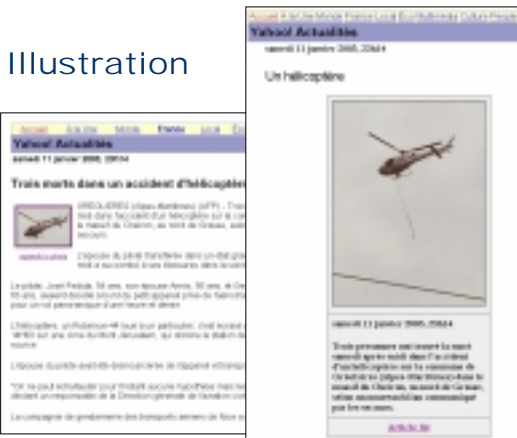


French President Jacques Chirac was the guest of honour at the Paris airshow yesterday and used part of his time at Le Bourget presenting an award of courage.

The French Legion d'Honneur was given to Juan Garcia for his courage when tackling the South American mountain range, the Andes, back in 1930.

The President, accompanied by dignitaries and military heads representing countries around the globe, was later treated to an aerial spectacle by some of the best aircraft France currently has to offer. Before his flying visit came to an end he toured the static display of Canadian aircraft manufacturer Bombardier.

Illustration




Un hélicoptère

Le vol 101 de la compagnie Air Inter, un Airbus A300-600, a été impliqué dans un accident d'hélicoptère le 12 mai 1994, au-dessus de la mer Méditerranée, au large de la Corse, sans survivants.

Le vol 101 de la compagnie Air Inter, un Airbus A300-600, a été impliqué dans un accident d'hélicoptère le 12 mai 1994, au-dessus de la mer Méditerranée, au large de la Corse, sans survivants.

Le vol 101 de la compagnie Air Inter, un Airbus A300-600, a été impliqué dans un accident d'hélicoptère le 12 mai 1994, au-dessus de la mer Méditerranée, au large de la Corse, sans survivants.

En bref



- Une image
 - ne signifie pas grand chose par elle-même
 - possède toujours un contexte
 - est toujours interprétée dans ce contexte
 - de façon pertinente ou non

11

Images animées + son

- Temporalité du flux d'images
 - 25 images / secondes
 - persistance rétinienne
 - illusion de réalité
- Superposition de flux
 - son
- Séquentialité
 - montage
- Document audiovisuel

12

Quelques définitions

■ Plan

- ensemble d'images tournées sans arrêt de la caméra
- unité de montage vidéo
- 80 min vidéo = 500/1000 plans
- notion d'image(s) clé (keyframe)

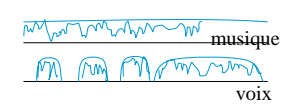


13

Quelques définitions (suite)

■ Montage

- superpositions de flux
 - vidéos : incrustations
 - audios : musique, parole, bruitages (Foley)
- structuration
- effets de contexte
- transitions



14

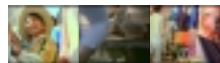
Quelques définitions (suite)

■ « Grammaire » audiovisuelle

- mouvements et raccords
- situations
 - dialogue : champ / contre-champ
- narration

■ Scène

- ensemble de plans contigus possédant une cohérence sémantique



*raccord sur le regard
*raccord sur un geste
*raccord dans l'axe
*raccord sur la bande-son
*raccord de mouvement
*raccord sur une couleur
*raccord sur une forme
*raccord « lumière »
*raccord sur une composition plastique
*raccord sur le décor
*...

15

Différence d'avec le texte

■ Spatialité vs temporalité

- exploration vs flux de conscience

■ Support et forme d'appropriation différents

- reconstruction obligatoire
 - dispositif technique
- la lecture contrôlée/donnée par le médium : processus temporel imposé par le document.

■ Question immédiates

- comment appréhender la globalité du flux ?
- comment faire pour aller y chercher une information ?

16

Retour à OK Coral

Durée ?

Nombre de plans ?

Technique ?

Bande-son ?

Personnages ?

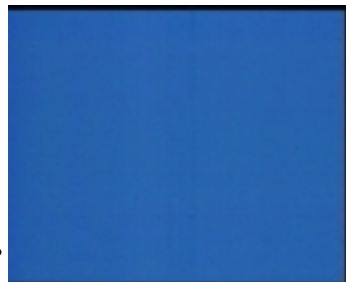
Objets ?

Mouvements de caméra ?

Texte ?

Couleurs ?

17



Retour à OK Coral

■ 30 secondes / 15 plans.

■ Plans tournés, palette, inserts

■ Bande-son

(Musique : guitare)

(Musique : démarrage chant)

Femme : - C'est mon jour, je sens que je vais trouver.

- Allez, je suis mon instinct.

- Oh, le voila, ce motif, ce bleu, j'adore.

Homme : - Vous l'aimez ? Alors confiez-le à Coral.

- Avec une lessive classique, lavage après lavage, les couleurs perdent leur éclat.

- Coral lui, protège activement les fibres et garde les couleurs éclatantes.

- Alors pour garder longtemps ce que vous aimez, faites confiance à Coral.

- Coral, le spécialiste des couleurs.

Histoires, formats et usages

- Cinéma
- Télévision
- Vidéo personnelle
 - super 8
 - VHS
 - DVD (les enregistreurs sont là)
 - Set Top Box (DD) : 30 Go, 15 h MPEG2
- Stockage distribué ?
- Loisirs, actualité
- Actualités, loisirs, documentaires, chaînes généralistes, puis spécialisées
- Films de vacances, enregistrement télé, vidéo club
- Nouveau usages, SIAV

19

Et puis tout le reste...

- Surveillance
- Visioconférence
- Éducation
 - support pédagogique
 - documentaire
 - analyse
 - manuels d'entretien
- Recherche : DAV comme
 - témoignage du passé, œuvre d'art
 - notes personnelles
 - acte de communication
- Le retour de la lecture active

20

Qui a le droit ?

- Principalement le droit d'auteur
 - des intervenants de tous types
 - des organismes multiples
 - des informations « distribuées »
- Dans la majorité des cas
 - réutiliser un DAV vous met hors-la-loi
- Pour la recherche
 - problème : conférences, démonstrations en ligne
 - tourner des DAV libres de droit ?

21

Plan

- Audiovisuel
- **Audiovisuel numérique**
- Systèmes d'information audiovisuelle
- Normalisations
- Problématiques actuelles
- Systèmes locaux
- Conclusions

22

Numérisation de base

- Une idée simple
 - un flux d'images
 - des flux de sons
 - une synchronisation
 - ... un ordinateur assez puissant pour permettre la visualisation
- Oui mais
 - 720 x 480 x 24 x 25 → 249 Mo/s, 112 Go/h
- Il faut compresser

23

Méthodes de compression

- Comprimer les images et le son
 - compression spatiale
 - GIF98a, MPEG1 (1,5 Mo/s : Cédérom)
- Tirer partie des propriétés du médium
 - compression temporelle
 - QT, MPEG2 (2-20 Mo/s : DVD, télé numérique), H263, RA (Web, streaming)
- L'audio suit
- Aller plus loin
 - objets, hybridation naturel synthétique, QoS
 - MPEG4, DivX

24

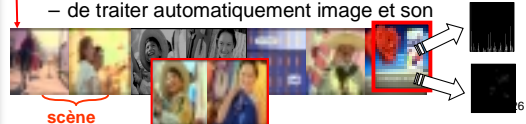
Facteurs importants pour la compression AV

- Débit
- Format / qualité
- Complexité de codage/décodage
- Résistance aux erreurs
- Accès aléatoire (random access)
- Interopérabilité

25

Et puis...

- La numérisation permet de mimer l'analogique par lecture
 - à partir d'un support
 - sur le réseau (streaming)
- mais elle permet également
 - d'avoir un accès direct à une partie d'un document
 - de documenter les flux
 - de manipuler, d'éditer
 - de traiter automatiquement image et son



6

...on veut tout...

- Indexation / recherche
- Manipulation
- Navigation
- Visualisation
- Filtrage, montage personnalisé

27

... pour tout faire

- Digital libraries (image or video catalog, musical dictionary,...)
- Multimedia directory services (e.g. yellow pages)
- Broadcast media selection (radio channel, TV channel,...)
- Multimedia editing (personalized electronic news service, media authoring)
- Education (sport, sociologie anthropologie, surveillance...)
- Journalism (e.g. searching speeches of a certain politician using his name)
- Tourist information
- Entertainment (e.g. searching a game, karaoke)
- Medical applications
- Shopping (e.g. searching for clothes that you like)
- Architecture, real estate, interior design
- Social (e.g. dating services)
- Film, Video and Radio archives
- Industrie (manuels d'entretien, vidéoconférence...)

28

Remarques

- Passage au numérique
 - La numérisation ne change pas le statut du document audiovisuel par rapport au dispositif technique.
 - Pour l'image et le texte, c'est un peu différent, car la peinture, le marquage sur le monde existaient avant le texte même, qui après une période de désignation de la réalité par des signes, est rapidement passée à la stabilisation de la parole.

29

Plan

- Audiovisuel
- Audiovisuel numérique
- **Systèmes d'information audiovisuelle**
 - Images
 - Sons
 - Vidéos
- Normalisations
- Problématiques actuelles
- Systèmes locaux
- Conclusions

30

Problématique

- Décrire des images avec des descripteurs sur lesquels on puisse faire des recherches à l'aide d'une machine.
- Remarque
 - un texte est sa propre indexation, ce n'est pas le cas pour une image « naturelle »

31

Grands domaines

- Bases d'images généralistes
 - Gestion des droits, etc.
 - ex : Corbis, Gettymage
- Web
 - ex. Google, Altavista, Webseek, Ditto, Amore...
- Bases d'images spécialisées
 - archéologie (ex. Maison de l'Orient)
 - marques (ex. INPI)
 - documents anciens (ex. facsimilés)
 - SIG (ex. municipalités)
 - médecine (ex. expertise analyse)
 - etc.
- Bases d'images personnelles

32

Indexation automatique image

- Approches statistiques
 - couleurs, textures, descripteurs globaux
- Approches structurales
 - image = ensemble de régions spatialement organisées
 - segmentation
 - détection visages, points d'intérêt, zones cellulaires...
 - classification
 - reconnaissance de visages

33

Recherche « par le contenu » "content-based" retrieval

- Requêtes
 - images
 - dessins
- Recherche
 - similarités
- Bouclage de pertinence
 - relevance feedback
 - exemples positifs et négatifs.
- Résultats
 - divers...
- Approches « perceptivo-centrées » ?

34

Indexation globale textuelle

- Automatique
 - Texte autour de l'image (web)
 - Marche pas si mal
- Manuelle
 - Cf. notices documentaires
 - DC++
 - Mots-clés
 - Vocabulaire possible : thésaurus iconographiques plus ou moins spécialisés
 - Textes descriptifs
 - les mots-clés mis manuellement sont plus pertinents que les termes extraits automatiquement
- Systèmes
 - Extensions de systèmes de RI standards, peu de fonctions avancées
 - [Rodden & Wood, 2003] : "Our results suggest that participants found their digital photos much easier to manage than their non-digital ones, but that this advantage was almost entirely due to the simple browsing features. The advanced features were not used very often and their perceived utility was low."

35

Annotations localisées

- Texte / formes géométriques
- Annotation manuelle
- Beaucoup de systèmes payants
 - marché entreprises / médical
- Stockage des annotations dans l'image ou sur serveurs dédiés

36

Et autour du web sémantique ?

- Annotea
 - Annotation d'images en passant par SVG
- RDFPic
 - descriptions RDF dans les images JPEG
- RDFWeb
 - <http://ibbering.com/svg/AnnotateImage.html>
 - Annotating Areas of Images with RDF - This is a HTML Form (for convenience) and SVG document which allows you to annotate areas of documents saying what the WordNet property of the area is, this can then later be searched, if the area is a person, then you can also add name and email address for codepiction
- Annotations à base d'ontologies
 - Cf. exposé

37

Plan

- Audiovisuel
- Audiovisuel numérique
- **Systèmes d'information audiovisuelle**
 - Images
 - **Sons**
 - Vidéos
- Normalisations
- Problématiques actuelles
- Systèmes locaux
- Conclusions

38

Remarques

- Domaine temporel
 - durée fixée d'audition d'un document
- Annotation
 - intemporelle (document)
 - temporellement située (fragment)
 - Visualisation ?
 - Utilisation ?

39

Grands domaines

- Bases de recherche
 - linguistique
- Bases musicales
 - institutionnelles (ex. IRCAM)
 - personnelles
- Bases d'entreprise
 - discours
- Bases radio
 - stations
 - archives

40

Approches manuelles

- Notices
 - Systèmes professionnels
 - RI standard + capacités de navigation
 - Partage
 - P2P
- Graphes d'annotations
 - linguistique
 - MATE, AGTK, ATLAS
 - outils de transcription
 - Transcriber

41

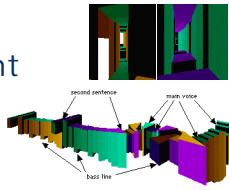
Approches automatiques

- Entour textuel
 - Web : moteurs de recherche altavista, webseek, mediasite.
- Reconnaissance de la parole
 - Base d'interviews ?
- Travaux sur la musique
 - siffler une mélodie...

42

Enjeux du moment

- Numérisation pour les DL / workflow
- Outils de visualisation
 - résumés
- Intégration
 - recherche classique
 - contenu
- Documentation des DL
 - liens entre partitions et interprétations diverses et variées (corpus)
- Plus amusant
 - Couloirs 3D
 - bijection entre éléments géométriques et éléments structurels de la musique (plancher/ligne de basse, plafond/mélodie)
 - « codage nécessitant l'annotation de la musique, car aucune des deux représentations, que ce soit les fichiers MIDI ou les fichiers audio, n'a véritablement de sens musical. Nous avons choisi le système de notation musicale GUIDO [HOOS96] pour annoter les fichiers pués. GUIDO est fondé sur un langage formel, indépendant de toute plateforme technique, préparé et stocké sous forme textuelle, directement lisible par l'utilisateur, axé sur des concepts musicaux (capable de représenter des agrégations comme des accords et des phrases), et non sur des caractéristiques graphiques. »



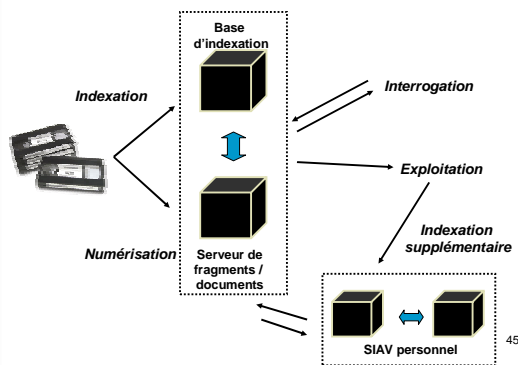
43

Plan

- Audiovisuel
- Audiovisuel numérique
- **Systèmes d'information audiovisuelle**
 - Images
 - Sons
 - **Vidéos**
- Normalisations
- Problématiques actuelles
- Systèmes locaux
- Conclusions

44

SIAV : généralités



45

Un SIAV

- Trouver des documents et les visualiser
 - numérisation si besoin, restauration
 - indexation (auto, manuelle)
 - stockage des fragments AV (formats AV)
 - stockage des index (formats de description)
 - requêtes sur les index
 - visualisation des résultats
 - visualisation de flux AV (évt. générés)
- Manipuler des documents / des fragments / des annotations
 - intégration dans son SI personnel
 - montage, génération de documents
 - hypervidéos

46

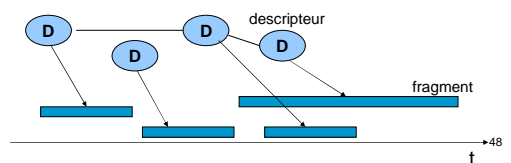
Indexation « à l'ancienne »

- Document AV en général sur bande
- Description = notice
- Fragment = document
- Usage large
 - RAMEAU par exemple est prévu pour indexer des documents audiovisuels
 - bibliothèques, médiathèques, INA
 - un SIAV qui marche très bien : le forum des images
- Mêmes avantages / inconvénients que les systèmes standards
 - plein texte, thésaurus, indexation à la main, etc.

47

Excursus : fragmentation

- Fragments de documents
 - ne pas nommer les unités temporelles *a priori*
 - créer ces unités au besoin, au moment où elles deviennent utiles, quand elles rentrent dans une description



48

Indexation automatique

- Image fixe
 - Approches statistiques / structurelles
 - Recherche
 - similarités
- Application au temporel
 - un plan est représenté par une image-clé
 - indexer l'image clé = indexer le plan
- Mais
 - la bonne vidéo ?
 - pas d'application (AMC)
- Intéressant
 - reconnaissances diverses
 - visages : animateur, journaliste
 - décor du JT de France2

49

Indexation automatique (suite)

- Image animée
 - détection de plans (cut)
 - proposition de scènes
 - mouvements
 - objets, caméra
 - requête = contraintes spatio-temporelles
- Son
 - segmentation (parole, musique)
 - reconnaissance de locuteur
 - extraction de texte, calage (timecode)
- Pas encore très convaincant
 - recherche de quoi ?

50

Problèmes

- Textes
 - machine et homme s'accordent sur l'alphabet et sur les mots
- Document vidéo
 - pas d'accord sur des unités minimales
 - interpréter les descripteurs extraits
- Objectif (Gaal)
 - descripteurs universels
 - mise en correspondance adaptée
- Conclusion
 - se limiter à de la recherche sur des choses précises (dont la sémantique peut être exprimée formellement)
 - assistance dans des domaines précis
 - primitives psycho-visuelles ?
 - mais...
 - TREC 2002 Video track : cut / classification, à suivre

51

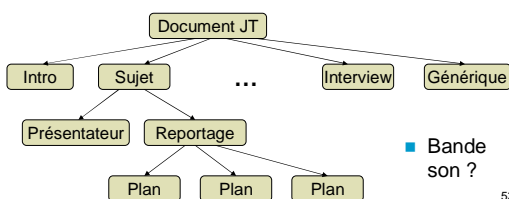
Le retour de la langue

- ... et de l'être humain
 - pas de réalité *a priori*
 - projection d'une attente (modèle)
- Paraphrase d'indexation
 - localisation : fragments
 - sémiotisation : langue (concepts, structure)
- Les unités de manipulation deviennent des unités de signification
 - de quoi ça parle ?
- Le cas du plan
 - accord humain/machine sur cette unité
 - une unité de signification tout le temps ?

52

Approches documentaires

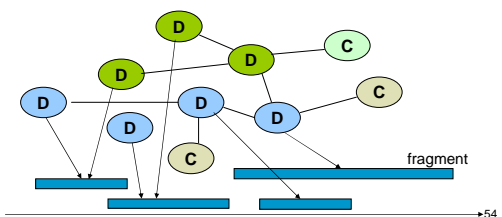
- A l'imitation des descriptions des documents textuels : descriptions structurelles
- Structure arborescente de description
- SGML/XML
 - ex. VANE



53

Approches « conceptuelles »

- Généralisation de l'approche documentaire
- Décrire les documents et leurs manipulations
 - multiples schémas, ontologies...
 - ex. filmage, structure montage, impressions...



54

Les connaissances

- Connaissances de description
 - traitements de corpus textuels
- Descriptions comme connaissances
 - ontologies / schémas de description
 - index, structures, connaissances : même combat
- Indexation « intelligente »
 - offrir dans son mode d'utilisation même la possibilité de l'interroger en tant que connaissance
- Système d'information « idéal »
 - descriptions structurées de documents (internes ou externes) instrumentant leurs exploitations
 - exploitation
 - inférences plus ou moins formalisées sur les index
 - système entièrement géré par les connaissances
- Cf. Web sémantique

55

Systèmes de gestion

- Diffuseurs / archiveurs
 - Workflow
 - Plans, KF, annotation manuelle
 - + evt. parole, reconnaissance de visage, identification de locuteur
 - Virage, Excalibur, Mediasite (Informedia)
 - Exemples
 - CNN, MGM, ABCNews, TF1, etc.
- Amateurs
 - bases libres : ex. <http://www.open-video.org/>
 - enregistreurs personnels
 - quelques fonctionnalités d'indexation
 - systèmes de montage perso
 - Imedia (Apple)
 - Premiere (Adobe)

56

Applications éducatives

- Cours sous la forme vidéo
 - Prise de note texte / audio
 - VSA : Video Sharing and Annotations
 - MRAS (Microsoft Research Annotation System)
 - texte / audio
 - des études d'usage !
 - Passage à l'hypertexte
 - E-learning
 - navigation texte / vidéo
- Pièces de théâtre en vidéo
 - discussions / comparaison

57

Applications SHS

- Les documents audiovisuels comme sources
 - lecture active
 - partage
- Exemples
 - Inathèque
 - Gestion de corpus : SLAV, Videoscribe / Mediacorpus
 - KRONOS
 - Événements : classification / analyse
 - OPALES
 - Portails pour des communautés de chercheurs
 - MMVIS
 - Analyse de tendances temporelles (matches de volley ball)
 - ANVIL, Eudico, Mediatagger
 - Linguistique

58

Un système intégré

- Modéliser de façon globale la gestion des flux
 - de la pré-production à la diffusion et à l'archivage, en passant par le montage
 - A4SM (Authoring System for Syntactic, Semantic and Semiotic Modelling)
 - Application aux news
 - 18 schémas !

59

Plan

- Audiovisuel
- Audiovisuel numérique
- Systèmes d'information audiovisuelle
- **Normalisations**
- Problématiques actuelles
- Systèmes locaux
- Conclusions

60

Normalisation images

- JPEG2000
 - possibilités d'avoir plusieurs couches (layers) → pour annotation
 - Pas de normalisation de descripteurs
- EXIF
 - <http://www.exif.org/>
 - Exchangeable Image File Format
 - appareil photos numériques / JPEG
 - schéma plutôt technique photo, proposition d'extension pour les lieux
- DIG35
 - http://www.ija.org/i_dig35.html
 - IJA : International Imaging Industry Association
 - description XML intégrée au flux
 - informations
 - création
 - description
 - historique
 - gestion des droits
- ...



Example Metadata
 Photographer: J.Hunt
 Date: 6/7/99
 Location: Steam River
 Subject: Golden Trout
 Focal Distance: 2 feet
 Zoom: 2-to-1

1

Normalisation AV

- Objectif n° 1
 - caler des informations sur un flux AV,
 - de façon standardisée
 - pour avoir des index
- Objectif n°2
 - normaliser les index
 - ce qu'il y a lieu de décrire
 - comment le décrire

62

MPEG 7 - Généralités

- Multimedia Content description for AV material
- Groupe MPEG
- Démarrage : octobre 1998
- Standard ISO : septembre 2001
- Deux mondes
 - traitement du signal (image, son)
 - document

63

MPEG 7 - Objectifs

- Standardiser les descriptions de contenu
 - décrire divers aspects du contenu : ensemble de descripteurs
 - caractéristiques de bas niveau
 - structure
 - sémantique
 - collections
 - proposer des structurations de descripteurs
 - permettre d'en définir de nouveaux
- Localisation des métadonnées
 - dans les flux
 - dans des bases de données.

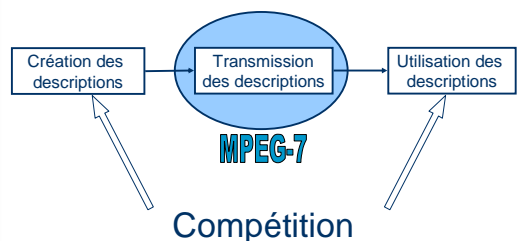
64

MPEG-7 - Usages visés

- Faciliter la gestion des documents AV
 - rechercher, filtrer, organiser, naviguer
- Son — Karaoke, vente, bibliothèques d'effets sonores, BD de discours historiques, recherche de scènes par événements auditifs, recherche de musique par mélodie
- Audiovisuel — Préférence utilisateur : guide/résumé des programmes/flux adaptés, Archives : descriptions et échanges de descriptions, Graphiques : dessiner quelques lignes et retrouver des images avec les mêmes graphique/logos/caractères, Mouvement : décrire des relations/mouvements d'objets AV, et retrouver des segments correspondants, Scénario : décrire des actions dans un flux, retrouver un document correspondant, Surveillance : détection de mouvements suspects, Navigation : voir un résumé de 15 minutes d'un match.

65

MPEG-7 - Portée du standard



- Ambiguïté quand on essaye de normaliser des descripteurs !

66

Éléments / terminologie

- **D : Descriptors**
 - syntaxe et sémantique de caractéristiques AV
 - DominantColor, AudioSpectrumCentroid
- **DS : Description Schemes**
 - structure et sémantique des relations entre les composants d'une description (D ou DS)
 - Segment, HierarchicalSummary

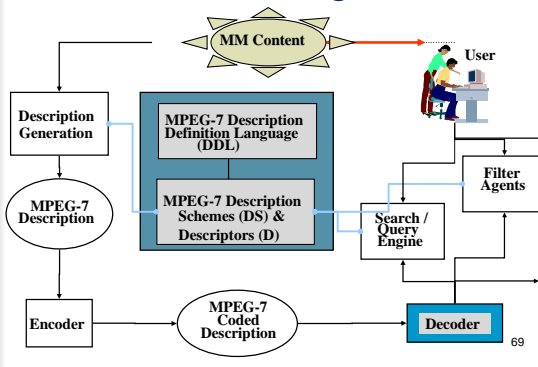
67

Éléments / terminologie (suite)

- **DDL : Description Definition Language**
 - langage permettant la création des DS et D
 - langage permettant l'extension des DS et D
 - XML Schema étendu
- **System tools**
 - Codage, décodage, multiplexage des descriptions
 - « Streaming » des descriptions

68

MPEG-7 – Schéma général



69

MPEG-7 Exemple 1



T. L'hermitte

```
<Person>
  <Name xml:lang="en">
    <GivenName>Thierry</GivenName>
    <FamilyName>L'hermitte</FamilyName>
  </Name>
  <Affiliation>
    <Organization>
      <Name>Independent cinema company</Name>
    </Organization>
  </Affiliation>
</Person>
```

70

MPEG-7 Exemple 2



Motion

```
<Segment xsi:type="MovingRegionType">
  <TextAnnotation>
    <FreeTextAnnotation xml:lang="en">
      Person</FreeTextAnnotation>
    </TextAnnotation>
  <MediaTime>
    <MediaTimePoint> 00:00:15 </MediaTimePoint>
    <MediaDuration> 00:00:30 </MediaDuration>
  </MediaTime>
  <ParametricObjectMotion model="Translational">
    ...
  </ParametricObjectMotion>
</Segment>
```

72

Parties de MPEG-7

- Systems - codage et transport (XML / BiM / streaming)
- DDL - Langage de schéma (XML Schema + ...)
- Audio - Descripteurs audios
- Visual - Descripteurs vidéos
- Multimedia DS - Structuration
- Reference Software - Logiciel de référence pour validation XM (eXperimentation Model)
- Conformance – guides et procédures pour tester la conformance des implantations de MPEG-7

MPEG-7 – pour finir

- Tout est synthétisé dans la norme
 - haut-niveau, bas niveau
 - un bon état de l'art
- Pas d'applications industrielles
 - prototypes de labo
 - ex. IBM alphaworks
 - norme jeune, mais...
- Utilisation ?
- Viabilité ?
 - Remarque : A4SM choisit de ne pas utiliser MPEG7
 - orientation hiérarchique / restrictive
 - distinction entre description complète ou partielle lourde
 - trop de schémas, trop liés entre eux

73

Autres standards

- EBU/P/META (Metadata exchange standards)
 - production essentiellement
 - association au flux
- Après MPEG7...
 - MPEG21!
 - watermarking
 - gestion de droits
 - ...

74

Plan

- Audiovisuel
- Audiovisuel numérique
- Systèmes d'information audiovisuelle
- Normalisations
- **Problématiques actuelles**
- Systèmes locaux
- Conclusions

75

Le fossé sémantique

- D'un côté le traitement de masse
 - extraction automatique et robuste
 - de descripteurs de bas-niveau
 - (actuellement) peu utiles pour l'indexation
- De l'autre les descriptions symboliques
 - mise en place à la main
 - de descripteurs structurels et conceptuels
 - extrêmement utiles
- La largeur du fossé varie
 - ainsi que la difficulté, voire la possibilité de construire un pont

76

Passerelles ?

- ASSAVID
 - Automatic Segmentation and Semantic Annotation of Sports Videos
 - domaine particulier
 - association de primitive (image, audio) à des termes d'un thésaurus
 - Résultats ?
- [Wenyin, Sun, Zhang, 2000]
 - importer une image, chercher les images similaires, utiliser leurs mots-clés comme descripteurs, confirmés ou infirmés par RF
- [Doom, Vries, 2000]
 - un thésaurus visuel de morceaux d'images comme des concepts, un thésaurus textuel, des clusters et des chaînes de Markov
- ...

77

Passerelles ???

- Tout dépend de images / vidéo et des besoins de l'applications
 - faisable sur domaine restreint (contrôlé)
 - illusoire sur domaine large
- Intérêt AV / MPEG7
 - carrefour des indexations
 - tout mettre dans le modèle
 - voir ce qui en sortira
 - mixer différentes méthodes automatiques (?)
 - paramétrer des algorithmes par des connaissances de haut-niveau
 - aider l'indexation manuelle
- Evidemment, restent les textes
 - extraction à partir des scripts de ce qui se passe à l'image
 - utilisation de textes de courriels avec photos pour annoter celles-ci
 - ...

78

Visualisations audiovisuelles

- Visualiser
 - un document : feuilletage
 - la bonne partie d'un document temporel
 - plusieurs documents en même temps
- Appréhender la globalité
 - délinéariser
 - un accès direct
- Bref
 - céder à la temporalité du flux
 - ou spatialiser cette temporalité
- Outils pour intégrer AV et métadonnées

79

Visualisations

- Fournir des descripteurs qui permettent
 - d'accéder au bon endroit
 - fragments décontextualisés
 - de construire des vues globales
 - résultats de requêtes
 - Synthèses
 - parcours de lecture plus ou moins spatialement temporalisés
 - preview, accès

80

<http://www.ai.sri.com/~oaa/chic/projects/Maestro.html>

Vue spatiale + texte



81

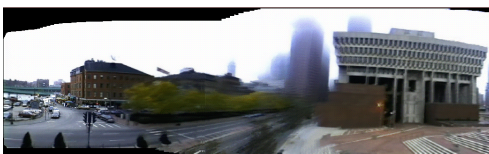
<http://www.fxpal.com/PapersAndAbstracts/papers/bor00.pdf>

Vue « bande dessinée »



82

Salient stills



(Media Lab MIT)



83

http://www.informedia.cs.cmu.edu/documents/ACMMM02_Collage.pdf

Collage (Informedia)



Figure 1. Magnifica, with camera panning and fast navigation for document perceptive effects.

Navigation

Annotations for the video player interface:

- Jump back/next (5 seconds, 10 seconds, Note, Slide Transition)
- Basic Controls
- Pause removal
- Time compression
- Duration
- Elapsed time indicator
- Table of contents
- Personal notes button
- Timeline Markers
- Timeline zoom
- Shot boundary frames

+ études d'usage...

DAVP

- Montage virtuel
 - génération automatique
 - Fiction, news...
 - règles de montage
 - pas trop convaincant pour l'instant
- Montage à la volée
 - en fonction des actions de l'utilisateur
 - ex. Frames

Hypervidéos

- Réelle intégration de la vidéo à l'hypermédia
 - vues centrées texte / flux
- Liens
 - Permanents, spatiaux, temporels, spatio-temporels
- Liens synchronisés, montage
- Etude de la navigation
 - conscience du lien
 - cartes,
 - historique
 - ...
- Vers une première théorisation



[Chambel & Guimarães, 2002]

Et puis ...

- Indexation de documents MM
 - MPEG7
 - pas évident
 - ...
- Annotation
 - et regard
 - et stylo
 - ...

Plan

- Audiovisuel
- Audiovisuel numérique
- Systèmes d'information audiovisuelle
- Normalisations
- Problématiques actuelles
- **Systèmes locaux**
- Conclusions

(E) Strates-IA

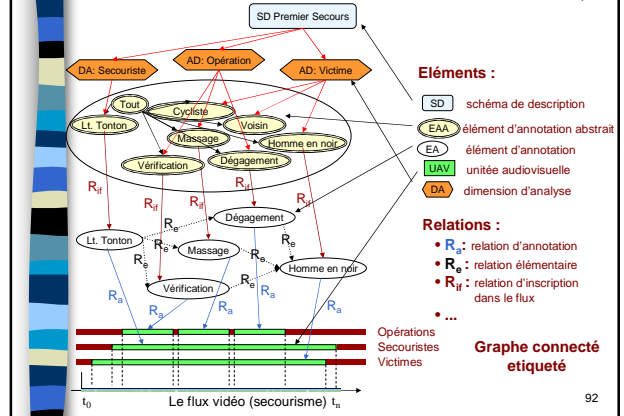
- Historique
 - SESAME 96-99 (RFV, LISI, CNET)
 - RECIS 99-02 (LISI, INRIA, FT-RD)
- Modèle simple
 - des fragments
 - unités audiovisuelles
 - annotés par des descripteurs
 - éléments d'annotation
 - termes : Chirac, Vache, Chocolat, Plan
 - que l'on peut mettre en relation
 - relation élémentaire
- Un graphe de description unique
 - une exploitation contextuelle
 - recherche de chemin dans le graphe
 - inférence contextuelle

(E)SIA Connaissances de description

- Dimensions d'analyse
 - ensemble d'éléments d'annotation abstraits utilisés dans une description
 - HP = {Chirac, Jospin}
 - SA = {Chirac, Vache}
 - Action = {Serrer la main, Marcher, Faire un discours}
- Schéma de description
 - graphe de DA
 - façon d'annoter
 - HP → Action
- Ontologies locales
 - inférences contextuelles
 - plus ou moins significatives

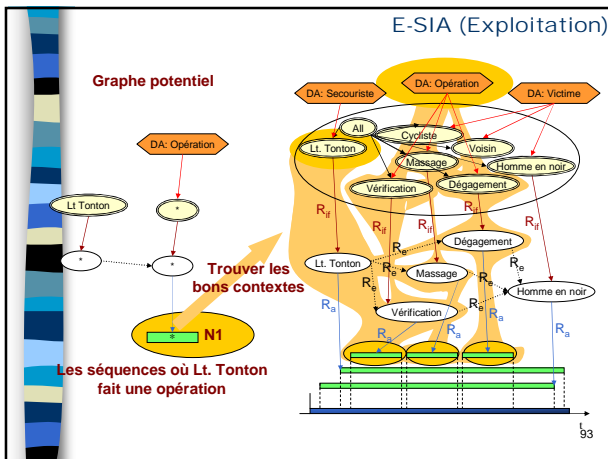
91

E-SIA Exemple



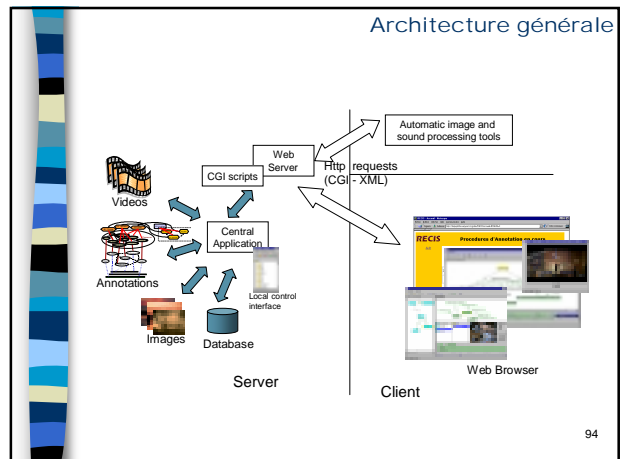
92

E-SIA (Exploitation)



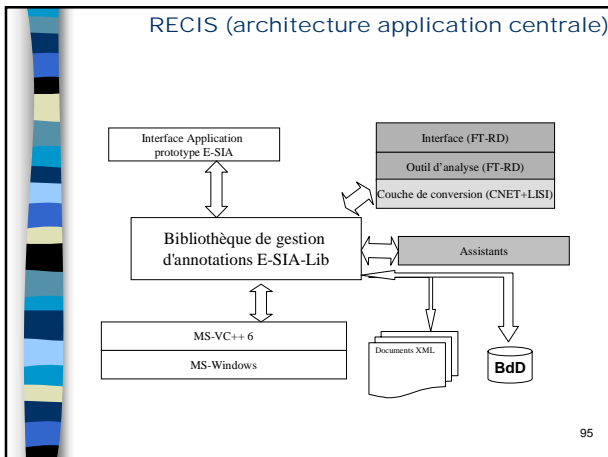
93

Architecture générale



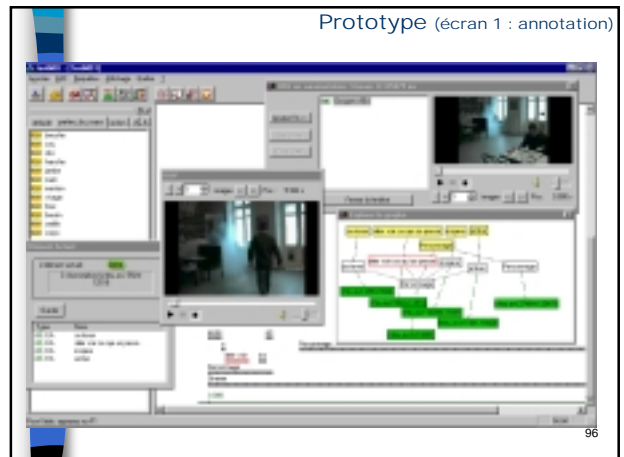
94

RECIS (architecture application centrale)

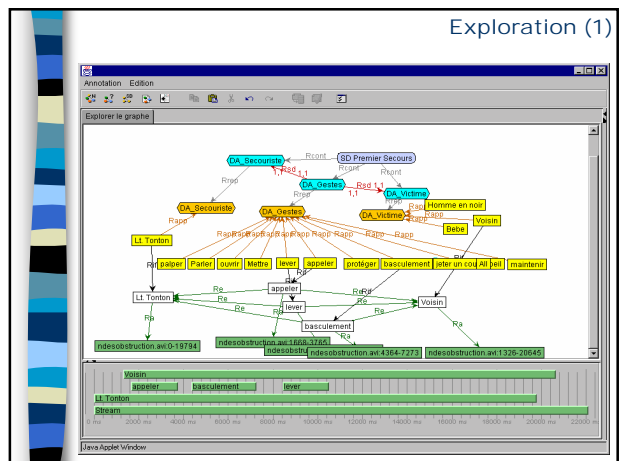
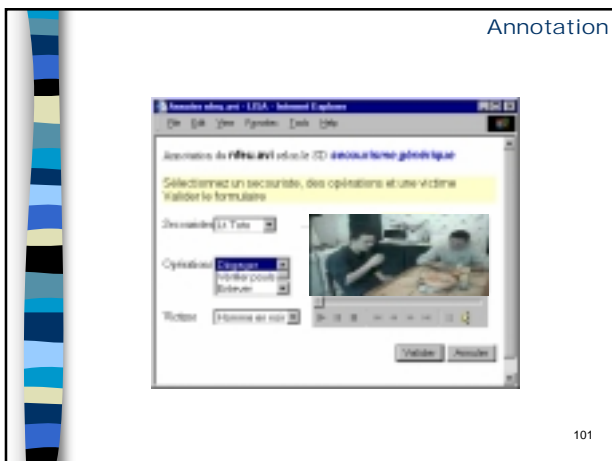
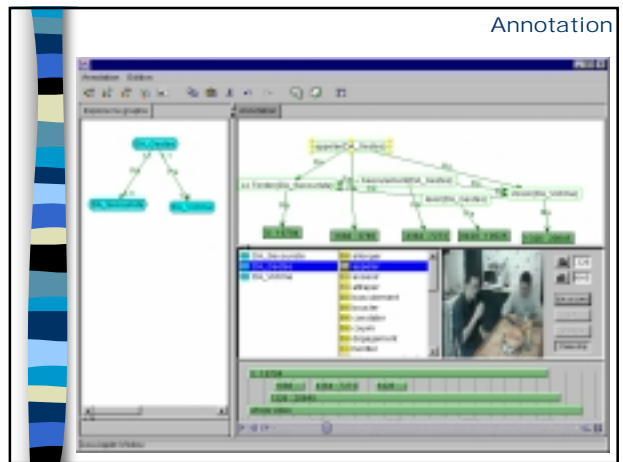
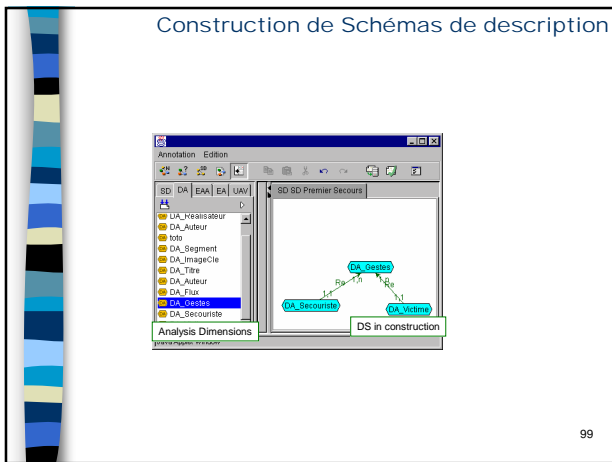
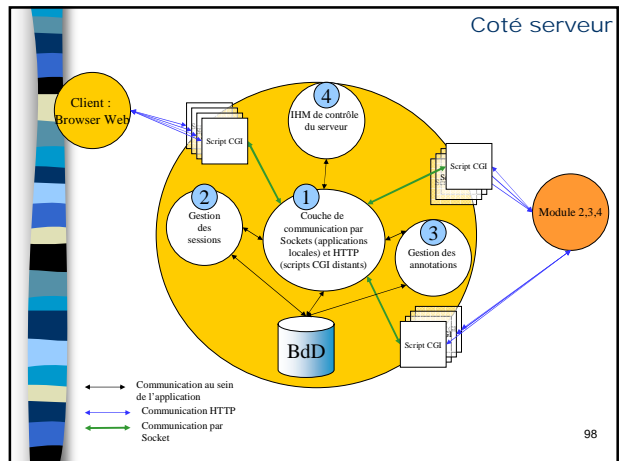
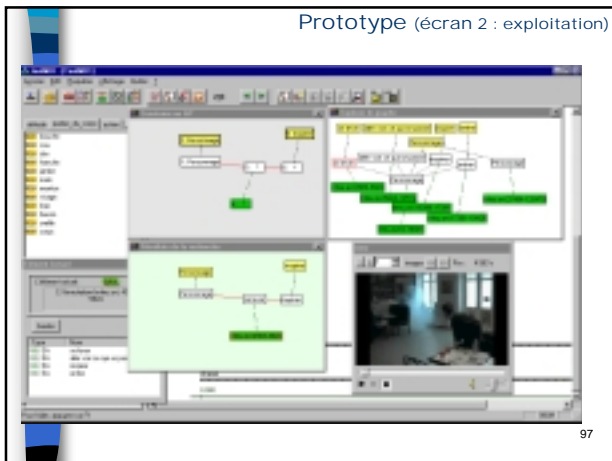


95

Prototype (écran 1 : annotation)



96



Exploration (2)

103

Projet : ADVENE

- Voir les transparents qui y sont consacrés.

104

Plan

- Audiovisuel
- Audiovisuel numérique
- Systèmes d'information audiovisuelle
- Normalisations
- Problématiques actuelles
- Systèmes locaux
- **Conclusions**

105

Conclusion

- Audiovisuel numérique
 - pas de « killer-app »
 - pas vraiment d'usages attestés
 - beaucoup reste à inventer
- Travaux en image
 - pas concluants pour l'indexation automatique généraliste
- Point de rencontre de multiples indexations
 - liens entre l'automatique et le manuel pour l'indexation
 - interdisciplinarité
 - rôle de MPEG-7 dans l'émergence de nouveauté ?

106

Conclusion

- Inspiration textuelle pour la modélisation et l'indexation
 - fertile jusqu'à un certain point
 - DAV != document textuel
 - perception / temporalité
 - format d'enregistrement / forme d'appropriation
 - spatialisatation / indexation
 - Graphes nécessaire (vs arbres)
- En retour
 - voir les documents textuels avec l'œil de l'AV ?

107

Conclusion

- Document = point nodal de l'activité
 - diverses interventions
 - toutes visent à inscrire une pratique du document dans son usage/contexte
 - des structures partout
- L'indexation comme action, écriture, pratique
 - sémiotisation
 - signification
 - sujet de l'indexation
- L'utilisation comme action, écriture, pratique
 - Cf. lecture active
 - nouveaux formats, nouveaux genres ?

108

Remerciements - emprunts

- Elöd Egyed-Zsigmond (INSA-Lyon)
- Claude Seyrat (ExpWay)
- Bruno Bachimont (INA)
- ...

109

Image satellite (SPOT)



10 meter natural color image of Harrisburg, Pennsylvania -- overview of mixed landcover and transition between urban, mountainous, and rural agricultural areas.

110

Effet Koulechov (1)



111

Effet Koulechov (2)



112

Effet Koulechov (3)



[Pas convaincu ?]

113

Compression temporelle

- Dégradation ou non
- Deux approches
 - codage par différence entre des images successives
 - pas forcément très efficace
 - codage par prédiction de mouvement
 - dégradation

114

Exemple MPEG

- I (Intra coded) frames**
 - compression spatiale --> 100%
- P (Predictive coded) frames**
 - compression spatiale et temporelle --> 50% du poids d'une I
 - extrapolation à partir de la précédente image de référence (P ou I)
- B (Bidirectionally predictive coded) frames**
 - compression spatiale et temporelle --> 15%
 - Interpolation à partir des I et P situées avant et après

115

Caractéristiques (exemples)

Histogramme de couleurs

Points d'intérêt⁶

Ikona (INRIA)

<http://www-rocq.inria.fr/cgi-bin/imedia/ikona>

117

FOAF le retour

Notice Inathèque (extrait Web)

- Numéro: 1707832.001
- Titre collection: Loft story
- Titre propre: Loft story : [19ème jour]
- Société de programmes: Métropole Télévision
- Chaîne de diffusion: M6
- Canal: Réseau 6
- Nature de production: Achat de droits commande
- Statut de diffusion: Première diffusion
- Extension géographique: National
- Date de diffusion: 15.05.2001
- Jour: mardi
- Heure de diffusion: 18:19:04
- Heure de fin de diffusion: 18:54:38
- Durée: 00:35:34
- Type de description: Emission simple
- Genre: Real TV;Jeu
- Médiamétrie: Variétés divertissement, jeu interactif
- Couleur: Couleur

119

Notice Forum des images (1)

- AUGUSTIN ROI DU KUNG-FU de Anne Fontaine fiction 1999 couleur 1h27min - VDP16770
- Augustin, apprenti comédien, enchaîne les castings ratés et les figurations désastreuses. Mais il se passionne pour le kung-fu, vénère Bruce Lee dont il connaît chaque film et la moindre acrobatie. Rêvant d'exotisme, il s'installe dans le quartier chinois de Paris (13e). Une comédie pleine d'humour, portée par l'interprétation de Jean-Christien Sibertin-Blanc, dont la gaucherie lunaire évoque celle de Buster Keaton.
- AUGUSTIN ROI DU KUNG-FU Principaux personnages : - Le Docteur Ling (Maggie Cheung), jeune chinoise séduisante et médecin acupuncteur débutant
Parmi les lieux de tournage parisiens : - un hôtel chinois modeste (13e), où Augustin emménage. - un dojo du 13e, où les enchaînements de kung-fu d'Augustin provoquent l'hilarité des pratiquants asiatiques. - un club de karaoké (13e) où se déroule la fête de l'association "Boudha vivant".
- titre, AUGUSTIN ROI DU KUNG-FU, Anne FONTAINE réalisation, Anne FONTAINE , scénario, Anne FONTAINE, Jacques FIESCHI,
- directeur de production, Sylvie BARTHET, montage, Christophe POLLOCK, son, Jean-Claude LOREU, décors, Katia WISCOP
- coproduction, Les Films Alain Sarde, Cine B, France2, 1999
- interprétation, Jean-Christien SIBERTIN-BLANC, Maggie CHEUNG, Darry COWL, Bernard CAMPAN
- film 35mm sonore 1h27min : couleur

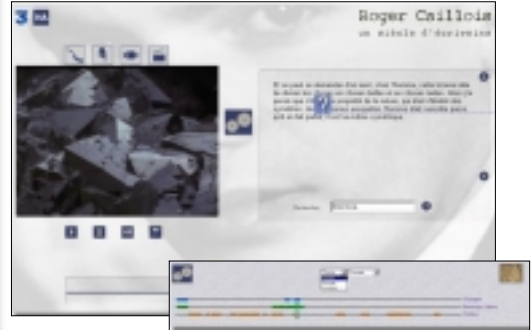
120

Notice Forum des images (1)

- LA GUERRE DU LOUVRE de Jean-Claude Bringuiet documentaire 2000 couleur 55min VDP17951 **Pendant l'Occupation, les œuvres d'art du musée du Louvre sont réparties dans différents châteaux hors de Paris, réquisitionnés comme centres de dépôt. Composé d'archives et de témoignages, notamment d'anciens conservateurs ou de gardiens, ce documentaire télévisé relate les étapes du déménagement des collections et leur périple à travers la France, jusque dans le Quercy, pour échapper à l'occupant allemand, puis leur retour à Paris à la fin de la guerre.** LA GUERRE DU LOUVRE Personnalités interrogées : - Christiane Desroches-Noblecourt, conservateur général honoraire du département égyptien du Louvre; - Frédérique Hébrard, écrivain; - Jean Leymarie, ancien directeur de l'Académie de France à Rome. titre, LA GUERRE DU LOUVRE
- réalisation, Jean-Claude BRINGUIER / image, Pascal SAULETEL / son, François DEVIN / montage, Alain ROBICHE
- coproduction, France3, CDP, Direction des Musées de France, Musée du Louvre, 2000
- vidéo Béta SP sonore 55min : couleur

121

Roger Caillois interactif



Requêtes spatio-temporelles

Visual Query The main paradigm that VideoQ operates is the visual one. Under the visual paradigm, the user can formulate a graphical query, sketching objects and adding features like color, texture and motion.

Text Search Each video clip has been annotated with 10-20 keywords. The text search engine is an effective way to search for a particular subject (sports), location (Paris), or object (lion).

Integrated Query The visual query is especially effective when combined with text search. The text search acts as a pre-filter that limits the number of initial videos to search. The visual query can further refine the search, locating specific details within a video shot.

SVT Semantic Visual Templates represent a small set of visual queries that can effectively retrieve a category of videos. SVTs have been generated from a two-way interaction between the user and system.

Video Navigator A video taxonomy has been developed, and the video clips have been organized within this taxonomy. If you would like a quick overview of all the subject categories that are available, visit the Video Navigator.

VideoQ

123

<http://www.ctr.columbia.edu/VideoQ/>

Souliers de satin - INA



124

Shakespeare Video Annotation Project



5

Hamlet on the Remparts



126

Les schémas de A4SM

- **Newscastr**
 - high level organisation scheme of a new cast, containing references to all related news clips and moderations
- **Newsclip**
 - high level organisation scheme of a new clip, containing all references such as links to relevant annotations and relations to other clips
- **Link**
 - link structure describing the connection between description scheme and the av-material to be described (data)
- **TSR**
 - relative (to a given link) temporal or spatial reference to the data
- **Relation**
 - structure describing the relation between descriptions
- **Formalides**
 - formal information about the news clip, such as broadcaster, origin, language, etc.
- **Bpinfo**
 - production and broadcasting information: when was the clip broadcasted (produced), on which channel, etc.
- **Subjective**
 - description of an event, such as comments of the audience
- **Mediadevice**
 - media specific technical information of the data, e.g. lens state, camera movement, etc.
- **Person**
 - persons participating in the production of the clip, such as reporter, cameraman, technicians, producer
- **Event**
 - the event covered by the description
- **Object**
 - object, existing or acting in the event
- **Character**
 - the relevant character
- **Action**
 - action of an object or character
- **Dialogue**
 - spoken dialogues and comments of the event
- **Setting**
 - setting information of an event, such as country, city, place etc.
- **Archive**
 - archiving value of the news clip according its content and compositing
- **Access**
 - access right info, IPR, rights management of the clip

127

Standards MPEG

- **MPEG-1** 11/1992
 - Storage and coding of moving picture and audio
- **MPEG-2** 11/1994
 - Digital Television
- **MPEG-4** (v2) 11/1999
 - Coding of natural and synthetic media objects for multimedia applications
- **MPEG-7** 08/2001
 - Multimedia Content description for AV material

128

Calendrier MPEG-7

- CFP Octobre 1998
- Evaluation Février 1999
- First WD Décembre 1999
- Committee Draft Octobre 2000
- Final Committee Draft Février 2001
- Draft International Std. Juillet 2001
- International Standard Septembre 2001

129

MPEG-7 et les autres standards MPEG

- **MPEG-7 et MPEG-2**
 - description insérée dans le flux AV
 - synchronisée avec le flux
- **MPEG-7 et MPEG-4**
 - une description pour chaque objet
 - une description pour chaque scène
 - synchronisée

130

MPEG-7 - Part 1 Systems

- **Codage et transport**
 - architecture d'un terminal
 - pré-requis pour la couche transport
 - syntaxe et sémantique des access units
 - format textuel
 - format binaire
 - codage binaire des chemins
 - codage binaire des sous-arbres

131

MPEG-7 - Part 2 Description Definition Language

- **MPEG-7 a adopté XML Schema**
- **Deux extensions simples :**
 - gestion des matrices à plusieurs dimensions
 - spécification des dimensions dans le schéma
 - spécification des dimensions dynamiquement dans la description (à la 'xsi:')
 - SimpleTypes pour la gestion du temps

132

MPEG-7 - Part 3 Audio Descriptors

- Outils généraux
 - "Scale Tree"
 - segment de silence uniforme !
- Quelques descripteurs
 - effets sonores
 - indexation et catégorisation des effets sonores
 - timbre des instruments de musique
 - caractéristiques perceptuelles
 - contenu parlé
 - des treillis de phonèmes
 - mélodies
 - représentation compacte de mélodies

133

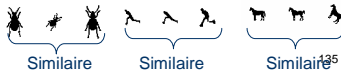
MPEG-7 - Part 4 Video Descriptors

- Outils généraux (localisation)
 - grid layout
 - 2D-3D Multiple View
 - spatial 2D Coordinates
 - ...
- Descripteurs de couleurs
 - color space - RGB, HSV...
 - dominant color(s) - petit nombre de couleurs
 - scalable color - histogramme
 - color layout - distribution spatiale des couleurs
 - ...

134

MPEG-7 - Part 4 Video Descriptors (suite)

- Descripteurs de textures
 - homogeneous texture
 - texture browsing
 - edge histogram - pour la recherche à partir de dessins
- Descripteurs de formes
 - region-based shape
 - contour-based shape
 - 3D shape
 - visages



MPEG-7 - Part 4 Video Descriptors (suite)

- Descripteurs de mouvements

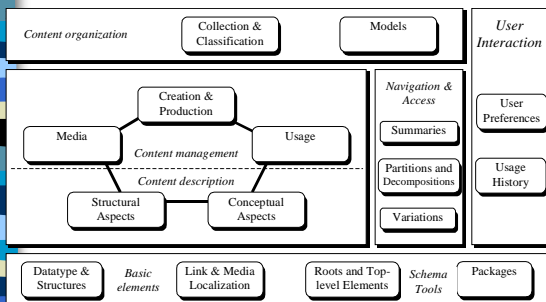
- camera motion



- motion trajectory - localisation précise d'un objet dans le temps et l'espace
- parametric motion - mouvement d'un objet
- motion activity
- quantité de mouvement

136

MPEG-7 - Part 5 Multimedia Description Schemes



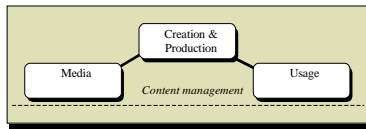
137

Eléments de base

- Constructions fondamentales
 - basic datatypes
 - vecteurs, matrices
 - lien et localisation de segments AV - timecode
- Time, place, person, individuals, ...
- Textual annotation : free, structured, ...

138

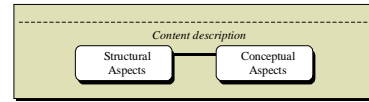
Gestion des contenus



- **Creation information**
 - Title, creators, creation location, classification (genre)
- **Usage information**
 - Rights, availability, financial
- **Media information**
 - Coding format, coding parameters, storage format...

139

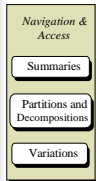
Description du contenu



- **Information structurelle**
 - Video segments, regions, table of content, ...
- **Description conceptuelle**
 - Events, objects, relationships,...

140

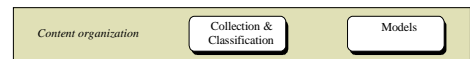
Navigation et accès



- **Résumé**
 - Browsing, navigation, visualization, ...
 - Hiérarchique ou séquentiel,
- **Partitions / décompositions**
 - Différentes vues du même document,
- **Variations**
 - Différentes versions,
 - Coding format, lossy compression.

141

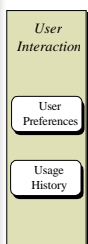
Organisation des contenus



- **Collections**
 - Gestion de collections de documents
 - Ensembles non ordonnés de contenu AV
- **Modèles**
 - Caractériser des collections ou classes de contenu AV
 - Probability, Cluster, Analytic

142

Interaction avec l'utilisateur



- **User Preferences**
 - Description des préférences utilisateurs
- **Usage History**
 - Historique des préférences

143

Effet Koulechov (1)



144

Effet Koulechov (2)



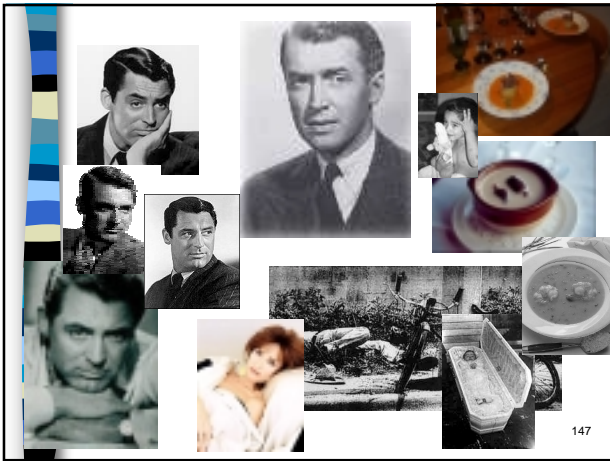
145

Effet Koulechov (3)



[[Retour](#)]

146



147