

Cours : Indexation d'informations MultiMedia

Chapitre 1: Introduction aux bases de données MultiMedia

présenté par:
Mohamed Sahbi Bahroun

Plan

- 1) Introduction : motivations,
- 2) Gestion des bases de données MM,
- 3) Recherche d'informations MultiMedia par le contenu
- 4) Types de bases d'images et description
- 5) Fonctionnalités d'un système de recherche d'informations MultiMedia par le contenu
- 6) Paradigmes de recherche

1. Introduction

Multimedia : définition

- ❖ Des contenus pouvant être de différents types, en particulier
 - ✓ Texte
 - ✓ Son
 - ✓ Image
 - ✓ Vidéos
- ❖ Qui peuvent être combinés
- ❖ Caractérisés par des besoins de stockage très importants

Contexte

❖ Croissance explosive de la production de contenus numériques

- ✓ Grand public : démocratisation de l'imagerie numérique
- ✓ Industrie audiovisuelle : mondialisation des programmes
- ✓ Spécialisés : observation de la terre, vidéosurveillance, imagerie médicale

❖ Stockage numérique

- ✓ Forte diminution du coût par gigaoctet
- ✓ Croissance de la vitesse d'accès

❖ Transmission / diffusion numérique

- ✓ Augmentation des débits
- ✓ Forte diminution du coût → démocratisation diffusion / partage

→ Très grands volumes de données multimédia à exploiter

Besoin des utilisateurs

- ❖ Avoir un aperçu du contenu de la base d'images
 - ✓ Partitionnement de la base > résumé visuel de la base
- ❖ Naviguer dans la base d'images/vidéos
 - ✓ « Chemins » pré-établis pour aider la navigation
 - ✓ Résumés suivant des critères complémentaires
- ❖ Détecter et reconnaître des composantes (visages, véhicules, piétons, monuments...) et des situations
- ❖ Retrouver des images suivant différents critères
 - ✓ Recherche d'une image spécifique ou d'un ensemble d'images
 - ✓ Recherche d'une partie d'image ou d'un objet dans les images

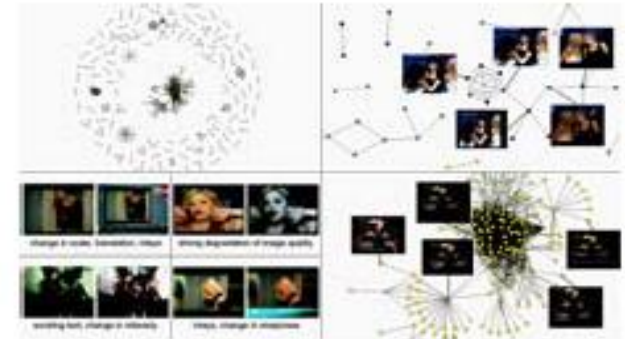
Besoin des utilisateurs

- ❖ La **recherche d'information** dans les documents multimédias est un problème crucial ...
- ❖ Elle a dopé les recherches dans le domaine de l'**indexation** et de la **recherche par le contenu** et en particulier celui de l'apparence visuelle

Domaines d'application

■ Audiovisuel

- ◆ Exemples : détection de copies (droits), retrouver un plan spécifique d'un programme, annotation automatique de vidéos

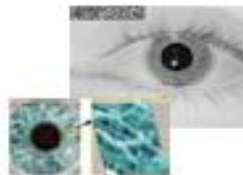
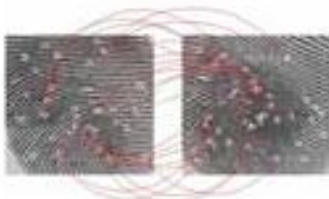


■ Sécurité

- ◆ Exemples : biométrie (empreintes, iris), vidéosurveillance



© M. Slaughter



Domaines d'application

■ Internet

- ◆ Exemples : commerce électronique, annotation et filtrage Web2.0



■ Design, publicité, architecture

- ◆ Exemples : rechercher texture spécifique pour l'industrie textile, illustrer une publicité par une photo adéquate, détecter des tendances



Titre : Trickshot_Lambert
Tags : le papillon
Chaine (categorie) : sport
Durée : 00:18
Pris le : 25 février 2007
Lieu : France



Titre : Produce3
Tags : BILARD POOL CLIP
VIDEO POUR LE PLAISIR
Durée : 02:23
Pris le : 03 novembre 2006
Lieu : France

■ Art, éducation

- ◆ Exemples : recherche encyclopédique d'illustrations, d'une œuvre d'art spécifique

Applications spécifiques

- Bases d'images satellitaires
 - ◆ Exemples : suivre l'évolution de la végétation, évaluer l'impact des feux de forêt
- Bases d'images botaniques
 - ◆ Exemple : identifier les gènes qui interviennent dans une même chaîne de synthèse protéinique
- Bases d'images médicales
 - ◆ Exemple : quelles images indiquent une pathologie (but éducatif ou diagnostic)



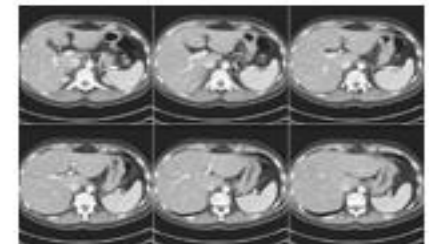
© INRIA et IFN



<http://atlas.nrcan.gc.ca>



© NASC <http://arabidopsis.info>



Chaine MultiMedia : Besoins

- ❖ Génération : outils de production et de création
- ❖ Représentation : utilisation de formats de représentations différentes
- ❖ Traitement des images
- ❖ Stockage
- ❖ Transmission : problème de réseaux, architecture
- ❖ **Recherche d'information : recherche basée sur le contenu**
- ❖ Distribution : conception de serveur de streaming

Applications types

- ❖ Films à la demande
- ❖ Informations médicales
- ❖ Systèmes d'informations géographiques
- ❖ Architecture/Design
- ❖ Géo-localisation
- ❖ Enquêtes policières
- ❖ Militaire
- ❖ Expérimentations scientifiques
- ❖ Enseignement
- ❖ Archivage, gestion des bases de données de contenu (personnelles ou professionnelles)
- ❖ Moteur de recherche (ex: Internet)

2. Gestion de bases de données MultiMedia

Base de données : standards

- ❖ Une base de données (DB) relationnelle est un ensemble de mise en relation.
- ❖ Un système de gestion de base de données (DBMS) est un logiciel facilitant la création, la maintenance et la manipulation d'une base de données. Il garantit sécurité et intégrité de la base.
 - ✓ Oracle, DB2, Sybase, PostgreSQL, Mysql
- ❖ Système de base de données = la base + le système de gestion
- ❖ Points clés :
 - ✓ Indépendance des données : données et logiciels de gestion séparés
 - ✓ Indépendant du mode de stockage ou de représentation de la base
 - ✓ Abstraction des données : représentation conceptuelle
 - ✓ Auto-suffisance de la base de données: Pas de relation avec des données externes.

Spécificités des bases de données multimédia (MMDB)

- ❖ Ex de requête impossible avec un SGDB “classique”:
 - ✓ récupérer toutes les images “qui ressemblent” à une image requête
- ❖ Grosky et Fotouhi (1) caractérisent le multimédia à partir du type d'activité impliquée dans la création des données

L'information portée par le multimédia est tout ce qui peut venir du monde réel, alors que l'information portée par une base de données classique ne peut être qu'une représentation symbolique de faits limités à l'univers de la base de données.
- ❖ Le développeur d'une MMDB ne peut expliciter tous les aspects des données qui seront importants pour l'utilisateur
 - ✓ les informations ne peuvent se limiter à un ensemble de champs prédéfinis
 - ✓ sauf pour certaines applications très spécifiques

Richesse des bases de données multimédia

- Apprendre un concept (ici les couleurs) à partir "du monde réel" [2]
→ c'est-à-dire Google ou tout autre base texte+images



3. Recherche d'informations MultiMedia par le contenu

Contenu sémantique Vs contenu physique

- Contenu sémantique:
 - Annotations/Ontologies: relations et structuration sémantique
=> « Human-based metadata »
- Contenu Physique:
 - Images fixes/animés, son/parole, ...
=> « Machine-based metadata »
- Dernière décennie => Pertinence de l'indexation et la recherche par le **contenu physique** pour l'accès aux archives multimédias (**Physical Content-based**)

Contenu et Méta-données

- ❖ Les données “brutes” (fichier image, fichier son) provenant d’objets multimédia contiennent des informations sémantiques.
- ❖ Grosky [4] distingue
 - ✓ La donnée brute (image, signal sonore)
 - ✓ Les méta-données décrivant le contenu

DEFINITION: le processus d'abstraction permet de décrire le contenu d'objets multimédia par des méta-données, soit manuellement, soit par des procédés (semi-)automatiques

- ❖ Ces meta-données proviennent soit de la description de l'utilisateur par du texte soit une annotation par un certain nombre de mots clés
- ❖ objets (couleur moyenne d'une image)

Description par le texte

- Par quoi indexer : mots-clés, graphes conceptuels, ...



Mots-clés : tournesol, soleil, Sud...

- Approche la plus ancienne et la plus répandue
- Inconvénients

- ◆ Ambiguïté inhérente
- ◆ Dépendance du contexte
- ◆ Subjectivité de l'indexation
- ◆ Dépendant de la langue
- ◆ Coût annotation manuelle
- ◆ Incomplétude inhérente



Mots-clés : écusson, ???

Description par le texte

Requête textuelle: "*Port du voile*"



Recherche d'information



Description par le texte



"Avocat" ?



Annotation / Méta-données

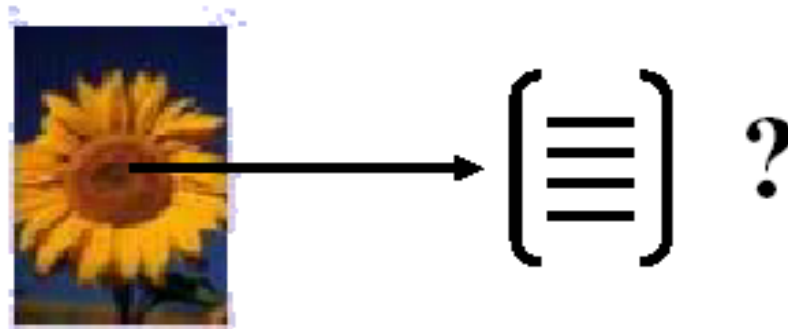
- ❖ Une annotation textuelle sera toujours trop restrictive, même si elle prend en compte à la fois des informations syntaxiques et symboliques.
- ❖ Malgré cela, l'approche la plus utilisée reste l'annotation textuelle et manuelle
- ❖ Avantage : recherche indépendante du type de media
- ❖ Inconvénients
 - ✓ Le coût d'annotation manuel est très important
 - ✓ Différentes personnes utilisent un vocabulaire différent pour signifier la même chose (ex: clair, lumineux). Connu sous le nom de "problème du vocabulaire". L'utilisation d'un thésaurus ne règle pas tout.
 - ✓ Différentes personnes peuvent décrire des aspects différents du média (polysémie du contenu), la même personne décrira différents aspects en fonction de la situation.
 - ✓ Le non-verbal ne peut être exprimé sans ambiguïté.

Annotation / Méta-données (corpus COREL)

- ❖ 10 000 images généralistes fournies par James Z. Wang <http://wang.ist.psu.edu>
- ❖ Chaque image est :
 - Annotée par de 1 à 5 mots-clés choisis manuellement dans un lexique de 250 mots-clés environ
 - Segmentée en 10 régions maximum
 - Chaque région (appelée « blob ») est décrite par un vecteur de 40 composantes visuelles

Indexation par le contenu visuel

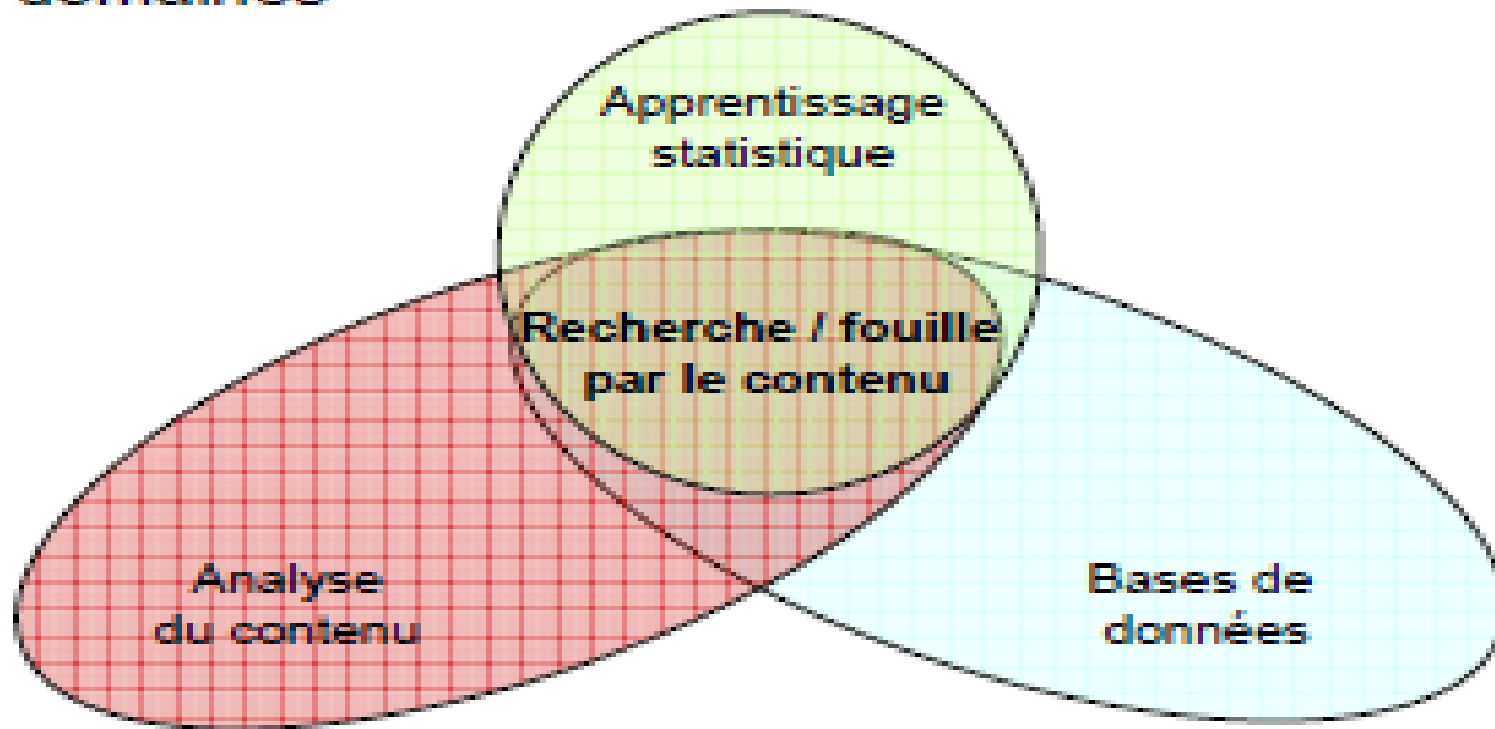
- Extraire automatiquement d'une image des descripteurs **significatifs** et **compacts**, qui seront utilisés pour la recherche ou la structuration



- Prise en compte directe du contenu visuel
... mais manifestation fréquente du **gap sémantique**

Recherche d'images par le contenu visuel

Thématique émergente à l'intersection de 3 grands domaines



4. Indexation

Indexer ?

❖ Indexer = extraire une information synthétique des images
afin de faciliter l'accès à leur contenu

✓ domaines scientifiques concernés : fouille de données (data mining), classification, ingénierie des connaissances, vision artificielle, SGBD, ...

✓ information = élément susceptible d'être codé pour être conservé, traité, communiqué

✓ index = clé d'accès à l'information contenue dans l'image

Exploitation d'une analyse fine du contenu du document visuel

Indexation par le contenu

○ Descripteur d'image

- Méthode d'extraction du contenu visuel de l'image (pertinent pour l'application)
 - Exple : histogramme couleur
- Mesure de similarité associée
 - Exple : intersection d'histogrammes

○ Signature d'image

- Vecteur numérique représentant le contenu visuel de l'image
 - Exple : 1 vecteur de dimension 216 pour l'histogramme couleur

○ Espace de description (de représentation) des images

- 1 image = 1 ou plusieurs points dans un espace multidimensionnel

○ Espace de recherche dans la base d'images

- Structurer l'espace de description pour une recherche efficace (index)

Contraintes

○ Deux contraintes fondamentales

1. L'invariance de la description

- Translation, rotation image
- Changement de point de vue
- Changement d'échelle
- Changement d'illumination (interne, externe)
- Occultations
- Arrière-plans différents
- Caméras différentes (grand angulaire, etc)
- Images de mauvaise qualité (processus d'acquisition non maîtrisé, codage JPEG, etc)



➤ Identifier les contraintes liées à l'application

- ✓ Exple : forts changement d'échelle ? D'illumination ?

➤ Déterminer un ensemble de grandeurs invariantes (au moins robustes) ou une mesure de similarité adaptés

- ✓ Attention : les descripteurs les plus invariants sont les moins discriminants !

Difficultés

- ❖ Transformations image : rotation, échelle, luminosité

rotation image



Difficultés

❖ Transformations image : rotation, échelle, luminosité

changement d'échelle



Difficultés

❖ Transformations image : rotation, échelle, luminosité



Propriétés (2)

○ Deux contraintes fondamentales (suite)

2. La dimension de la description

- ✓ Dimension de la signature d'image / Nombre de signatures par image
- Exple 1 : 500 000 images décrites par un histogramme couleur décrit par un vecteur de dimension 216 => recherche dans un espace de dimension 216 contenant 500 000 points.
- Exple 2 : 500 000 images décrites par 20 régions décrites par un vecteur de dimension 40 => recherche dans un espace de dimension 40 contenant 10 millions ($500\,000 \times 20$) de points.

➤ Déterminer le plus petit ensemble de grandeurs invariantes

- ✓ Problématique du passage à l'échelle (bien connu en bases de données)

5. Types de bases d'images et description

Types de bases d'images

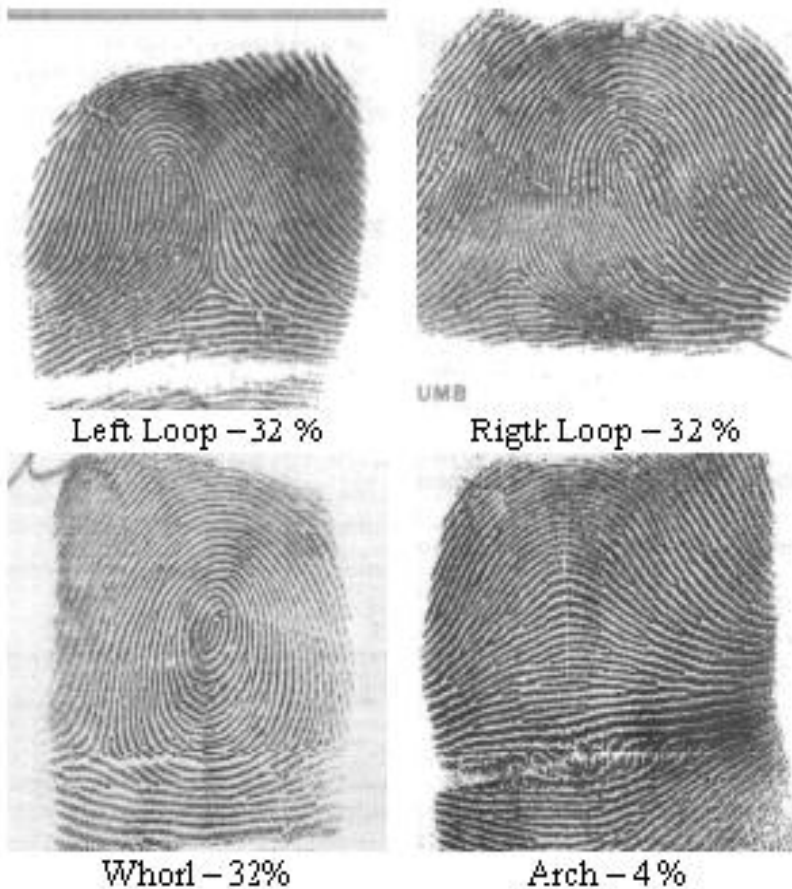
■ Bases **génériques**

- ◆ Contenu hétérogène (bases grand public, Internet, archives généralistes)
- ◆ En général sans vérité-terrain naturelle

■ Bases **spécifiques**

- ◆ Spécifiques à un domaine d'application
- ◆ Contenu homogène (visages, empreintes digitales, monnaies...)
- ◆ En général avec vérité-terrain naturelle

Bases spécifiques : exemples



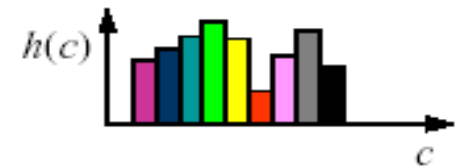
Description du contenu MM

- Séparation et description de caractéristiques
 - ◆ Image : couleurs, textures, formes
 - ◆ Vidéo : couleurs, textures, formes, mouvements
 - ◆ Audio : temporelles, spectrales
 - Identification de « composantes »
 - ◆ Image : régions homogènes, points d'intérêt, configurations ; objets, visages, etc.
 - ◆ Vidéo : *shots* ; scènes, *key-frames*, régions à mouvement cohérent, objets, personnes, etc.
 - ◆ Audio : séparation/segmentation voix/musique/bruit, événements sonores (ex. but dans un match, explosion, etc.)
- Manifestation du **gap sémantique** (signal ↔ signification) !

Description du contenu

■ Séparation et description de caractéristiques

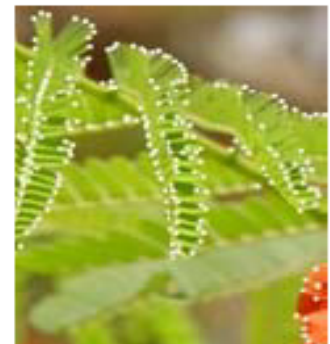
- ◆ Image : couleurs, textures, formes
- ◆ Vidéo : couleurs, textures, formes, mouvement



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

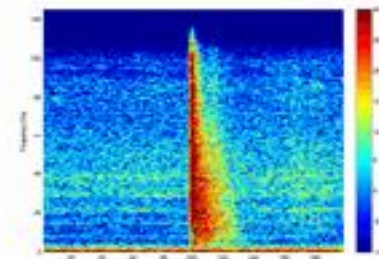
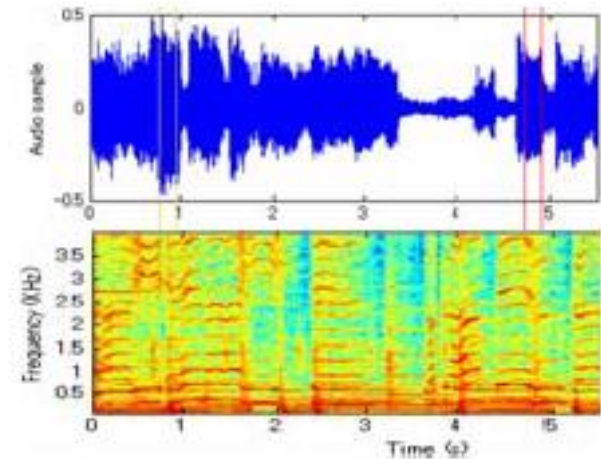
■ Identification de « composantes »

- ◆ Image : régions homogènes, points d'intérêt, configurations ; objets, visages, ...
- ◆ Vidéo : *shots*, régions à mouvement cohérent; scènes, personnes, objets, ...



Description du contenu (2)

- Séparation et description de caractéristiques
 - ◆ Audio : temporelles, spectrales
 - ◆ Description résumée ou séquence de descriptions
- Identification de « composantes »
 - ◆ Audio : séparation/segmentation voix/musique/bruit, événements sonores (exemples : but, explosion)



Description des images: typologie

■ Descripteurs **génériques**

1. Description **globale** (approximative) de l'image
 2. Description **locale** (potentiellement plus précise)
 - Description de régions homogènes
 - Description autour de points d'intérêt
- ◆ Caractéristiques décrites : couleurs, textures, formes...

■ Descripteurs **spécifiques**

- Empreintes digitales : minuties
- Visages : EigenFaces, DSW...

Typologie des descripteurs d'images

○ 3 classes de descripteurs

● Description **globale** de l'image

- Description **approximative** de toute l'image (recherche d'une **ambiance**)
- Solutions :
 - Couleur, Texture , Forme
- 1 point dans un espace de description de grande dimension

● Description **locale** de l'image

- Description **précise** des parties de l'image (recherche **précise**)
- Solutions :
 - Modèles pour la reconnaissance d'objets
 - Génériques (sans modèle) : segmentation en régions, détection de points d'intérêt
- Plusieurs points dans un espace de description de dimension modérée

● Descripteurs **spécifiques** (essentiellement biométrie)

- Empreintes digitales : minuties
- Visages : EigenFaces, ...

Description globale de l'image



Description
des couleurs



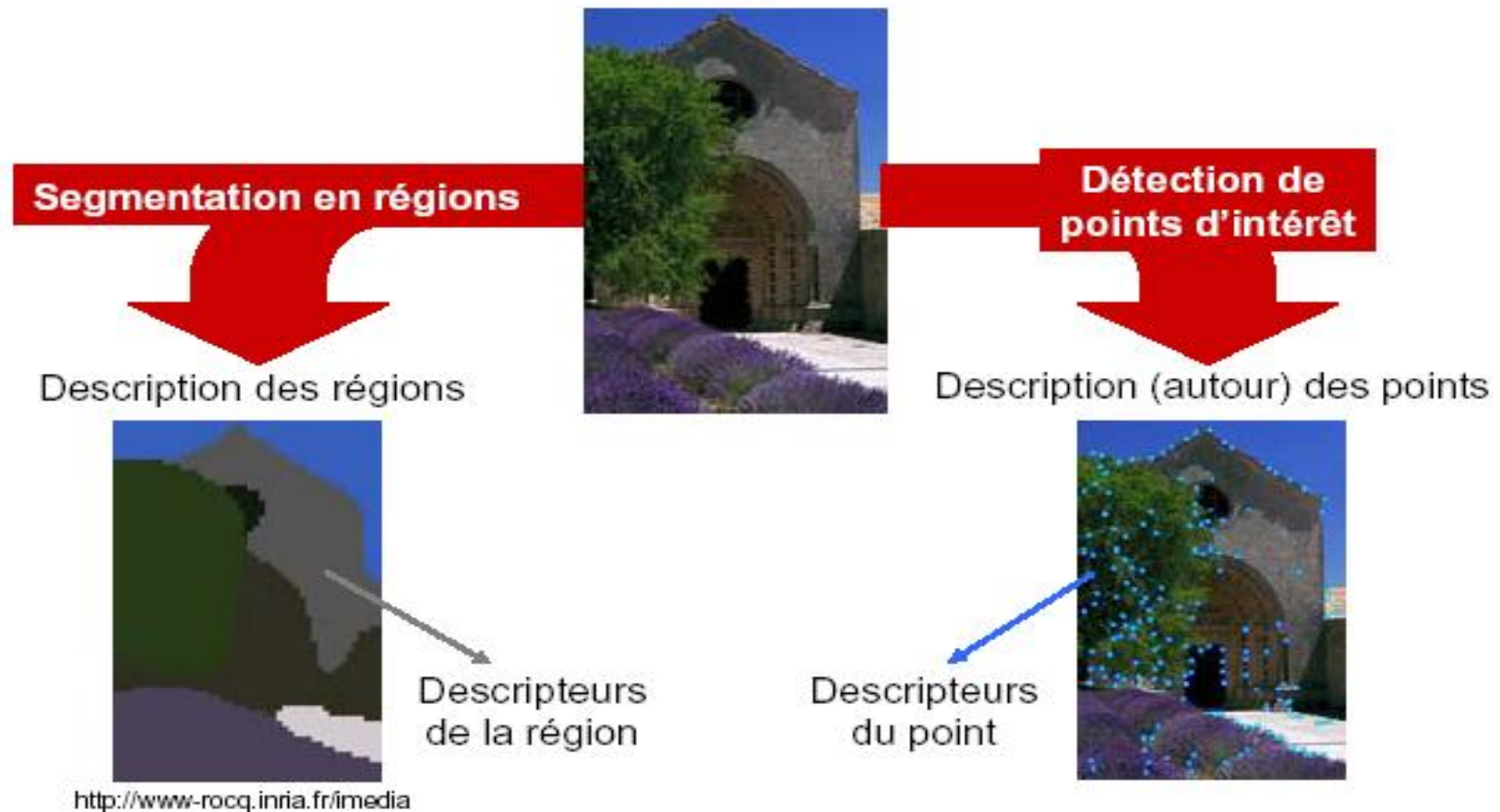
Description
des textures



Description
des formes

Assembler les différents types de description pour
décrire de façon plus complète chaque image

Description locale de l'image



5. Fonctionnalités d'un système de recherche d'informations MultiMedia par le contenu

Recherche approximative : par similarité

- ❖ L'idée clé dans la recherche de média est la recherche approximative
- ❖ Utilise la notion de proximité, de similarité, de distance entre objets
- ❖ Généralement, la requête est exprimée sous la forme d'un ou de plusieurs vecteurs dans un espace multi-dimensionnel.
 - ✓ définition d'une distance (ou mesure de dissimilarité) sur cet espace
 - ✓ recherche des objets dont la distance est minimale
- ❖ Les vecteurs d'attributs sont extraits de contenu de l'image

Recherche par le contenu (CBIR: content based information retrieval)

- ❖ Possibilité d'utiliser un retour de pertinence

Recherche par l'exemple

- Objectif : retrouver les images similaires à une image exemple
- Paradigme le plus répandu et le mieux traité actuellement, de nombreuses applications y ont recours
- Méthode
 1. Choix d'un exemple (requête) par l'utilisateur (image, partie d'image)
 2. Comparaison du descripteur de la requête avec ceux de la base
 - Calcul de distances entre la requête et la base
 - Parfois plusieurs descripteurs exemples (cas des descripteurs locaux ou des premières méthodes de contrôle de pertinence)
 3. Recherche des images les plus proches
 - Retourner l'image la plus proche
 - Retourner les k images les plus proches
 - Retourner les images situées à une distance inférieure à une distance donnée



Fonctionnalité d'un système de recherche

- ❖ Ce qu'il doit (*devrait*) *permettre*
 - ✓ Les objets multimédia peuvent être actifs
 - ✓ Le processus de requête est interactif
 - ✓ Le processus de requête peut utiliser plusieurs modes de représentation (ex: texte, puis image)
- ❖ Il doit posséder deux modes
 - ✓ Gestion du contenu (ajout, modification des objets présents, suppression)
 - ✓ Interrogation de la base
- ❖ Il doit pouvoir combiner des informations provenant de différents modes de représentation, utiliser le retour de pertinence pour affiner la recherche
- ❖ Nécessité d'une interface utilisateur permettant la manipulation des objets existants, la présentation de résultats, le retour de pertinence

Recherche globale par l'exemple

Requête (ici, l'image entière)

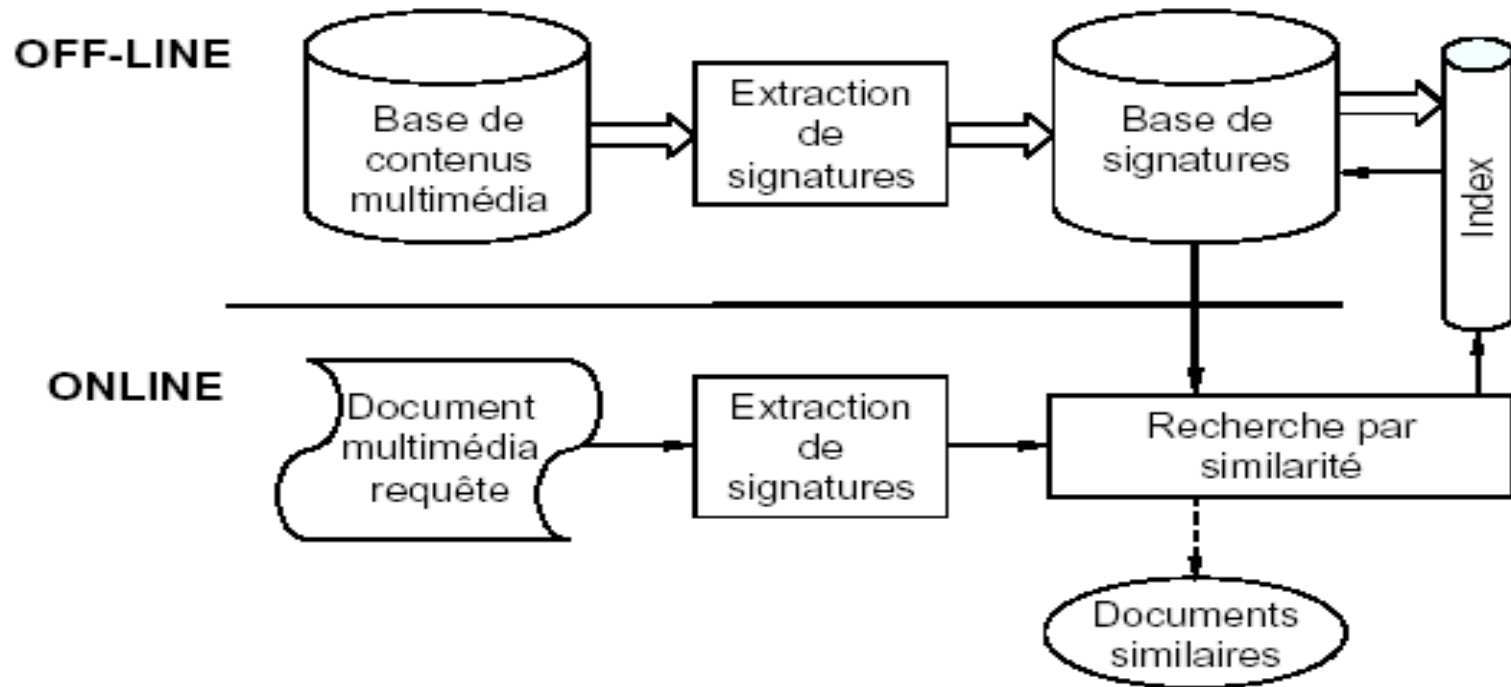
Réponses classées par ordre de similarité visuelle décroissante

Recherche locale par l'exemple



Architecture générale d'un SRM

- SRM = Système de Recherche Multimédia
 - ◆ Architecture client-serveur



6. Paradigmes de recherche

Recherche par le contenu : paradigmes

- Initialisation de la recherche (problème de la page zéro)
 - ◆ Requête externe (image fournie par l'utilisateur)
 - ◆ Résumés visuels
- Recherche par l'exemple
 - ◆ Globale, approximative
 - ◆ Locale, précise (requêtes partielles)
- Recherche itérative avec contrôle de pertinence
 - ◆ Raffinement itératif exploitant les réponses de l'utilisateur
- Recherche par image mentale
 - ◆ Recherche par composition à partir d'un thésaurus visuel
 - ◆ Recherche itérative d'une image cible

Initialisation aléatoire

■ Présentation d'un échantillon aléatoire d'images

- ◆ Problème : les images montrées ne sont pas représentatives de la base

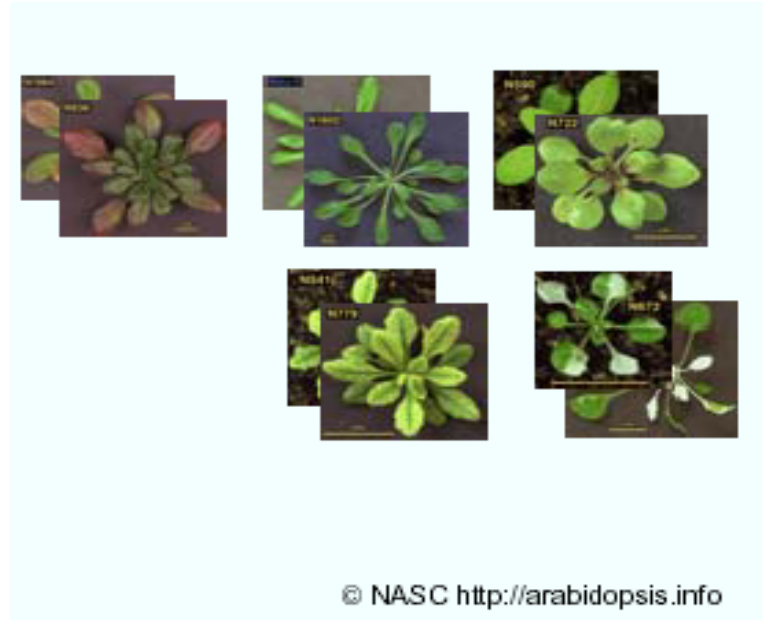
Exemple : base Columbia
(1440 images, 20 objets
sous 72 points de vue)



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

Résumé visuel

1. Partitionner la base en classes
 - ◆ Nombre de classes en général inconnu
 2. Présenter à l'utilisateur un bon représentant pour chaque classe
 - ◆ Fournit un résumé de la base
 - ◆ Facilite la recherche d'une bonne image de départ
- Partitionnement et choix de représentants : traitements hors ligne (*off-line*)



© NASC <http://arabidopsis.info>

<http://www-rocq.inria.fr/media>

Contrôle de pertinence

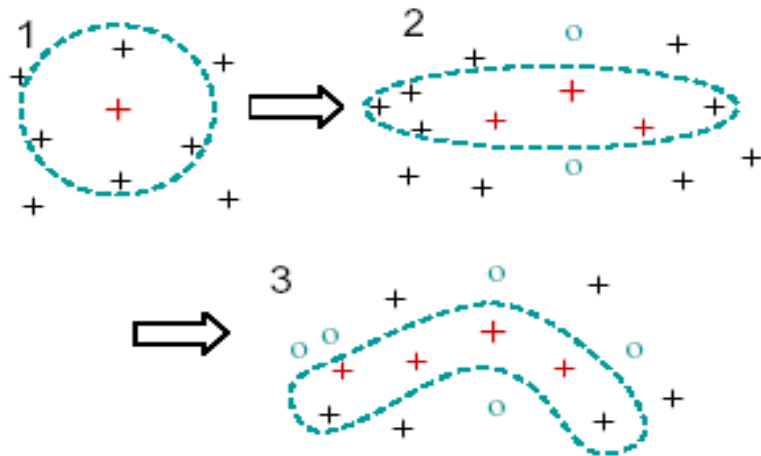
- Objectif : retrouver un ensemble d'images pertinentes par rapport à la recherche en cours (caractérisées en général par des relations de similarité plus complexes)
- Principe : raffinement itératif exploitant les réponses de l'utilisateur (qu'on peut voir comme une recherche avec reformulation de requête)
- Étapes lors de chaque itération :
 1. Le système montre des images candidates à l'utilisateur
 2. L'utilisateur marque certaines des images comme étant des exemples positifs (« ce que je cherche ressemble à ça ») ou négatifs (« ce que je cherche ne correspond pas à ça »)
 3. Le système affine son estimation de l'ensemble d'intérêt

Contrôle de pertinence : exemple

Objectif de la recherche :
retrouver des portraits

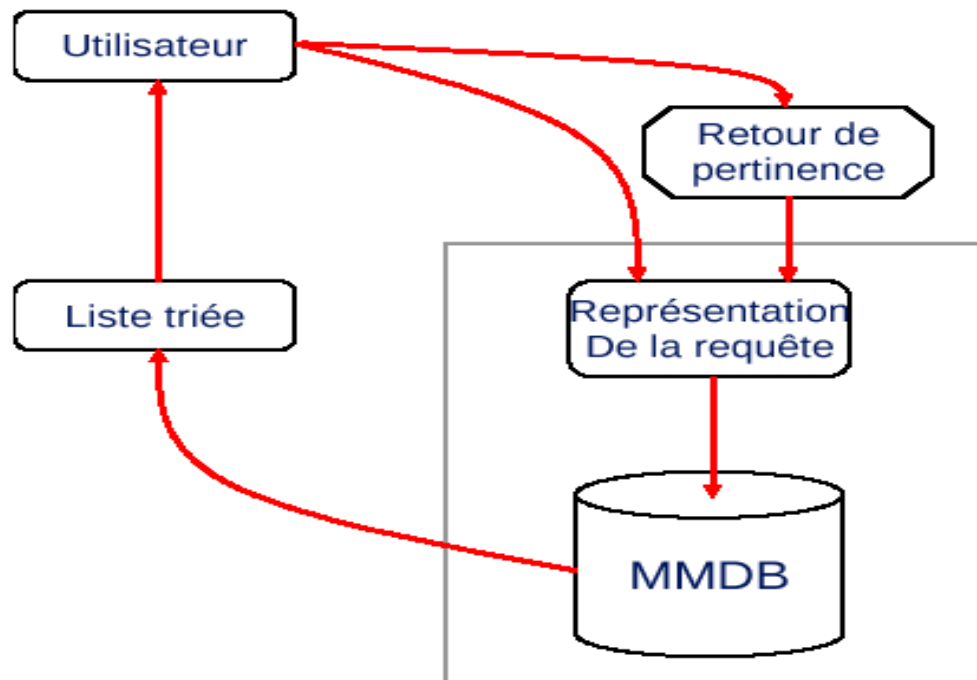
Base de 7500 images, dont
110 portraits

Disponible : description
globale (couleur, texture,
forme) → 120 dimensions



Retour de pertinence

- Possibilité de mettre l'interaction au coeur du système
 - parmi les réponses, le système identifie ce qu'il trouve intéressant
 - puis affinage de la requête par l'utilisateur



Composition de requête par thesaurus

- Principe : construction d'un représentant de l'image mentale (la requête) par composition de prototypes de régions (thesaurus de régions)
- Exemple : recherche d'un paysage urbain



Formulation de la requête



Catégorie « régions à dominante verte »

Composition logique correspondante



Exemple de recherche par thesaurus

Résultat de la requête



Exemples d'images rejetées



Moteurs de recherche basées sur le contenu

■ Systèmes industriels

- ◆ QBIC (IBM, depuis 1995) : <http://www.qbic.almaden.ibm.com>
- ◆ ImageFinder (Attrasoft) : <http://www.attrasoft.com>
- ◆ Virage (Virage Technologies, 1996) : <http://www.virage.com>

■ Systèmes académiques

- ◆ IMEDIA (INRIA Rocquencourt) : <http://www-rocq.inria.fr/imedia/>
- ◆ SAPIR (projet Européen) : <http://milos.isti.cnr.it:8080/milos/album/>
- ◆ Tiltomo : <http://www.tiltomo.com>

Bibliographie

- [1]. W.I. Grosky, F. Fotouhi and Z. Ziang. *Multimedia data Management. Using metadata to integrate and apply digital media*, chapter *Using metadata for the intelligent browsing of structured media objets*, pages 123-148. In Sheth and Klas, 1998
- [2]. J. Van-der-Weijer, C. Schmid and J. Verbeek, *Learning Color Names from Real-world Images*, CVPR 2007
- [3]. **A. Hampapur and R. Jain. *Multimedia data management. Using metadata to integrate and apply digital media*, chaper Video data management systems: metadata and architecture, pages 245-286.,1998.**
- [4]. W.I. Grosky, F. Fotouhi and Z. Jiang. *Multimedia data management. Using metadata to integrate and apply digital media*, pages 123-149, 1998.