

# Bases de données multimédia

Séance 1 : Introduction à la recherche par contenu visuel  
dans les banques d'images

Valérie Gouet-Brunet  
valerie.gouet@cnam.fr

17 Octobre 2007



## Plan de la séance

- Contexte, Motivations
- Domaines d'application
- Définitions
- Descripteurs d'images
- Paradigmes de recherche
- Méthodes d'évaluation

Bases de données multimédia / Introduction 2



## Contexte

- Démocratisation de l'imagerie numérique
  - Grand public
    - PC familial multimédia
    - Appareil photo numérique
    - Magnétoscope numérique
  - Stockage de l'information
    - Plus pérenne que les bandes magnétiques
    - Accès à l'information plus facile
    - Exemple INA : 240 000 h de vidéos numérisées, 800 000 h en 2015
- Évolution des réseaux de communication
  - Internet
  - Télévision numérique



Océan d'images  
(fixes, animées)

Bases de données multimédia / Introduction 3



## Besoins des utilisateurs

- Avoir un aperçu du contenu de la base d'images
  - Partitionnement de la base → résumé visuel de la base
- Détecter et reconnaître des composantes d'images (visages, véhicules, ...)
- 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

Bases de données multimédia / Introduction 4

## ● ● ● | Indexation 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
  - Subjectivité de l'indexation
  - Dépendant de la langue
  - Dépendant du contexte
  - Coût annotation manuelle
  - Incomplétude inhérente

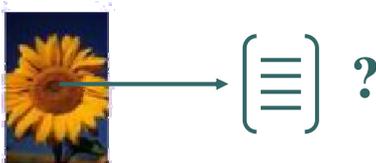


Mots-clés : écusson, ???

Bases de données multimédia / Introduction 5

## ● ● ● | Indexation par le contenu visuel

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

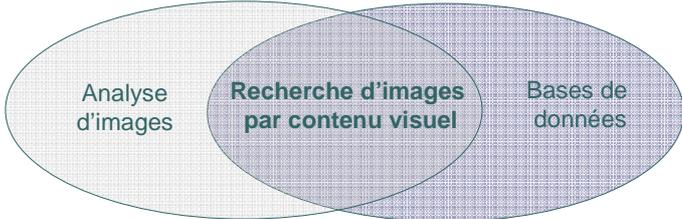


→ Prise en compte directe du contenu visuel  
... mais manifestation fréquente du **gap sémantique**

Bases de données multimédia / Introduction 6

## ● ● ● | Recherche d'images par le contenu

- Domaine émergent à l'intersection de 2 grands domaines de l'Informatique



The diagram consists of two overlapping ovals. The left oval is light blue and labeled 'Analyse d'images'. The right oval is a darker blue and labeled 'Bases de données'. The overlapping area in the center is a medium blue and labeled 'Recherche d'images par contenu visuel'.

- En anglais : CBIR pour *Content-Based Image Retrieval*

Bases de données multimédia / Introduction 7

## ● ● ● | Domaines d'application

- Applications scientifiques
  - Bases d'images médicales
    - Exemple : retrouver les images présentant un caractère pathologique, dans un but éducatif ou diagnostic
  - Bases d'images botaniques
    - Exemple : identifier les gènes qui interviennent dans une même chaîne de synthèse protéinique
  - Bases d'images satellitaires
- Audiovisuel
  - Exemple : retrouver un plan spécifique d'un film ou d'un journal télévisé (TV, INA), annotation automatique de vidéos, détection de copies (droits)
- Authentification
  - Exemple : détecter les contrefaçons de modèles déposés, identifier un visage (Visionics) ou des empreintes digitales (Thomson Idmatics), supports d'enquête (Police Judiciaire)

Bases de données multimédia / Introduction 8

## ● ● ● | Domaines d'application

- Le design, la publicité, l'architecture
  - Exemple : rechercher une texture spécifique pour l'industrie textile, illustrer une publicité par une photo adéquate, mettre en lumière des tendances
- L'art, l'éducation
  - Exemple : recherche encyclopédique d'illustrations, d'un tableau ou d'une œuvre d'art spécifique (Ministère de la culture, numismatique)
- Les télécommunications
  - Coder et transmettre les images par leurs index (MPEG-4, MPEG-7)
- Internet
  - Exemple : butinage web (lbazar.com), commerce électronique

Bases de données multimédia / Introduction 9

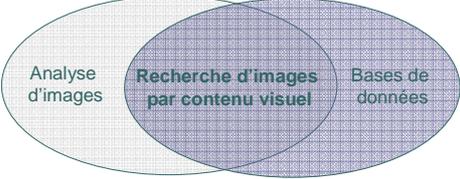
## ● ● ● | Définitions : descripteur, signature

- **Descripteur d'image**
  - Méthode d'extraction du contenu visuel de l'image (pertinent pour l'application)
    - Exemple : histogramme couleur
  - Mesure de similarité associée
    - Exemple : intersection d'histogrammes
- **Signature d'image**
  - Vecteur numérique représentant le contenu visuel de l'image
    - Exemple : 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**
  - Structuration de l'espace de description pour une recherche efficace (index)

Bases de données multimédia / Introduction 10

## ● ● ● Définitions : index

- Qu'est ce qu'un index ?



- En bases de données : l'index d'une base est une structure de données (multidimensionnelle) permettant de structurer la base pour un accès efficace aux données
  - « Indexer une base d'images »
- En image : l'index d'une image est son descripteur
  - « Indexer une image »

Bases de données multimédia / Introduction 11

## ● ● ● Définitions : contraintes fondamentales

### 1. L'invariance de la description

- Invariance possible à
  - 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
  - Exemple : forts changement d'échelle ? D'illumination ?

→ En inférer les changements auxquels les descripteurs doivent être **invariants** (ou au moins robustes) et les changements par rapport auxquels ils doivent être **discriminants**



Bases de données multimédia / Introduction 12

## ● ● ● Définitions : contraintes fondamentales

### 2. La dimension de la description

- Dimension de la signature / Nombre de signatures par image
  - Exemple 1 : 500 000 images décrites par un histogramme couleur décrit par un vecteur de dimension 216  $\Rightarrow$  recherche dans un espace de dimension 216 contenant 500 000 points
  - Exemple 2 : 500 000 images décrites par 20 régions décrites par un vecteur de dimension 40  $\Rightarrow$  recherche dans un espace de dimension 40 contenant 10 millions ( $500\,000 \times 20$ ) de points
- Déterminer le **plus petit** ensemble de grandeurs invariante
  - Problématique du passage à l'échelle (bien connue en bases de données)

Bases de données multimédia / Introduction 13

## ● ● ● 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 de données multimédia / Introduction 14

## ● ● ● Bases spécifiques : exemples

Left Loop – 32 %

Right Loop – 32 %

Whorl – 32%

Arch – 4 %

Bases de données multimédia / Introduction

15

## ● ● ● Indexation des images

- Traitement *off-line*
- Choix d'un espace de représentation
  - Calcul des *descripteurs* de l'image dans cet espace
    - Exemple : histogramme couleur de l'image

- Représentation **compacte** (– de données, + de sémantique)
- Définition d'une mesure de similarité dans cet espace

Bases de données multimédia / Introduction

16

## Indexation de la base d'images

- Construction de l'index
  - Stockage efficace des descripteurs en machine
  - Structurer le nuage de points dans l'espace des caractéristiques, pour réduire ultérieurement le coût de la requête
  - Compromis entre
    - Stockage (mémoire)
    - Recherche (temps de réponse)
- Index multidimensionnels
  - B-tree, Kd-tree, R-tree, X-tree, etc
  - tables de hachage
  - VA-files
  - SASH
  - ...

Bases de données multimédia / Introduction 17

## Architecture générale d'un SRI

- SRI = Système de Recherche d'Images
- Architecture client-serveur

Bases de données multimédia / Introduction 18

## Les SRI les plus connus

- Systèmes industriels
  - QBIC (IBM, 1995) : <http://www.qbic.almaden.ibm.com>
  - ImageFinder (Attrasoft) : <http://www.attrasoft.com>
  - Excalibur (Excalibur Technologies, 1996) : <http://www.excalib.com>
  - Virage (Virage Technologies, 1996) : <http://www.virage.com>
- Systèmes académiques
  - Ikona (INRIA Rocquencourt – IMEDIA) : <http://www-rocq.inria.fr/imedia/>
  - Blobworld (Université de Californie – Berkeley)
  - Photobook (Massachusetts Institute of Technology)
  - Viper (Université de Genève – Computer Vision Group)
  - SIMPLcity (Stanford University)
- Quelques comparaisons
  - <http://www.jtap.ac.uk/reports/htm/jtap-054.html>

Bases de données multimédia / Introduction 19

## Exemple d'interface : Ikona

The screenshot shows the IKONA v0.9.3 interface. It features a main window titled 'IKONA v0.9.3 (Main Window)' with a menu bar (Files, Options, Help) and a toolbar. The main area is divided into two panels: an 'Image Panel' and a 'Query Panel'. The 'Image Panel' displays a grid of 16 image thumbnails, each with a small '0.0' score and an 'IDS' label. The 'Query Panel' contains buttons for 'Visual', 'Feedback', 'Object', and 'Quit', along with a 'Keywords' input field. A red box highlights one of the images in the grid, with an arrow pointing to it from the text 'Image requête'. Below the grid, there is a 'Navigation Panel' with buttons for 'Prev', 'Next', 'Back', and 'Shuffle'. At the bottom, a status bar shows the current database ('[Database: IDS]'), the filename of the selected image ('[Filename: mer\_a\_000652.37.jpg]'), and the distance ('[Distance: 0.0]').

Visualisation aléatoire de la base

Outils de navigation

Image requête

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

Bases de données multimédia / Introduction 20

## Exemple d'interface : Ikona (2)

Requête (ici l'image entière)

Réponses classées par ordre de similarité visuelle décroissante

Navigation Panel: Prev, Next, Back, Shuffle

Query Panel: Visual, Feedback, Object, Quit

Keywords: \_\_\_\_\_

[Database: IDS] [Filename: mer\_a\_000652\_37.jpg] [Distance: 0.0]

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

Bases de données multimédia / Introduction 21

## Typologie des descripteurs d'images

- Description globale de l'image
  - Description approximative de toute l'image
  - Solutions
    - Couleur, Texture, Forme
- Description locale de l'image
  - Description précise des parties de l'image
  - Solutions
    - Modèles pour la reconnaissance d'objets
    - Sans modèles (segmentation en régions, détection de points d'intérêt)
- Descripteurs spécifiques (souvent en biométrie)
  - Empreintes digitales : minuties
  - Visages : EigenFaces, ...

Bases de données multimédia / Introduction 22

## ● ● ● Description globale de l'image



signature de couleur

↓

Histogramme,...



signature de texture

↓

Fourier, ondelettes ...



signature de forme

↓

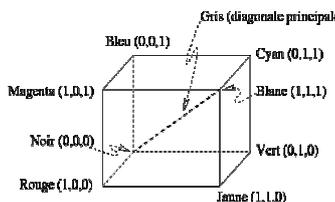
Moments, EOH, ...

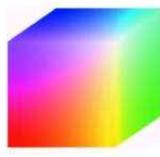
A combiner pour une description plus riche

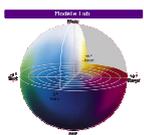
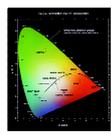
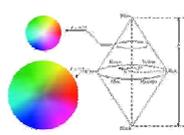
Bases de données multimédia / Introduction
23

## ● ● ● Signatures de couleur

- Choix d'un espace de représentation des couleurs
  - RVB
  - CMY
  - HSV
  - CIE Lab
  - CIE Luv
  - ...





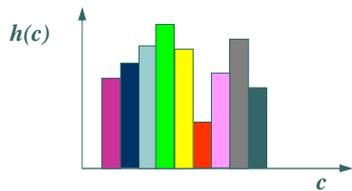
- Certains espaces colorimétriques sont mieux adaptés à la recherche d'images par contenu
  - Deux couleurs proches en distances le sont aussi pour l'œil.

Bases de données multimédia / Introduction
24

## Exemple : histogrammes couleur

- Échantillonnage des couleurs de l'image
  - Exemple : image dont les pixels sont codés sur 24 bits, échantillonnage en 216 couleurs (= 6×6×6)
- Calcul de l'histogramme pour chaque couleur  $c$ 

$$\forall c \in C \quad \tilde{h}(c) = \frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} \delta(f(i, j) - c)$$



Bases de données multimédia / Introduction 25

## Similarité entre histogrammes couleur

- Intersection d'histogrammes
 
$$\bigcap(H, M) = \frac{\sum_i \min(h_{c_i}, m_{c_i})}{\sum_i m_{c_i}}$$
- Distances
  - $L_1$
  - $L_2$  généralisée  $d_{L_2}(H, M) = \sqrt{(H - M)A(H - M)^T}$
  - $L_\infty$   $d_{L_\infty}(H, M) = \max_i |h_{c_i} - m_{c_i}|$

Bases de données multimédia / Introduction 26

## ● ● ● Histogrammes couleur : exemple

(avec  $L_1$ )

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

Bases de données multimédia / Introduction 27

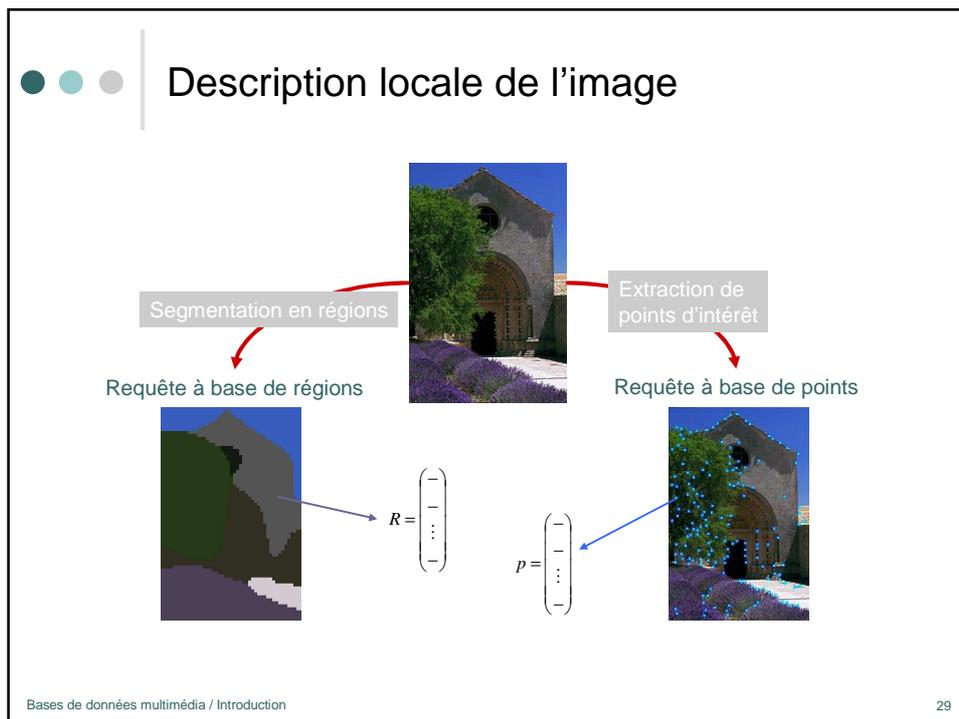
## ● ● ● Histogrammes couleur

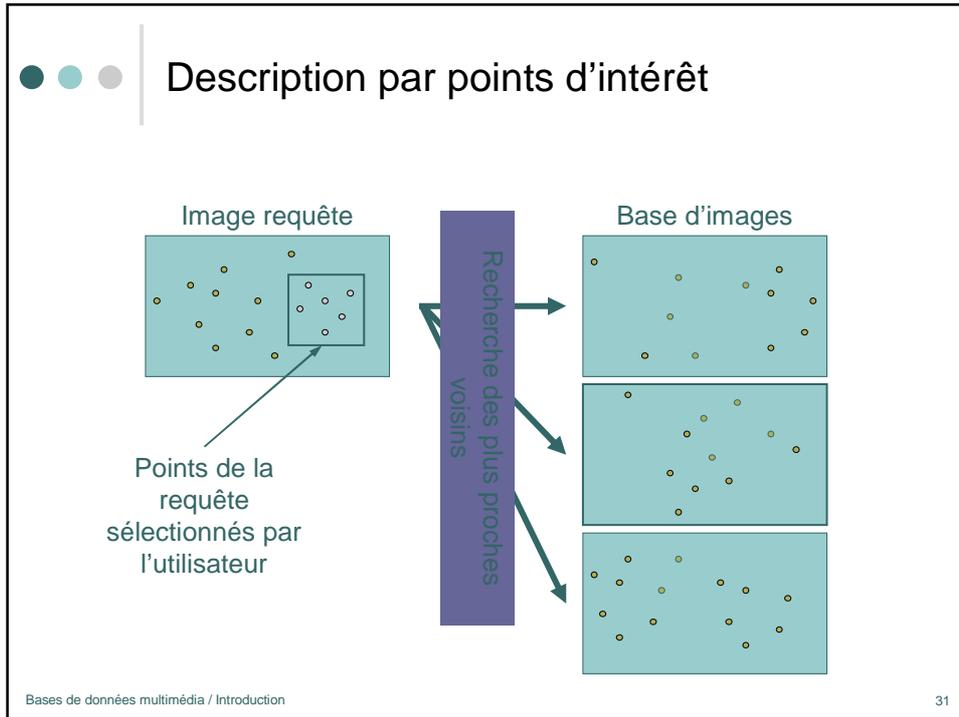
- Avantage
  - Robustes à certaines transformations géométriques de l'image
- Limitations

Contenu visuel différent

mais histogrammes similaires

Bases de données multimédia / Introduction 28





### Description par points d'intérêt

The screenshot shows a software interface for image search by interest points. On the left, a **Requête** (query) window displays a sunflower image with a blue box highlighting a region of interest. The main window, titled **IKONAWO3 (Main Window)**, displays a grid of search results. Each result consists of a small image and its corresponding ID and distance from the query image. The results are as follows:

0.0 flore_a_000213 90.jpg	35.561395 flore_a_000214 25.jpg	39.603794 flore_id_178a 117.jpg	45.370824 flore_a_000302 04.jpg
45.446329 flore_a_000488 84.jpg	46.139389 flore_a_000335 17.jpg	46.661137 flore_a_000426 40.jpg	46.956783 flore_a_000427 54.jpg
47.038002 516072.jpg	48.149837 516057.jpg	48.158958 flore_00031553.jpg	48.490685 flore_a_000488 83.jpg
48.725494 516081.jpg	48.841908 516038.jpg	48.87896 flore_id_180a 132.jpg	48.997284 516063.jpg

The interface also includes a **Navigation Panel** with buttons for **Prev**, **Next**, **Back**, and **Shuffle**. A **Query Panel** contains buttons for **Visual**, **Feedback**, **Object**, and **Quit**, along with a **Keywords** input field. The status bar at the bottom shows: **[Database: Demo\_500ppil] [Filename: 516072.jpg] [Distance: 47.038002]**. The URL <http://www-rocq.inria.fr/imedia> is displayed at the bottom right.

32

## Description par points d'intérêt

Requête

Close

IKONOS v1.3 (Main Window)

Files Options Help

Image Panel

Navigation Panel

Query Panel

Database: Demo\_500ppi | Filename: aim1mb05-0011056.jpg | Distance: 1.699821

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

Bases de données multimédia / Introduction

33

## Description par points d'intérêt

- Avantages
  - Recherche de parties ou d'objets (élimination de l'arrière-plan)
  - Sélection dynamique de la requête (dite **partielle**)
  - Robustesse aux transformations image
  - Robustesse aux occultations
- Inconvénients
  - Coûteux en temps de recherche
    - Exemple : 500 000 images décrites par 300 points d'intérêt décrits par un vecteur de dimension 17  $\Rightarrow$  recherche dans un espace de dimension 17 contenant 150 millions de points ( $500\,000 \times 300$ )
  - $\rightarrow$  Nécessité de **structurer l'espace de recherche** avec un index multidimensionnel

Bases de données multimédia / Introduction

34

## ● ● ● | Paradigmes de recherche

- **Initialisation** de la recherche (problème de la page zéro)
  - 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

Bases de données multimédia / Introduction 35

## ● ● ● | 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)



Bases de données multimédia / Introduction 36

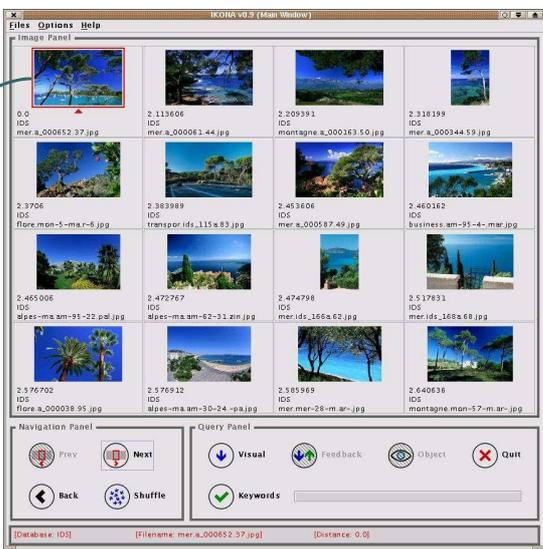
<http://www-rocq.inria.fr/media>



## Recherche globale par l'exemple

Requête (ici l'image entière)





Réponses classées par ordre de similarité visuelle décroissante

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

## Recherche locale par l'exemple

Requête (ici une partie d'image)





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

## ● ● ● Contrôle de pertinence

- Objectif : retrouver un ensemble d'images pertinentes par rapport à la recherche en cours
  - En général caractérisées par des relations de similarité plus complexes
- Principe : raffinement itératif exploitant les réponses de l'utilisateur
  - Peut être vu 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

Bases de données multimédia / Introduction 41

## ● ● ● Contrôle de pertinence : exemple

Objectif : retrouver des peintures de Cézanne

Disponible : description globale (couleur, texture, forme)

■ Exemples positifs

■ Exemples négatifs



Bases de données multimédia / Introduction 42

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

## ● ● ● Contrôle de pertinence : exemple (2)

Réponses après plusieurs itérations :  
12 images de peintures de Cézanne sur la première page

Files Options Help  
IKORA v0.5 (Main Window)

Files Panel

Navigation Panel

Query Panel

Database: IqDatabase | Filename: cezanne.bord-route.jpg | Existence: 1.7004075

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

43

## ● ● ● Composition à partir de thésaurus

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

Formulation de la requête

Composition logique correspondante

Catégorie « régions à dominante verte »

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

44

## ● ● ● Composition à partir de thésaurus (2)

Résultat de la requête



Exemples d'images rejetées



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

Bases de données multimédia / Introduction 45

## ● ● ● Indexation multimédia

- Combiner le texte et l'image
  - Lors de l'indexation
  - Lors de la recherche

Bases de données multimédia / Introduction 46

## Indexation multimédia

### Scénario d'indexation

- Propagation de mots-clés :


→ Similarité visuelle →


Mots-clés : Claire Chazal, gp. ← Propagation → Mots-clés : Claire Chazal, gp.

- ✓ Indexation textuelle automatique
- ✓ Auto-complétion

Bases de données multimédia / Introduction 47

## Indexation multimédia

### Scénario de recherche

- Caractéristique de chaque mode :
  - Indexation textuelle : sémantique de haut niveau.
    - Ex : Fashion, Sarah Moon, 1998.
  - Indexation d'images : propriétés syntaxiques de l'image
    - Ex : couleurs, disposition des zones...
- Apport de la combinaison texte/image :
  - Réduire les ambiguïtés ou lacunes textuelles par l'apparence visuelle,
  - Comblent l'insuffisance sémantique des descripteurs image.



Fashion, Sarah Moon, 1998

Exemple de requête : *« je cherche la photo dont l'auteur est Sarah Moon et dans laquelle le mannequin porte une jupe blanche et une veste noire ».*

Bases de données multimédia / Introduction 48

## Évaluation

1. Évaluation directe avec les utilisateurs
  - Coûteuse
  - Complexe à mettre en place
2. Évaluation à partir d'une vérité-terrain
  - En général, la base est partitionnée en classes de même effectif, mutuellement exclusives
  - Notations :  $N$  images,  $C$  classes,  $m (= N/C)$  images par classe
  - Mesures de performances
    - Courbes de précision/rappel
    - Matrices de confusion
    - Courbes ROC (*Receiver Operating Characteristics*)
    - Rapidité de convergence de la recherche itérative

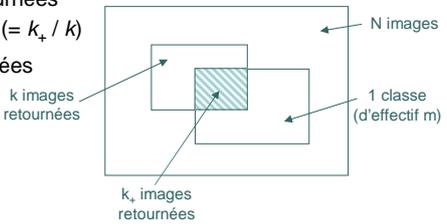
Bases de données multimédia / Introduction 49

## Courbes précision/rappel

- Base vérité-terrain
- Pour  $k$  images retournées et classes d'effectif  $m$  :
  - **Précision** : proportion d'images retournées faisant réellement partie de la classe ( $= k_+ / k$ )
  - **Rappel** : proportion d'images retournées de la classe ( $= k_+ / m$ )
- Courbe précision/rappel :
  - Elle comptabilise la précision pour chaque rappel  $1/m, 2/m, \dots, 1$

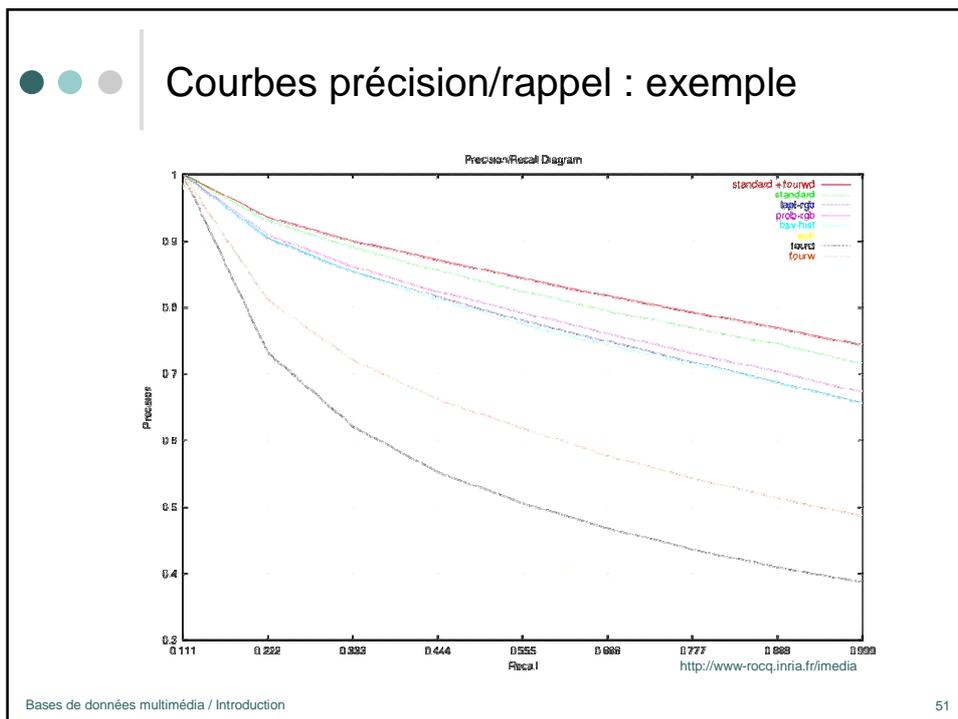


Exemple de classe




→


Bases de données multimédia / Introduction 50



### Courbes ROC : sensibilité, spécificité

- Évaluation de la discrimination entre présence d'une classe et absence de la classe (ex. : détection d'objets)
- Table de vérité
 

		Vérité terrain	
		Classe présente	Classe absente
Décteur	Classe détectée	Vrai Positif	Faux Positif
	Classe non détectée	Faux Négatif	Vrai Négatif
- Sensitivité et spécificité pour un seuil de décision fixé
 
$$\text{Sensitivité} = \frac{VP}{\text{Total positif}} = \frac{VP}{VP + FN}$$

$$1 - \text{Spécificité} = \frac{FP}{\text{Total négatif}} = \frac{FP}{VN + FP} = 1 - \frac{VN}{VN + FP}$$

Bases de données multimédia / Introduction 52

### Courbes ROC

- o Courbe ROC (*Receiver Operating Characteristics*) : sensibilité vs. 1-spécificité pour différents seuils de décision

The graph plots Sensibilité (Sensitivity) on the y-axis against 1-Spécificité (1-Specificity) on the x-axis, both ranging from 0 to 1. A dashed diagonal line represents a random classifier. Three points on the curve are marked with green dots and labeled: 'Seuil = 0.2' at approximately (0.05, 0.25), 'Seuil = 0.4' at approximately (0.1, 0.55), and 'Seuil = 0.6' at approximately (0.15, 0.75). The curve itself is a solid green line that starts at (0,0) and approaches (1,1) as the threshold increases.

Bases de données multimédia / Introduction 53

### Rapidité de convergence

- o Proportion des images de la classe retrouvées après  $n$  itérations (gauche) ou *clicks* (droite)

The left graph shows Mean precision (y-axis, 0.4 to 1.0) versus Iterations (x-axis, 0 to 25). The right graph shows Mean precision (y-axis, 0.2 to 0.9) versus Clicks (x-axis, 20 to 200). Both graphs compare seven methods: STO (crosses), ANN (asterisks), GRE (squares), GCO (open squares), MIN (filled squares), OPT (open circles), and TIR (filled circles). In both graphs, precision increases with iterations/clicks and converges towards 1.0. The TIR method generally shows the highest precision, followed by GRE and MIN.

Bases de données multimédia / Introduction 54