


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

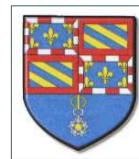
- Par quoi indexer : mots-clés, graphes conceptuels, ...



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, ???



Indexation par le contenu visuel

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

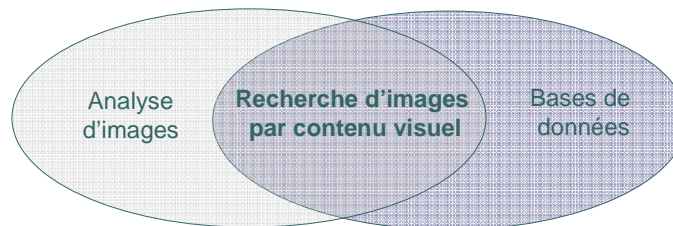


- Prise en compte directe du contenu visuel
... mais manifestation fréquente du **gap sémantique**



Recherche d'images par le contenu

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



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



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)



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

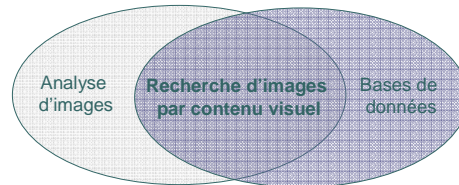


Définitions : descripteur, signature

- **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)

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 »

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**





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 invariantes
 - Problématique du passage à l'échelle (bien connue en bases de données)

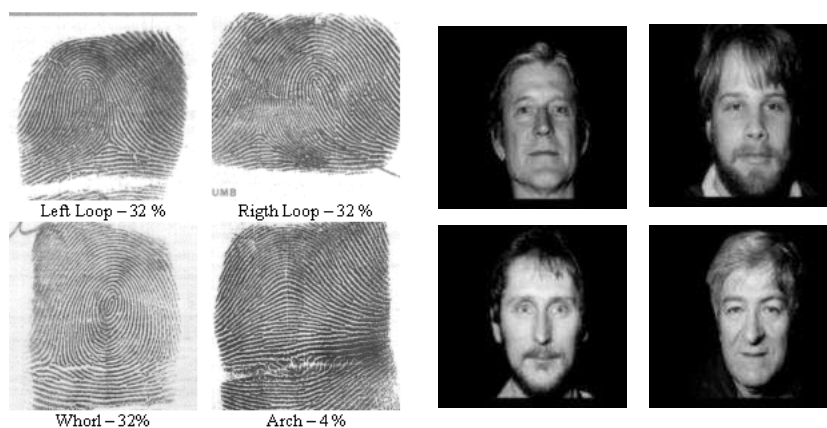


Types de bases d'images

- 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



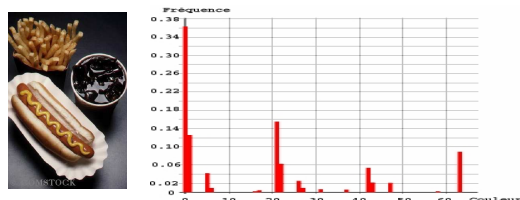
Bases de données multimédia / Introduction

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



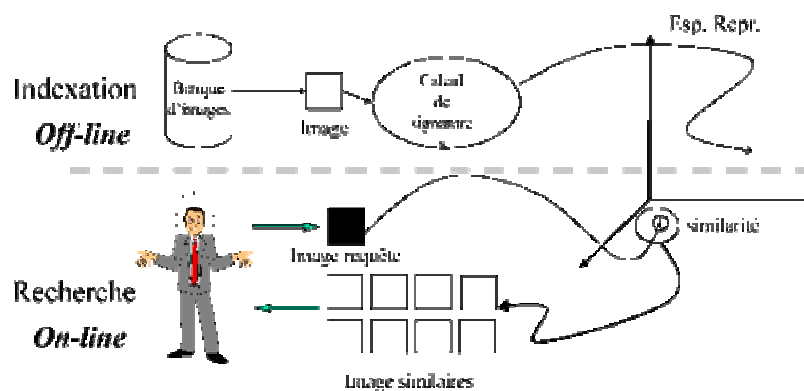
Indexation de la base d'images

- 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
 - ...



Architecture générale d'un SRI

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



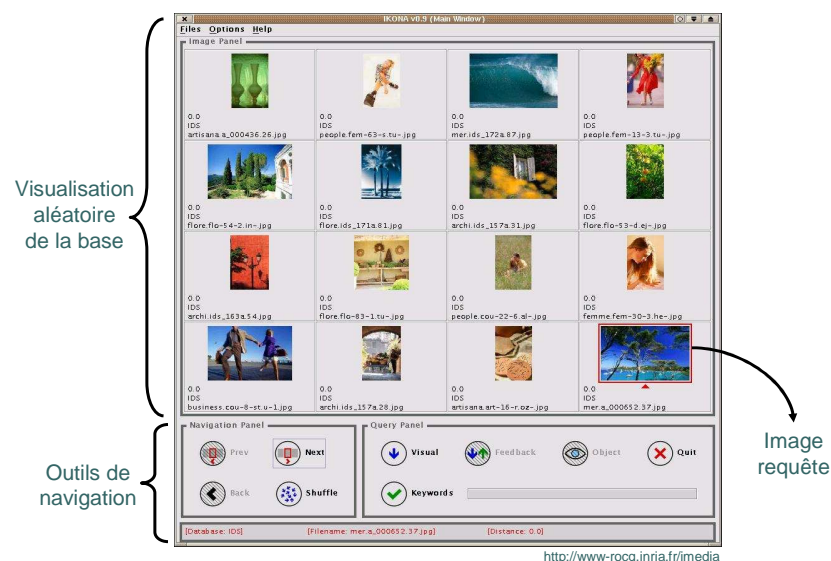


Les SRI les plus connus

- 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>



Exemple d'interface : Ikona



Exemple d'interface : Ikona (2)



Bases de données multimédia / Introduction

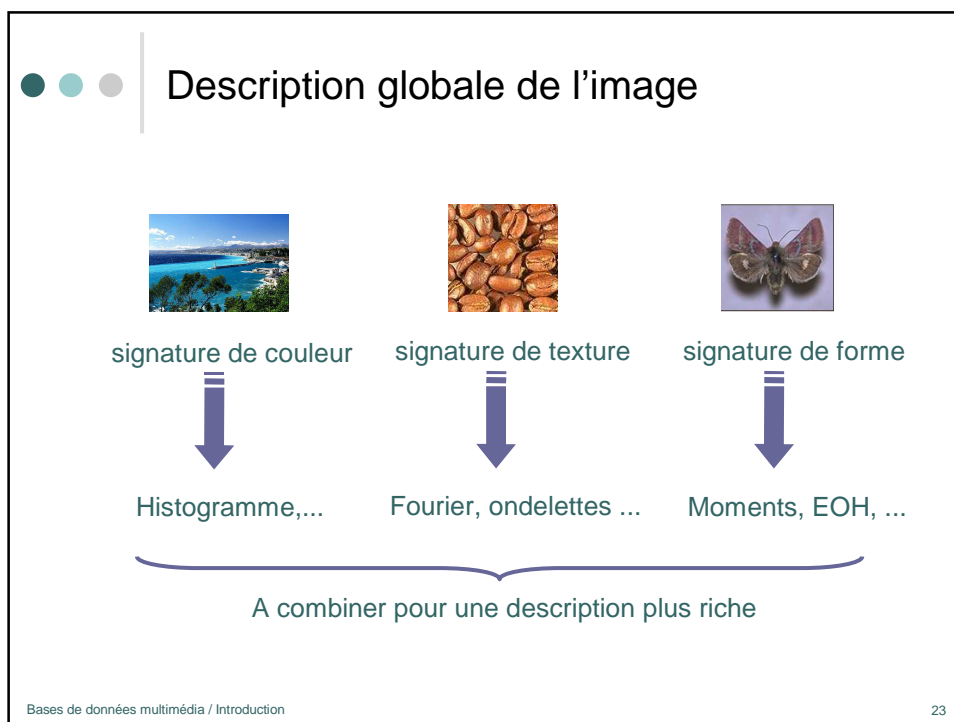
21

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, ...

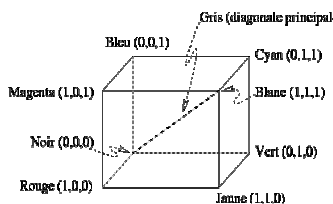

Bases de données multimédia / Introduction

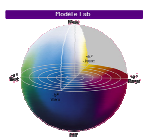
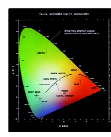
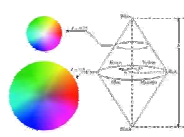
22



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.

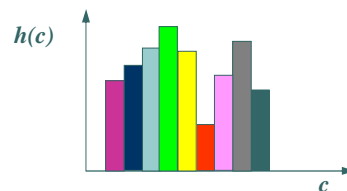
Bases de données multimédia / Introduction 24



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

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



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_∞ $d_{L_\infty}(H, M) = \max_i |h_{c_i} - m_{c_i}|$

Histogrammes couleur : exemple

(avec L_1)



Bases de données multimédia / Introduction

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

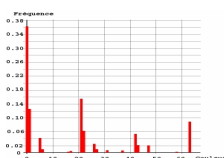
27

Histogrammes couleur

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



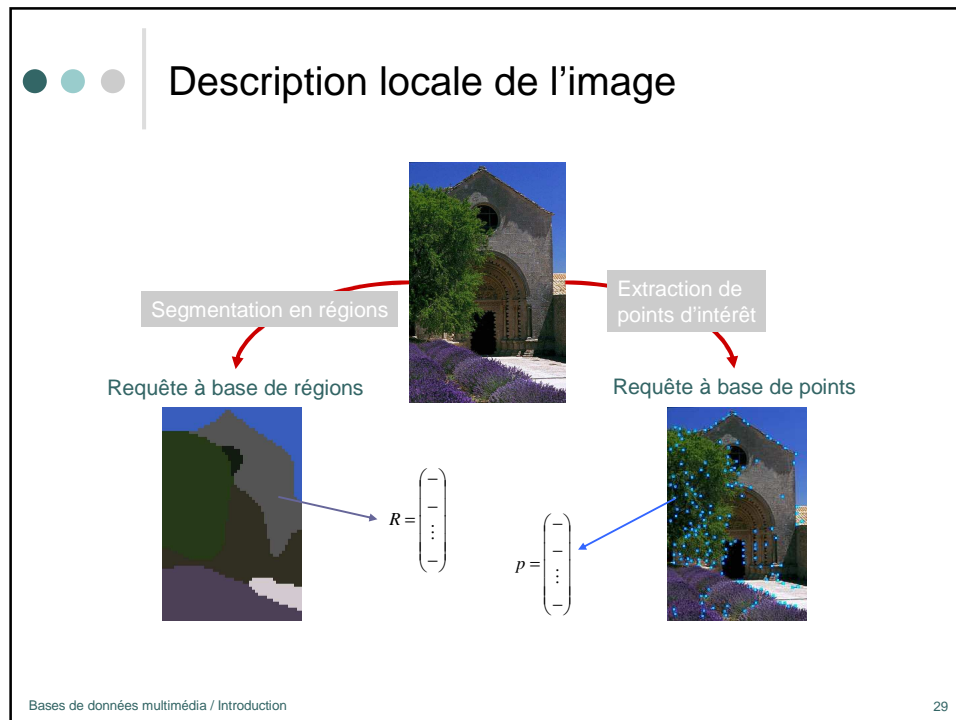
Contenu visuel différent



mais
histogrammes similaires

Bases de données multimédia / Introduction

28



Description par points d'intérêt

1. Extraction automatique de points d'intérêt dans chaque image
 - 1 point d'intérêt = 1 site informatif de l'image
2. Description locale de chaque point d'intérêt
 - Exemple : un point peut être décrit par 8 invariants couleur à la translation et à la rotation
3. Mise en place d'une mesure de similarité entre descripteurs de points d'intérêt

Bases de données multimédia / Introduction 30

Description par points d'intérêt

Image requête

Points de la requête sélectionnés par l'utilisateur

Recherche des plus proches voisins

Base d'images

Bases de données multimédia / Introduction

31

Description par points d'intérêt

Requête

IKON-Web (Main Window)

Image Panel

0 0
flore_a_000233_90.jpg

35 561295
flore_a_000234_25.jpg

39 603794
flore_id_178a_117.jpg

45 370824
flore_a_000302_04.jpg

45 445229
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_id_1553.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
516065.jpg

Navigation Panel

Prev

Next

Back

Shuffle

Query Panel

Visual

Feedback

Object

Quit

Keywords

Database: Demo_500ppil

Filename: 516072.jpg

Distance: 47.038003

Bases de données multimédia / Introduction

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

32

CNAM / Valérie Gouet-Brunet

Description par points d'intérêt

Requête

Bases de données multimédia / Introduction

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

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 ($500\,000 \times 300$)

→ 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



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)



Résumés visuels

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 *off-line*



NoiseCluster - 64 images

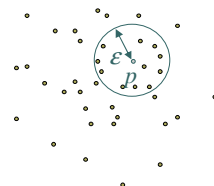
<http://www-rocq.inria.fr/imedia>

Bases de données multimédia / Introduction

37

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



Bases de données multimédia / Introduction

38



Recherche globale par l'exemple

Requête
(ici
l'image
entière)

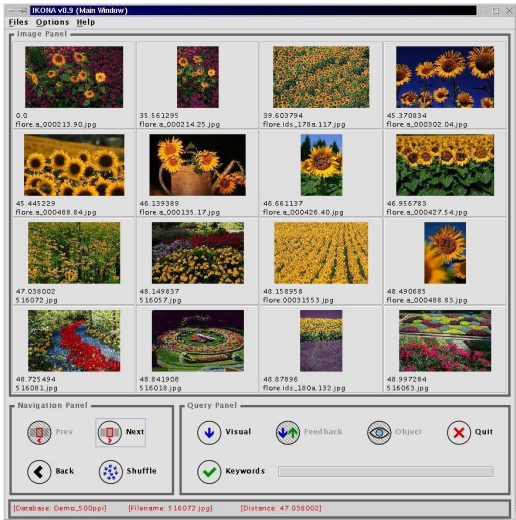
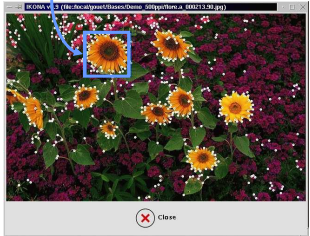


Réponses
classées
par ordre
de
similarité
visuelle
décroissante



Recherche locale par l'exemple

Requête (ici une
partie d'image)





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



Contrôle de pertinence : exemple (2)

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



Bases de données multimédia / Introduction

<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

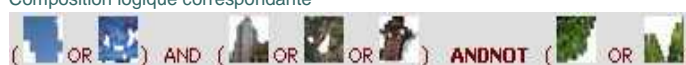


Formulation de la requête



Catégorie « régions à dominante verte »

Composition logique correspondante



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

Bases de données multimédia / Introduction

44



Composition à partir de thésaurus (2)

Résultat de la requête



Exemples d'images
rejetées



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



Indexation multimédia


○ Combiner le texte et l'image

- Lors de l'indexation
- Lors de la recherche

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



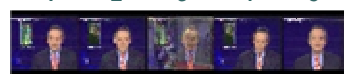
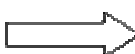
Courbes précision/rappel

- Base vérité-terrain
- Pour k images retournées et classes d'effectif m :

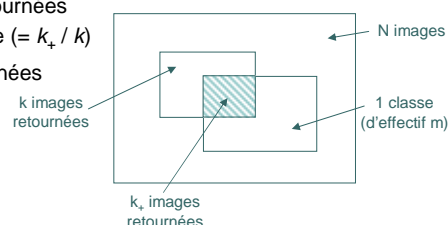
- **Précision** : proportion d'images retournées faisant réellement partie de la classe ($= k_+ / k$)
- **Rappel** : proportion d'images retournées de la classe ($= k_+ / m$)

- Courbe précision/rappel :

- Elle comptabilise la précision pour chaque rappel $1/m, 2/m, \dots, 1$



Exemple de classe

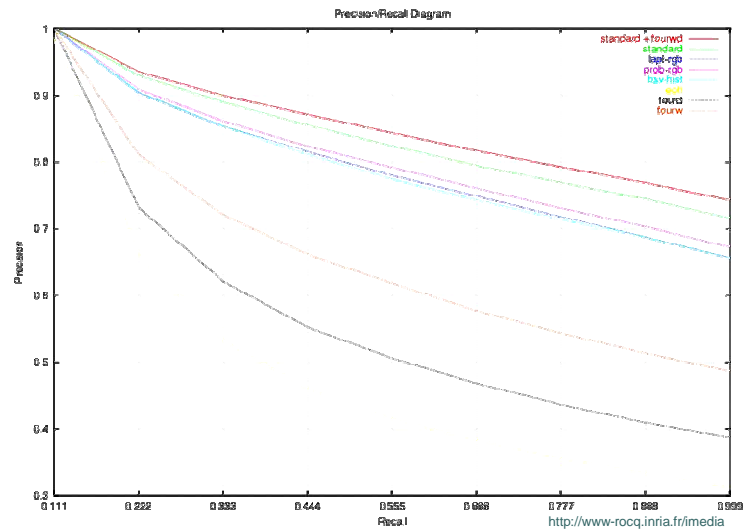


Bases de données multimédia / Introduction

Requête

50

Courbes précision/rappel : exemple



Bases de données multimédia / Introduction

51

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

- Sensibilité et spécificité pour un seuil de décision fixé

$$\text{Sensitivité} = \frac{VP}{\text{Total positif}} = \frac{VP}{VP + FN}$$

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

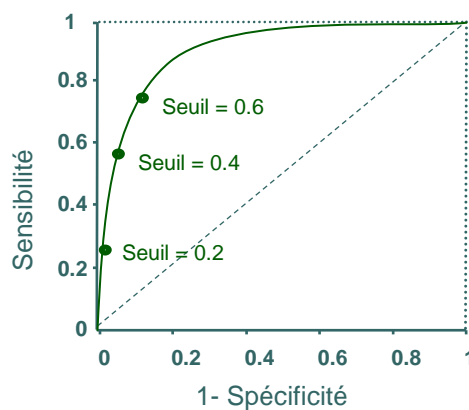
Bases de données multimédia / Introduction

52



Courbes ROC

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



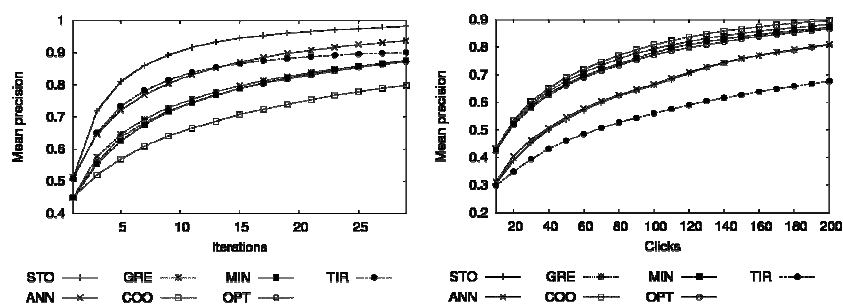
Bases de données multimédia / Introduction

53



Rapidité de convergence

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



Bases de données multimédia / Introduction

54