Indexation d'images

Cours Master IA&D

Techniques du traitement d'images

UNIVERSITE PIERRE & MARIE CURIE
Antoine Manzanera - ENSTA /Unité d'Électronique et d'Informatique
ENSTA

Indexation - Introduction

Le sujet de ce cours est la *recherche automatique* de documents *visuels* (images, séquences video), dans des bases de données de grande taille, à partir de requêtes relatives au *contenu* de ces documents.
Ce problème fait actuellement l'objet de recherches très abondantes dans le domaine du traitement d'images et de la vision par ordinateur. En effet, la *généralisation des supports numériques*, l'apparition de *formats video compacts*, la *chute du coût des média de stockage* a engendré une augmentation vertigineuse de la *quantité des données multimedia*. Pour que ces données soient exploitables, il faut qu'elles puissent être consultées efficacement comme par le biais d'un catalogue.
Les techniques présentées ci-après, dite d'*indexation*, se proposent d'attacher à une image ou à une video un ensemble de *descripteurs de leur contenu*, dans le but de mesurer la *ressemblance* avec les descripteurs correspondant à la requête.
Mais cette requête peut prendre des formes très différentes, elle peut être conceptuelle (ex : mot), symbolique (ex : schéma) ou instancielle (ex : une autre image).
De la même façon, l'indexation sera *sémantique* (on attache des descripteurs de niveau conceptuel au document) ou *visuelle* (on attache des descripteurs de niveau visuel au document).

Applications et enjeux

BD Images et video :

· Collections et catalogues des particuliers, entreprises
· Médiathèques
· Agences de photographie
· Archives audiovisuelles (ex. INA)
· Internet (ex.AltaVista/Virage)

Applications :

· Médiamétrie (ex. empreintes digitales)
· Propriété des oeuvres
· Reconnaissance de visages, d'objets...
· Données biomédicales
· Imagerie satellitaire, aérienne
· Video de télésurveillance

Plan du cours *Indexation*

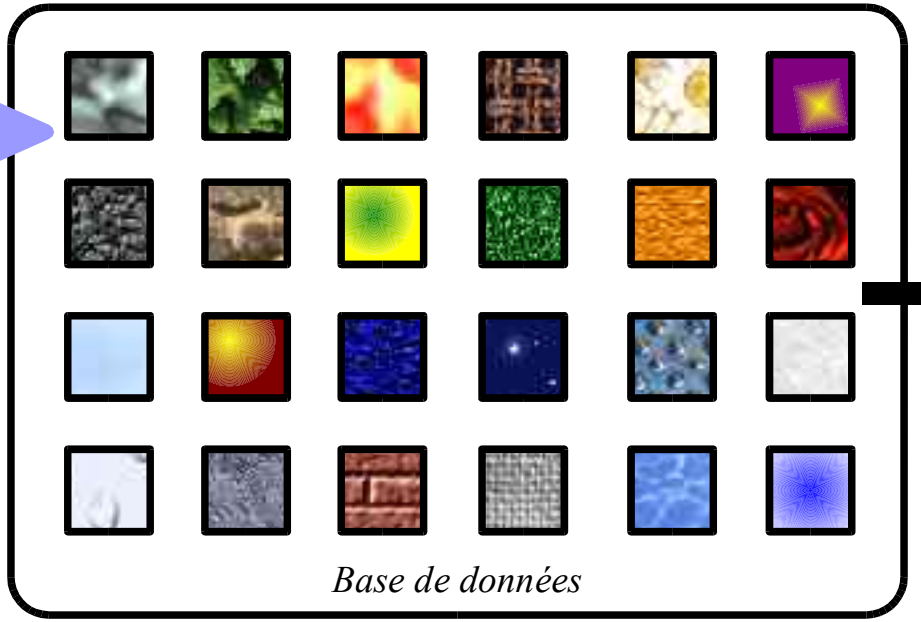
- Indexation multimedia : Etat actuel et Perspectives
  - Recherche de documents multimedia par le contenu
  - Indexation sémantique manuelle
  - Indexation visuelle automatique
- Aide à l'indexation manuelle
  - Sémantique de l'indexation video
  - Découpage en plans
  - Détection d'objets
- Indexation automatique
  - Images structurées et texturées
  - Extractions des descripteurs
  - Appariement d'images
  - Métriques de similarité

Recherche multimedia par le contenu

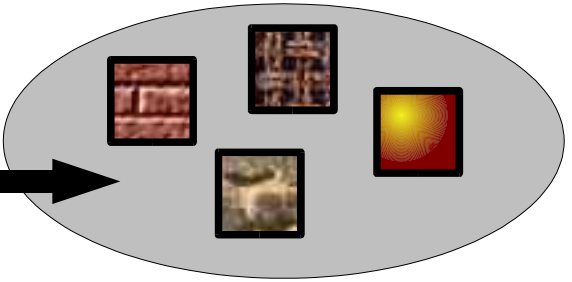
Requête

Types de requête :

- Mot
- Image
- Dessin, schéma
- Modèle CAO
- Carte
- Plan
- .../...



Résultat



Evaluation des résultats :

$$Précision = \frac{\#\text{ réponses pertinentes}}{\#\text{ réponses}}$$

$$Rappel = \frac{\#\text{ réponses pertinentes}}{\#\text{ données pertinentes}}$$

Difficultés :

- à la différence de données textuelles, le contenu sémantique n'est jamais explicite.
- les requêtes sont difficiles à exprimer, donc en général ambiguës, incomplètes.

Dimension multidisciplinaire :

Multimedia : texte, image, son - Problèmes de gestion de bases de données - Problèmes hardware - Problèmes liés à l'apprentissage - Problèmes linguistiques,...

Etat actuel : Indexation explicite

Les outils de recherche de documents multimedia qui fonctionnent actuellement sont basés sur une recherche de mots clefs *explicitement attachés* au document ou *indexés automatiquement à partir du texte environnant* (Ex : Google.)
Les documents video telles que les archives audiovisuelles sont *indexés manuellement* par des opérateurs spécialisés, à partir d'un descriptif très précis lié à un thesaurus.
Mais cette indexation manuelle s'avère une tâche pénible et longue (jusqu'à 10 fois la durée d 'une séquence, alors que par exemple le fonds de document télévisuel de l'INA représente 350.000 heures de programmes...)
De plus une donnée intéressante à une date donnée ne l'était pas forcément à la date de l'indexation...

Il faut également citer l'émergence de nouveaux standards de codage video tels que *Mpeg7* qui intègre dans le codage des données explicites relatifs aux contenus audiovisuels, dans le but de faciliter à la fois la recherche d'information dans une base de données video, et la navigation « intelligente » dans une video.

Indexations sémantique et descriptive

Par nature, l'indexation *manuelle* est *sémantique*. L'opérateur d'indexation attache au document des données de haut niveau relatifs à la *signification* du contenu de l'objet. Les requêtes associées sont en général des *mots*, désignant un *objet*, une *action*, le nom d'un *personnage* ou d'un *événement*.

Par opposition, l'indexation *automatique* est essentiellement *descriptive* ou *visuelle*. L'algorithme d'indexation attache des données de *bas niveau* sémantique, relatifs aux contenus *géométrique*, *spectral*, de l'image, à un niveau *local* ou *global*. Les requêtes associées se font en général par l'exemple, ou par modèle.

Mais l'analyse automatique de documents peut également être utilisée pour rendre plus facile (plus rapide, moins pénible) le travail de l'opérateur d'indexation manuelle. Cela concerne typiquement :

- pré-tri de grosses bases de données images.
- indexation automatique aiguillée par opérateur.
- découpage de video et simplification en image-clefs.

Sémantique des séquences d'images

Avant d'analyser automatiquement ou manuellement une séquence d'images, il faut avoir défini précisément la façon dont la video va être *structurée*. La structuration classique d'une video est celle d'un découpage en *scènes* avec titre, résumé, mots-clefs.
Les outils d'aide à l'indexation video peuvent se fonder sur une structuration précise des videos, utilisant plusieurs niveaux d'analyse.
Le premier objectif est de fournir un *cadre riche et rigoureux* pour faciliter l'indexation manuelle.
Le second objectif est de *diminuer le niveau sémantique* du découpage par scènes pour permettre l'utilisation d'outil d'indexation visuelle automatique.

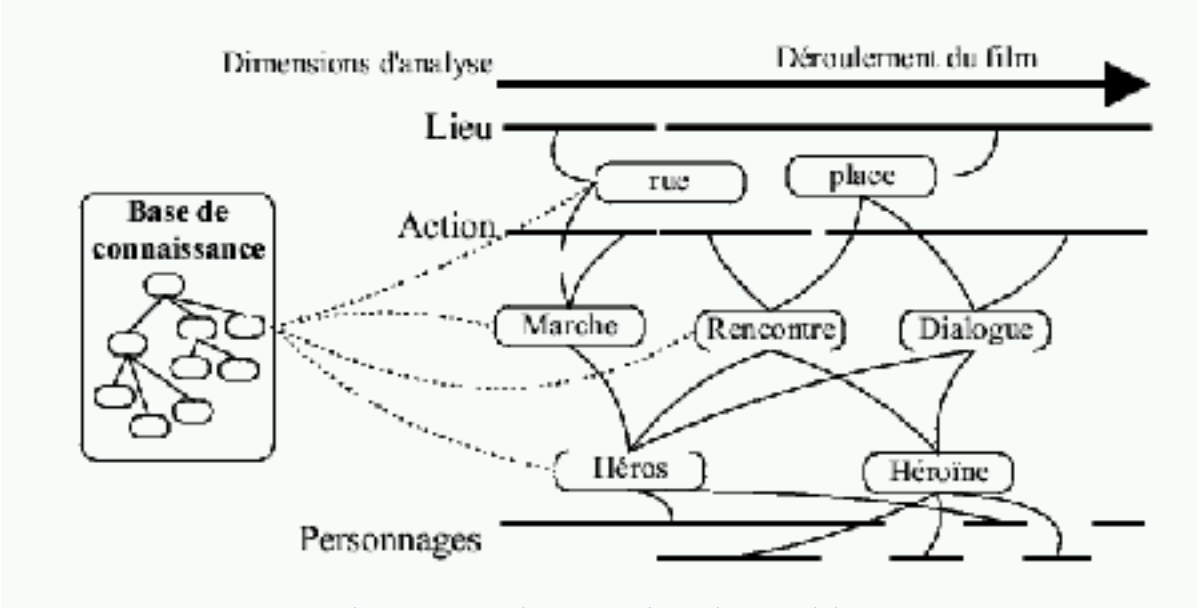


Diagramme d'annotation d'une video
(Projet Sesame – Insa Lyon / RFV)

Scène :

- Nature du lieu
- Présence d'un objet, d'un personnage
- Plan

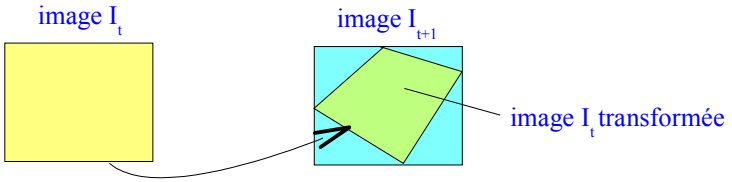
Aide à l'indexation video

Exemple : Découpages en plans (cuts) d'une video

Les techniques employées sont diverses ; elles sont en général basées sur la détection de discontinuités temporelles d'un ou de plusieurs descripteurs globaux associés à :

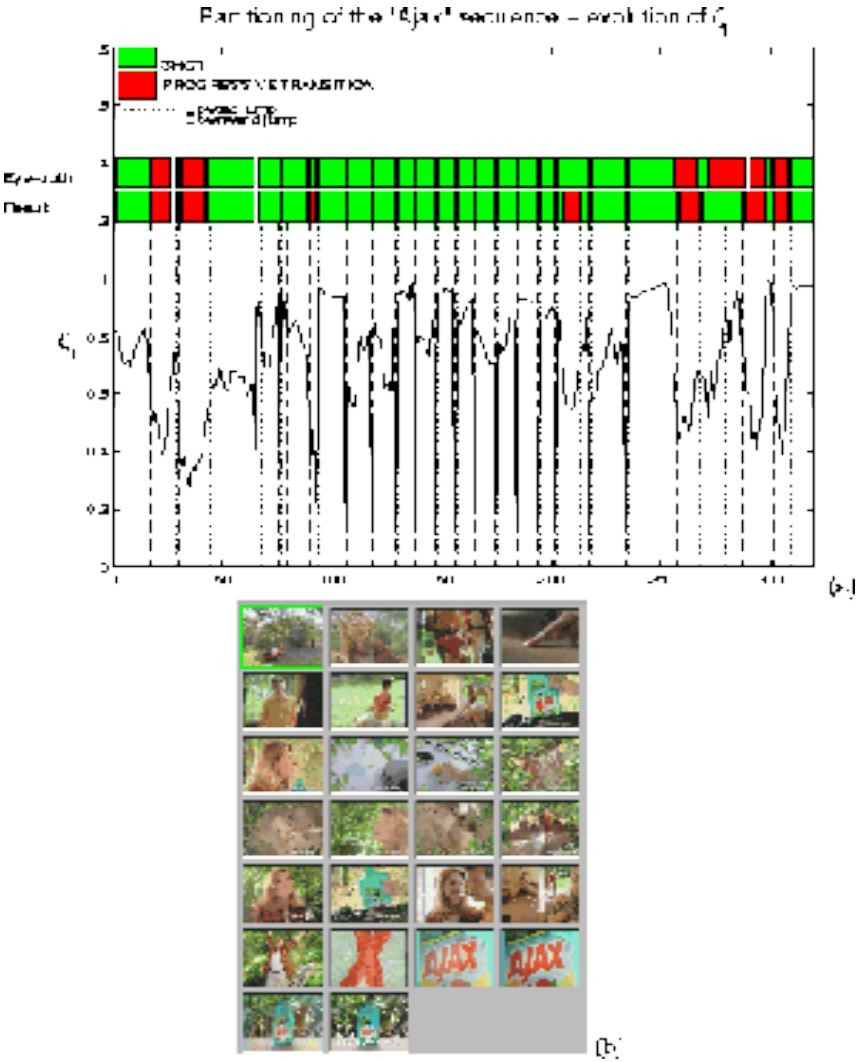
* La *couleur*. Ex : moments d'histogrammes couleurs.

* Le *mouvement*. Ex, ci-contre : extraction du mouvement dominant (transformation affine 2d), et mesure du taux de recouvrement entre image et image transformée.



Difficultés :

- Fondu-enchainés,
- Mouvements brusques,...



Logiciel *MD-shots* (IRISA Rennes / projet VISTA) de découpage video, basé sur un descripteur global du mouvement dominant (axe vertical)

Aide à l'indexation video

Un problème qui accompagne souvent celui de la segmentation en plans pour l'aide à l'indexation video est l'extraction d'*images-clefs* dans chaque plan, c'est-à-dire d'images « les plus représentatives » du plan. Les techniques utilisées actuellement reposent généralement sur des statistiques liées aux descripteurs utilisés pour le découpage en plans. On peut alors utiliser l'image médiane, les images extrêmes,...

Enfin, certaines *techniques spécifiques* de *détection*, *reconnaissance*, *identification* sont utilisées pour effectuer des tâches particulières d'aide à l'indexation. Ce sont typiquement :

* La détection et le suivi des objets mobiles.

* La détection d'objets particuliers :
  visages, véhicules, texte incrusté pour indentifier le type de scène

* Identification : le visage d'un personnage, un véhicule particulier,...

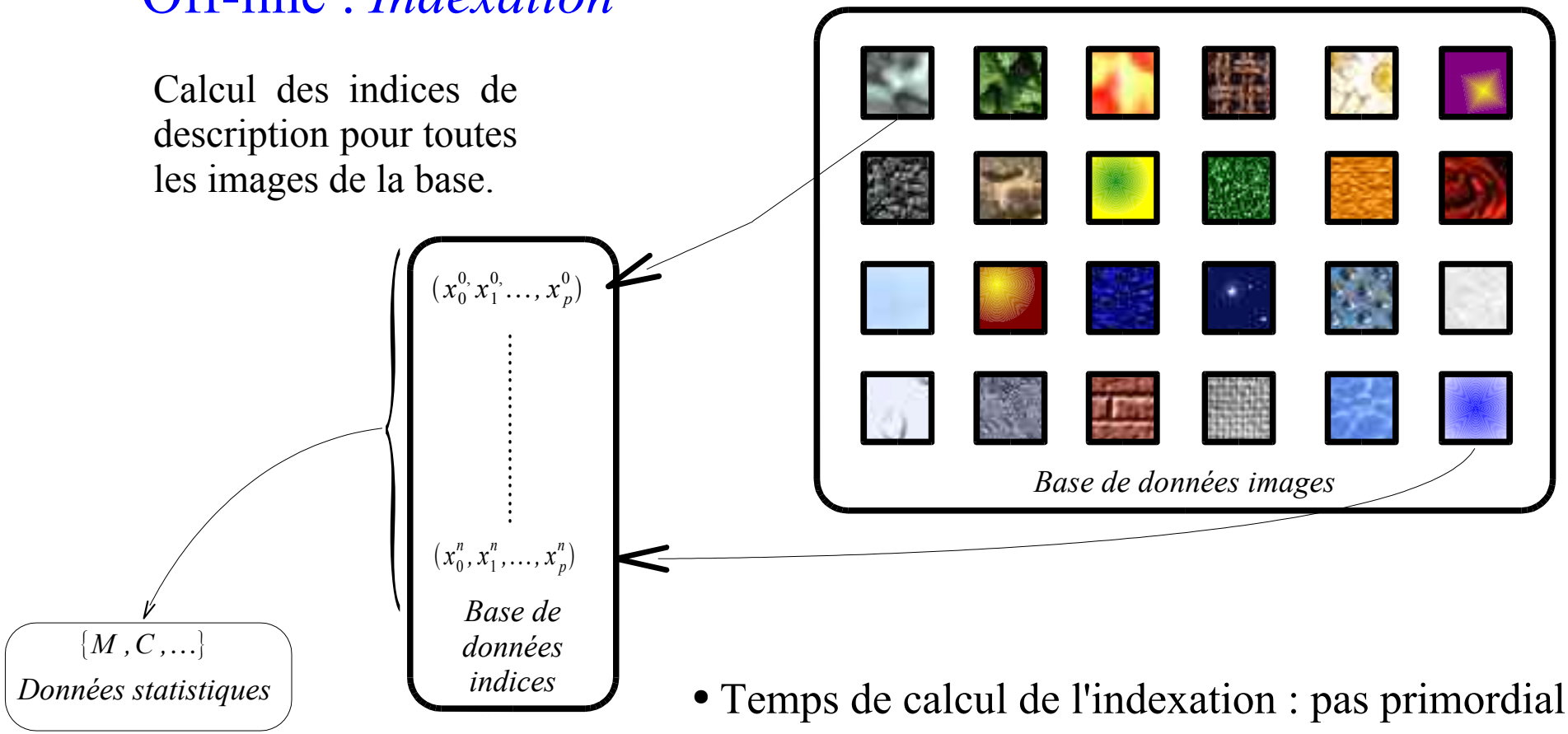
Ex : video cliquable (INRIA)

Interface de l'outil de segmentation video développé à l'INRIA Rhône-Alpes – projet MOVI

Indexation automatique d'images

Off-line : *Indexation*

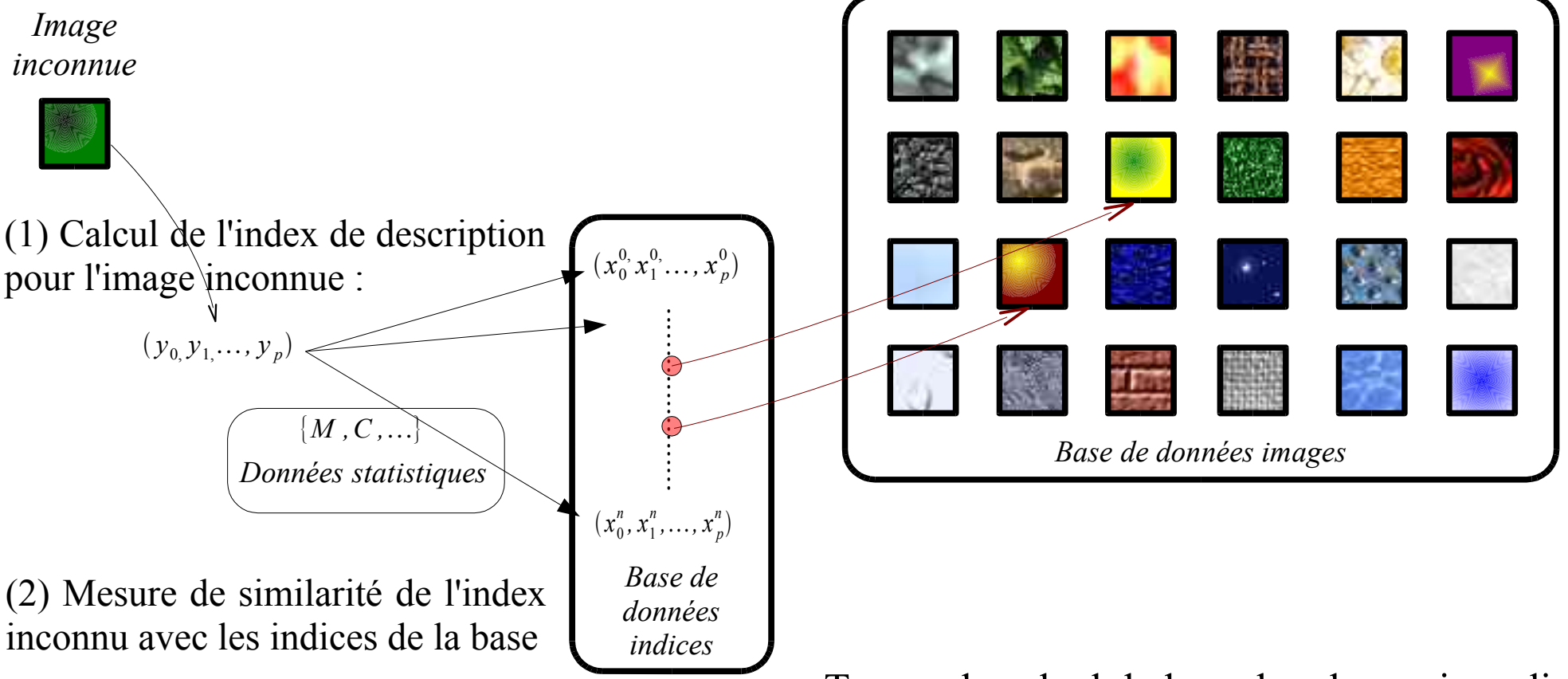
Calcul des indices de description pour toutes les images de la base.



- Temps de calcul de l'indexation : pas primordial
- Indices + données statistiques : calcul incrémental
- Stockage : bases de données images et indices
- Représentation des indices : primordial

Indexation automatique d'images

On-line : *Recherche*



(2) Mesure de similarité de l'index inconnu avec les indices de la base

(3) Résultat : adresse des meilleurs images au sens de la mesure de similarité

- Temps de calcul de la recherche : primordial
- Mesure de similarité : indice de confiance
- Quels descripteurs ?
- Quelles mesures de similarité ?

Indexation automatique d'images

- Requête par un exemple : recherche d'images semblables





- Recherche d'un objet, ou d'un type d'objets particulier




Difficultés :

- Variabilité : rotation, translation, homothétie,...
- Reconnaissance 2d ou 3d
- Visibilité partielle
- Changement de luminosité
- .../...




Appariement d'images structurées

Images comportant des structures géométriques « simples » :
contours rectilignes, elliptiques...

Mise en correspondance de structures 2d

↕

Reconstruction 3d




Techniques de traitement d'images :

- Détection + chaînage de contours
- Détection de formes paramétrées (transformée de Hough)




Indices : listes de coordonnées des structures (segments, ellipses,...)

Métrique d'appariement : basée sur l'appariement des structures
Ex : Calcul de la transformation + Distance de Hausdorff

distance de Haussdorff entre deux ensembles P et Q : $H(P,Q)=max\{h(P,Q),h(Q,P)\}$ *avec :* $h(X,Y)=\max_{x\in X}\ \min_{y\in Y}\ d(x,y)$

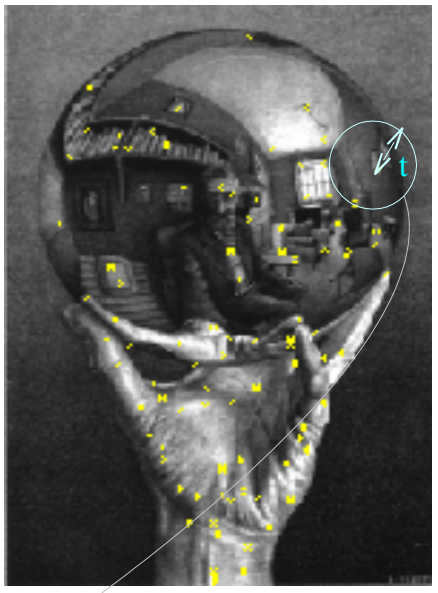
lien avec la morphologie mathématique : $H(P,Q)=min\{\lambda\in\mathbb{R};\delta_{B_\lambda}(P)\subset Q \ \ et \ \ \delta_{B_\lambda}(Q)\subset P\}$ δ_{B_λ} : *dilatation par une boule de rayon* λ

Appariement d'images texturées

Dans ce cas, on ne recherche pas de structures particulières, mais des ressemblances *globales* (histogrammes, spectres de Fourier), ou *locales*...
Si l'on cherche des ressemblances locales, il est essentiel de réduire l'espace de représentation, pour deux raisons majeures :

- réduction du temps de calcul
- augmentation de la robustesse

→ Utilisation des points d'intérêt :



Points d'intérêt (méthode de Harris)

On extrait des *descripteurs locaux* uniquement aux voisinages des points les plus « intéressants ».
Puis on représente le comportement local au voisinage de ces points par les *descripteurs différentiels* :

Jet local : $L^t_{ij} = G^t_{ij} * I$ avec : $G^t_{ij} = \frac{\partial^{i+j}}{\partial x^i \partial y^j} G^t$

et : $G^t(x,y) = \frac{1}{2\pi t^2} \exp\left(\frac{-(x^2+y^2)}{2\,t^2}\right)$ *t : facteur d'échelle*

On notera : $\{L^t_{ij}; 0 \leqslant i+j \leqslant 3\} = \{L, L_x, L_y, L_{xx}, L_{xy}, L_{yy}, L_{xxx}, L_{xxy}, L_{xyy}, L_{yyy}\}$

(dérivées jusqu'au 3e ordre)

Invariants géométriques et photométriques

Le principe de calcul des invariants est de *combiner* les différentes composantes du jet local de manière à obtenir des grandeurs qui soient invariantes à divers changements d'aspect, notamment transformation affines et changement d'illumination.

1 - *Invariance par déplacement*

Invariants différentiels de Hilbert :

→ quantités invariantes par rotation (Notez : invariance par rotation du noyau gaussien)

$$\Psi = \begin{pmatrix} L \\ L_i L_i \\ L_i L_{ij} L_j \\ L_{ii} \\ L_{ij} L_{ij} \\ \varepsilon_{ij}\left(L_{jkl} L_i L_k L_l - L_{jkk} L_i L_l L_l\right) \\ L_{iij} L_j L_k L_k - L_{ijk} L_i L_j L_k \\ -\varepsilon_{ij} L_{jkl} L_i L_k L_l \\ L_{ijk} L_i L_j L_k \end{pmatrix}$$

Avec : $\varepsilon_{xx} = \varepsilon_{yy} = 0$

$\varepsilon_{xy} = -\varepsilon_{yx} = 1$

Notations d'Einstein : sommation sur les indices

Par ex :

$$\Psi_2 = L_i L_{ij} L_j = L_{xx} L_x L_x + 2\ L_x L_{xy} L_y + L_{yy} L_y L_y$$

$$\Psi_7 = -\varepsilon_{ij} L_{jkl} L_i L_k L_l = L_{xxy}\left(-L_x L_x L_x + 2 L_x L_y L_y\right)$$
$$+ L_{xyy}\left(-2 L_x L_x L_y + L_y L_y L_y\right) - L_{yyy} L_x L_y L_y + L_{xxx} L_x L_x L_y$$

Invariants géométriques et photométriques

2 - *Invariance photométrique*

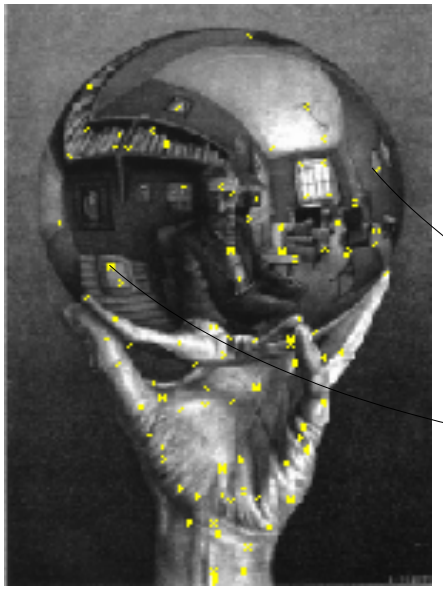
L'objectif est d'être invariant à une modification affine de la fonction d'illumination : $f(I) = aI + b$

→ Normaliser par l'un des invariants (par ex. Ψ_1).

3 - *Invariance par changement d'échelle*

→ Utiliser des invariants à plusieurs échelles.

Un vecteur d'invariants est donc calculé pour chaque point d'intérêt dans toutes les images. Ce sont ces vecteurs qui seront comparés par la suite.



$x = \{x_1, \ldots, x_n\}$

$y = \{y_1, \ldots, y_n\}$

.../...

Métriques d'appariement

Le problème consiste donc à comparer des descripteurs qui sont des vecteurs imprécis :

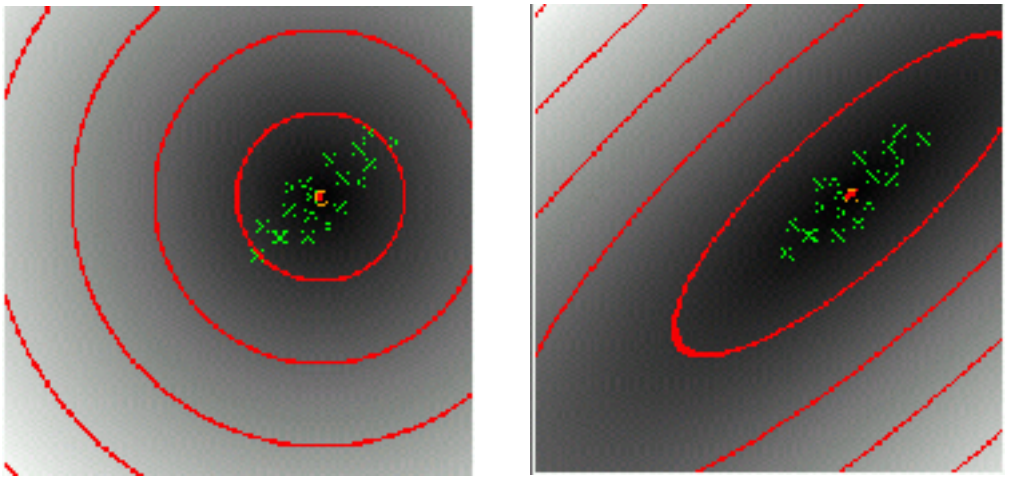
$$x=\{x_1,\ldots,x_n\}$$

Métriques d'appariement :

Distance euclidienne

$$\delta_e(x,x')=\sqrt{{}^t(x-x')(x-x')}$$

La distance euclidienne ne tient compte ni des *différences d'amplitude* ni des *corrélations* entre les différentes composantes du vecteur de description.
On utilise plutôt la distance suivante :



Distance des points de l'espace au centre d'inertie d'un nuage de points, en distance euclidienne (à gauche) et en distance de Mahalanobis (à droite).

Distance de Mahalanobis

$$\delta_m(x,x')=\sqrt{{}^t(x-x')C^{-1}(x-x')}$$

avec : $C=\begin{pmatrix} var(x_1) & cov(x_1,x_2) & \cdots & cov(x_1,x_n) \\ cov(x_2,x_1) & var(x_2) & \cdots & cov(x_2,x_n) \\ \vdots & \vdots & \ddots & \vdots \\ cov(x_n,x_1) & cov(x_n,x_2) & \cdots & var(x_n) \end{pmatrix}$

$$cov(x_i,x_j)=\langle(x_i-\mu_i)(x_j-\mu_j)\rangle=\langle x_i x_j\rangle-\langle x_i\rangle\langle x_j\rangle$$

$$var(x_i)=cov(x_i,x_i) \qquad \mu_i=\langle x_i\rangle$$

...où < . > désigne la moyenne.

Métriques d'appariement

La matrice de covariance C est calculée et mise à jour off-line.
Si on diagonalise C^{-1}, on peut se ramener à un calcul de distance euclidienne par rapport aux vecteurs descripteurs :

$$C^{-1} = {}^{t}P\,D\,P \longrightarrow \sqrt{{}^{t}(x-x')C^{-1}(x-x')} = \left\| \sqrt{D}\,Px - \sqrt{D}\,Px' \right\|$$

normalisation — distance ellipsoïdale

A chaque mise à jour de la base on doit donc :

- mettre à jour la matrice de covariance C.
- calculer et diagonaliser C^{-1}.
- normaliser tous les vecteurs : $x \rightarrow \sqrt{D}\,Px$

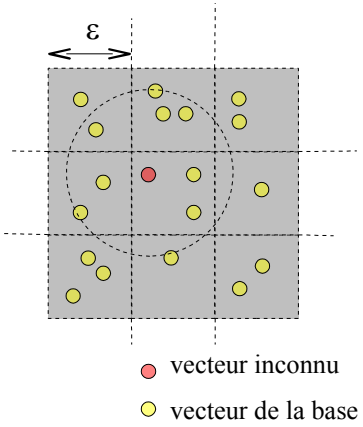
Le problème de la recherche s'exprime maintenant comme suit : étant donné une donnée inconnue de vecteur descriptif x, et un seuil ε, trouver toutes les données de la base dont les vecteurs descriptifs y sont tels que :

$$\delta_m(x,y) = \delta_e(\sqrt{D}\,P\,x, \sqrt{D}\,P\,y) \leqslant \varepsilon$$

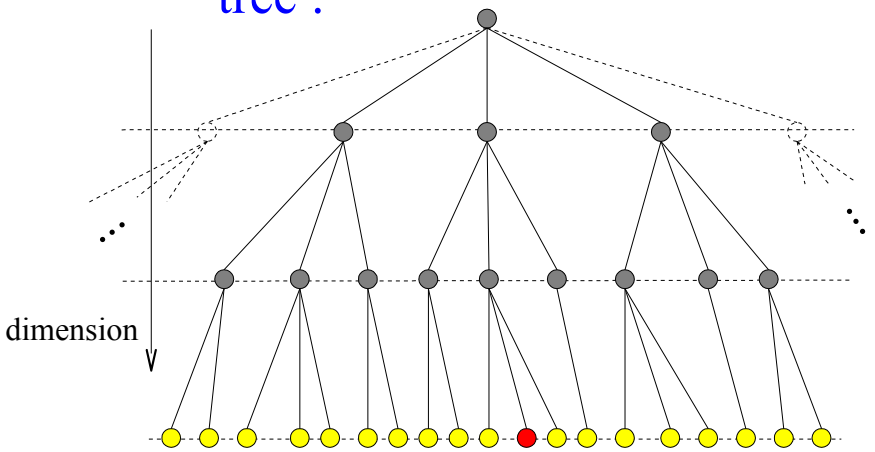
Parcours de l'espace de recherche

Pour limiter le temps de recherche dans une grosse base d'indices, on cherche à limiter la recherche à un certain « voisinage » de l'index inconnu. Ce problème est intimement lié au stockage des vecteurs descriptifs de la base.

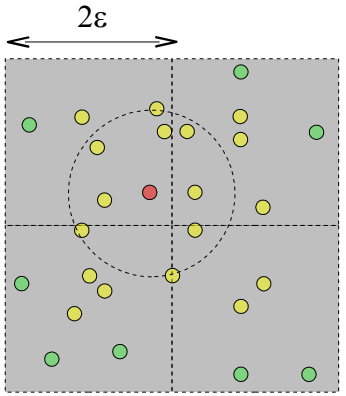
Découpage de la base d'indices en hypercubes :



Représentation de la base d'indices sous forme de Kd-tree :



représentation du voisinage du vecteur inconnu dans un Kd-tree



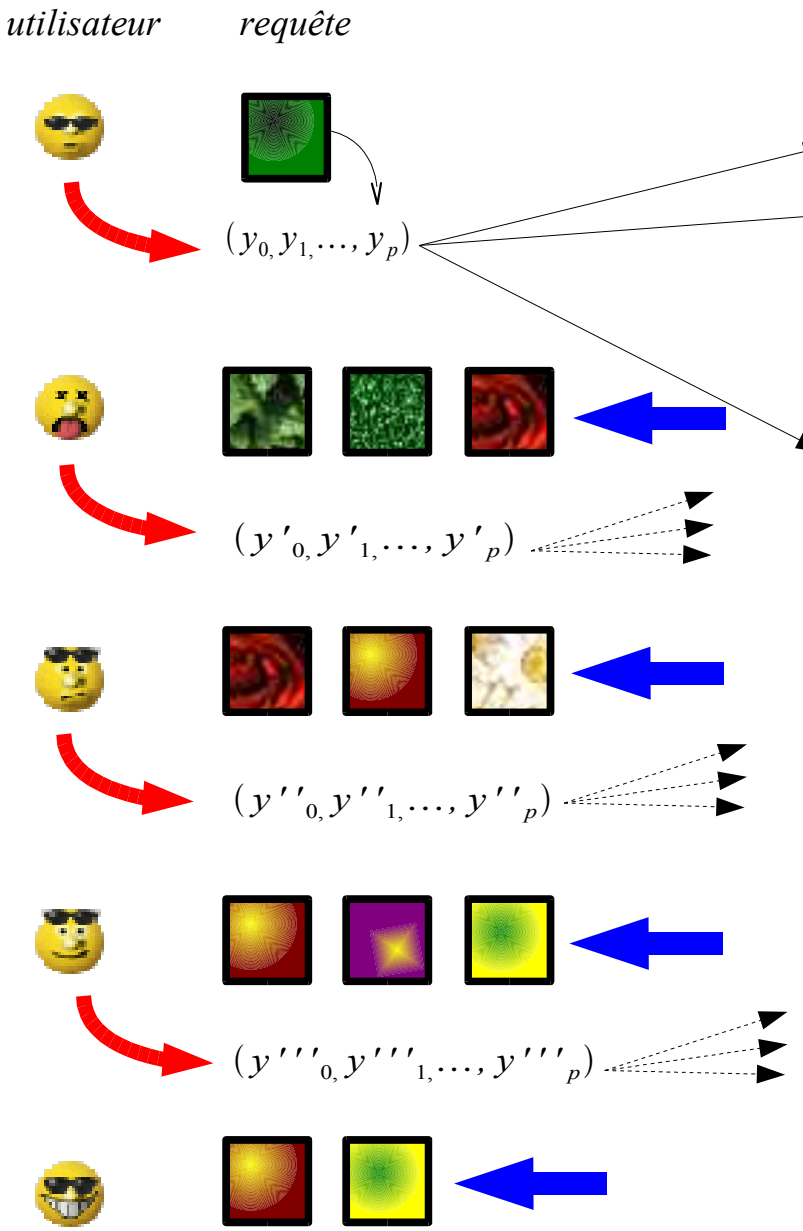
un découpage alternatif de la base d'indices

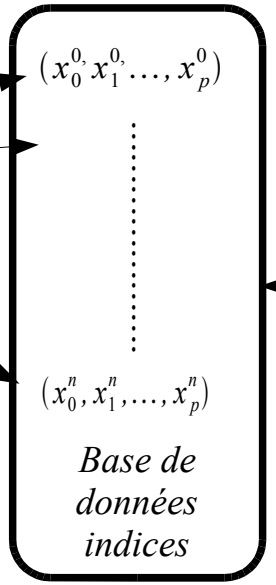
Complexité de la recherche :

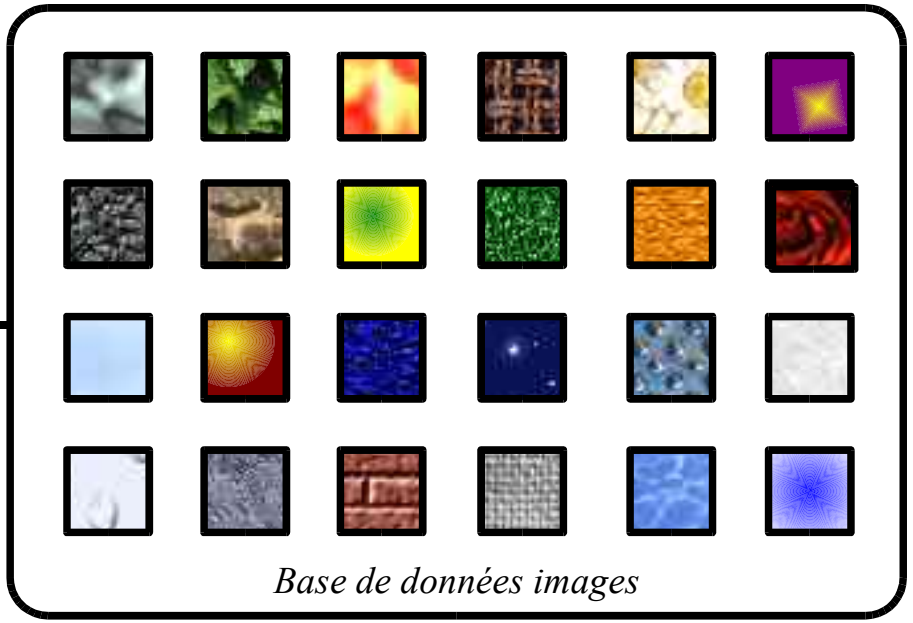
$$\underbrace{\frac{m^2 N 3^d}{k^d}}_{\text{coût de l'appariement}} + \underbrace{m 3^d}_{\text{coût du parcours du Kd-tree}}$$

N = nombre d'images de la base
m = nombre d'invariants par image
k = nombre d'hypercubes par dimension
d = dimension des invariants

Bouclage de pertinence et apprentissage







Le bouclage de pertinence (relevance feedback) est une technique utilisant une évaluation interactive de l'utilisateur pour modifier le descripteur de la requête ou la métrique d'appariement, par exemple en modifiant les poids des composantes, ou bien en ne conservant que les invariants jugés les plus pertinents au regard des évaluations.

Bibliographie et sources

- P. Gros : *Traitement des images par le contenu* - document de cours - IRISA 1999.
- C. Schmid : *Appariement d'images par invariants locaux de niveaux de gris* - thèse de doctorat - INPG 1996.
- J.M. Jolion et al : *Projet Sesame / Rapport final* - INSA 1998
- R.C. Veltkamp, M. Tanase : *Content-based image retrieval : a survey* - Utrecht University

- IRISA / TEXMEX : http://www.irisa.fr/texmex/index.htm
- INRIAlpes / LEAR : http://www.inrialpes.fr/lear/index.html
- INSA Lyon / RFV : http://telesun.insa-lyon.fr/kiwi/
- Univ. Stanford / SIMPLICITY : http://www-db.stanford.edu/IMAGE/
- Univ. Texas / CIRES : http://amazon.ece.utexas.edu/~qasim/research.htm
- .../...