# Indexación de imágenes y vídeos

Antoine Manzanera
ENSTA – ParisTech
Unité d'Électronique et d'Informatique

ENSTA

ParisTech

# Indexación - Introducción

El tema de este curso es la *búsqueda automática* de documentos *visuales* (imágenes, secuencias video), en bases de datos de gran tamaño, a partir de consultas sobre el *contenido* de estos documentos.
Este problema está recibiendo mucha atencion en el área de procesamiento de imágenes y visión artificial. De hecho, el uso generalizado de los medios digitales, la aparición de formatos de vídeo compactos, la disminución del costo de los medios de almacenamiento ha creado un vertiginoso aumento en la cantidad de datos multimedia. Para que estos datos sean utilizables, deben ser vistos de forma eficaz, como a través de un catálogo.
Las técnicas presentadas en este curso, denominadas técnicas de *indexación*, tienen como objetivo de adjuntar a una imagen o un vídeo, una serie de *descriptores del contenido* con el fin de medir la *similitud* con los descriptores correspondiente a la consulta.
Pero esta consulta puede tomar formas muy distintas, puede ser conceptual (por ejemplo, palabra), simbólica (por ejemplo, diagrama) o « instancial » (por ejemplo: otra imagen).
De la misma manera, la indexación sera *semántica* (se adjunta descriptores de nivel conceptual al documento) o *visual* (se adjunta descriptores visuales al documento ).

# Interés y aplicaciones

•*BD Imagenes y videos:*

· Colecciones y catalogos individuales, o de empresas
· Bibliotecas / Mediatecas
· Agencias de fotografia
· Archivos audiovisuales (ej. Francia: INA)
· Internet (ej. AltaVista / Virage / Google / Dailymotion...)

•*Aplicaciones:*

· Mediametria
· Propriedad intelectual / artistica
· Reconocimiento de rostros, de objetos...
· Datos biomedicales
· Imagens aeriales, por satelite...
· Videos de televigilancia

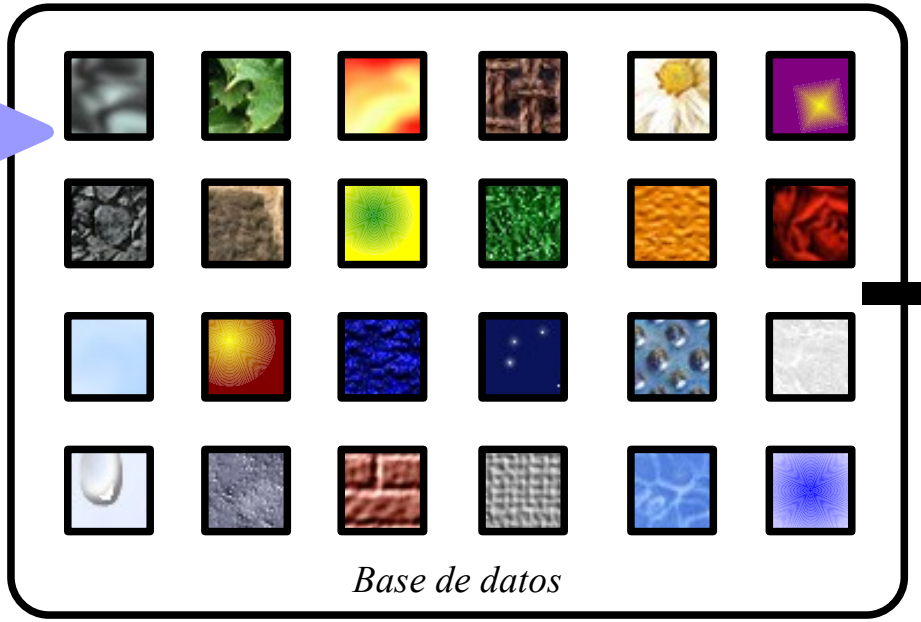
# Indice del curso *Indexacion*

- Indexacion multimedia: Estado actual y Perspectivas
  - Busqueda de documentos multimedia por el contenido
  - Indexacion semantica manual
  - Indexacion visual automatica
- Ayuda a la indexacion manual
  - Semantica de la indexacion video
  - Cortada en planos / secuencias
  - Detection de objetos
- Indexacion automatica
  - Imagenes estructuradas y texturadas
  - Extraccions de los descriptores
  - Apareamiento de imagenes
  - Metricas de similitud

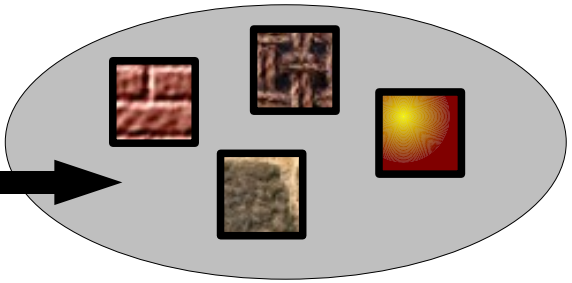
# Busqueda multimedia por el contenido

Consulta

•*Tipos de consulta:*
- Palabra
- Imagen
- Dibujo, esquema
- Modelo CAD
- Mapa
- Plano
- .../...




Resultado



Evaluacion de los resultados :

$$Precision = \frac{\#\text{ respuestas relevantes}}{\#\text{ respuestas}}$$

$$Recuerdo = \frac{\#\text{ respuestas relevantes}}{\#\text{ datos relevantes}}$$

•Dificultades:
- A diferencia de los datos textuales, el contenido semantico nunca es explicito.
- Las consultas son dificiles expresar, entonces a medudo ambiguas, incompletas.

Dimension multidisciplinaria:
Multimedia : texto, imagen, sonido - Gestion de bases de datos - Hardware - Aprendizaje - Lingüistica,...

# Estado actual: Indexacion explicita

La mayoria de las herramientas de búsqueda de documentos multimedia que funcionan actualmente se basan en una búsqueda de palabras claves *explícitamente relacionadas* con el documento o *indexadas automáticamente a partir del texto que lo rodea* (Ej: Google).
Los documentos vídeo como archivos audiovisuales *se indexan manualmente* por operadores especializados, a partir de una descripción muy exacta vinculada con un tesauro.
Pero esta indexación manual es una tarea ardua y larga (hasta 10 veces la duracion de la secuencia, mientras que por ejemplo el fondo de documentos de la televisión del INA almacena mas de 350.000 horas de programación...)
También datos interesantes en una fecha determinada no lo eran necesariamente en la fecha de indexación...

También se debe mencionar la aparición de nuevas normas, como la codificación de vídeo Mpeg7 que incluye en el codigo datos explícitos sobre los contenidos audiovisuales, con el fin de facilitar la busqueda de información en una base de datos de vídeo, y permitir la navegación « inteligente » en un vídeo.

# Indexaciones semantica y descriptiva

•Por naturaleza, la indexación *manual* es *semántica*. El operador de indexación atribuye al documento datos de alto nivel sobre la *significacion* del contenido. Las consultas asociadas son generalmente *palabras* designando un *objeto*, un *acción*, el nombre de un *personaje* o *acontecimiento*.

•Por el contrario, la indexación *automática* es esencialmente *descriptiva* o *visual*. El algoritmo de indexación atribuye datos de *bajo nivel* semántico, sobre el contenido *geométrico*, o *espectral*, de la imagen, a un nivel *local* o *global*. Las consultas suelen ser realizadas por ejemplo, o por modelo.

•Pero el análisis automático de documentos también se puede utilizar para hacer más fácil (más rápido, menos arduo) el trabajo del operador de indexación manual. Esto es típicamente:

- Pre-selección de grandes bases de datos de imágenes.
- Indexación semi-automática con interaccion operador/computadora.
- División de vídeo en planos y simplificación en « imágenes claves ».

# Semantica de las secuencias de imagenes

Antes de analizar manualmente o automáticamente una secuencia de imágenes, se debe definir con precisión la forma en que el vídeo *se estructura*. La estructuracion clásica de un vídeo es un corte en *escenas* con el título, resumen, palabras claves.
Las herramientas para ayudar a la indexación de vídeo se hacen sobre la base de una estructuracion precisa de videos, utilizando varios niveles de análisis.
El primer objetivo es proporcionar un marco *rico y riguroso* para facilitar la indexación manual.
El segundo objetivo es *reducir el nivel semántico* de corte en escenas para permitir el uso de herramienta de indexación visual automática.

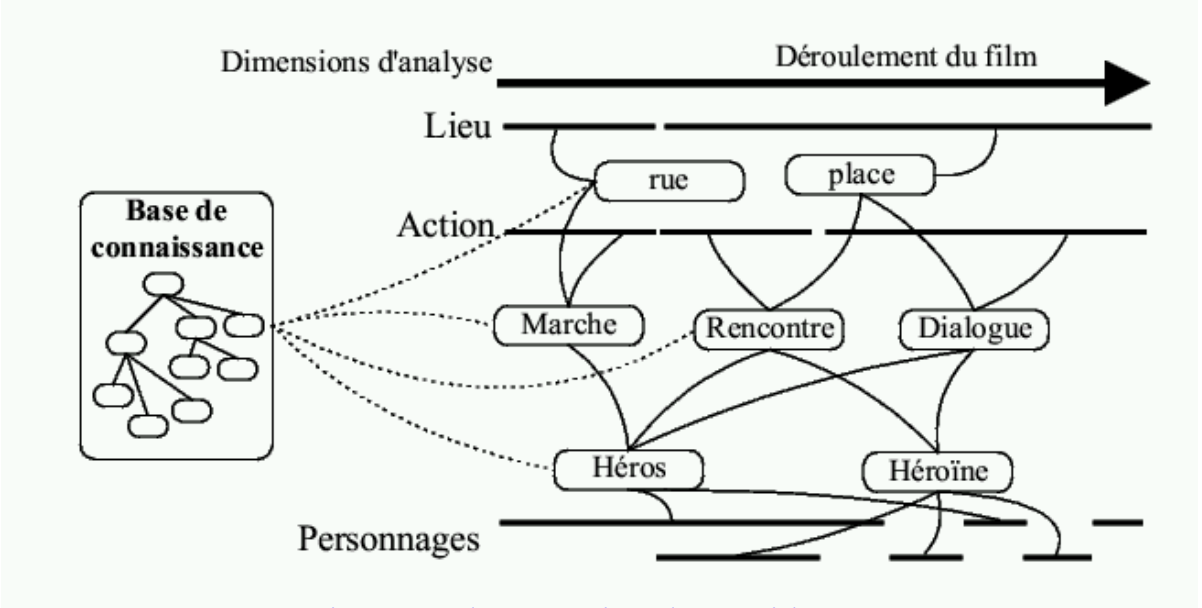



Diagrama de anotacion de un video
(Projet Sesame – Insa Lyon / LIRIS)

Escena:

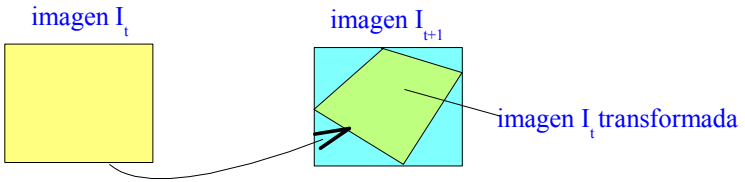
- Naturaleza del sitio
- Presencia de un objeto, de un personaje
- Tipo de plano
- Tipo de accion
- .../...

# Ayuda a la indexación video

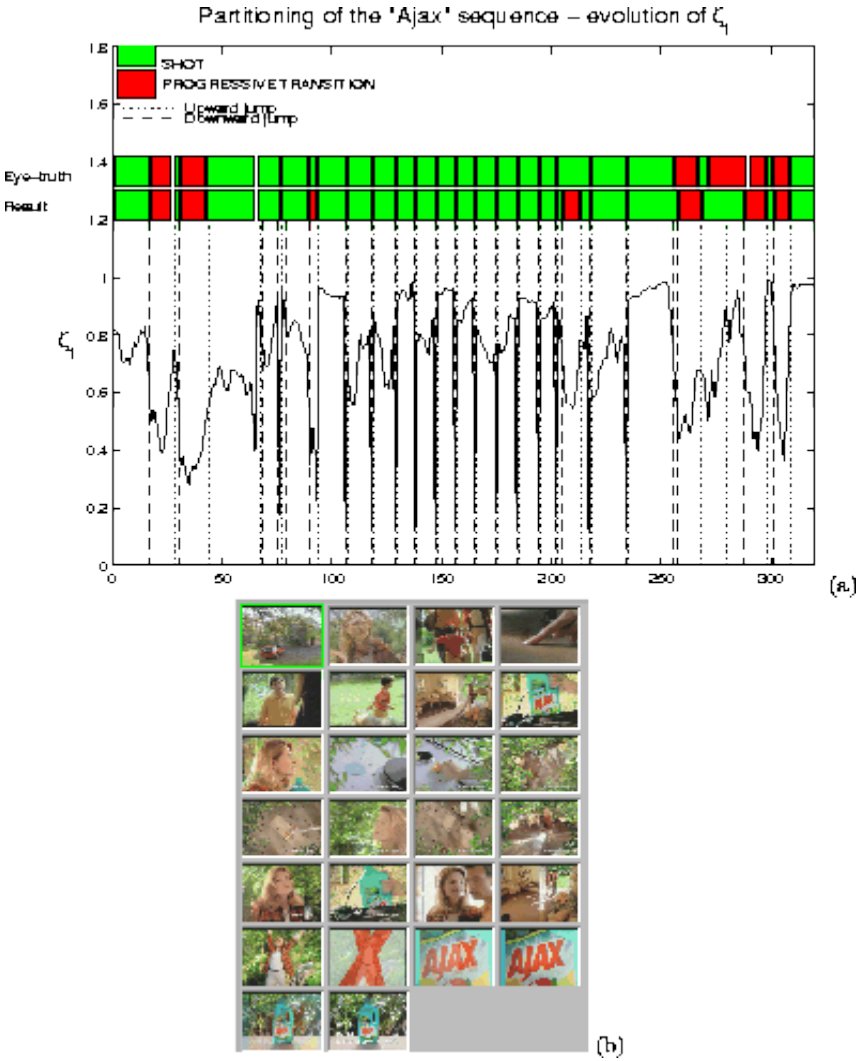
*Ejemplo*: Corte en planos (cuts) de una video

Las técnicas utilizadas son variadas ; se basan generalmente en la detección de discontinuidades temporales de uno o más descriptores globales relacionados con:

* El *color*. Ej: momentos de histogramas color.
* El *movimiento*. Ej. a la derecha: extracción del movimiento dominante (transformación afín 2d), y medida de la tasa de recubrimiento entre la imagen inicial y la imagen transformada.




•Difficultades:
- Transicion por sobreimpresion,...
- Movimientos bruscos,...




Software *MD-shots* (IRISA Rennes / projet VISTA) de cortada de video, basado sobre un descriptor global del movimiento dominante (eje vertical)

# Ayuda a la indexación video

Una técnica que acompaña a menudo a la segmentación en planos para la ayuda a la indexación de video, es la extracción de *imágenes claves* en cada plano, es decir, las imágenes "más representativas" del plano. Las técnicas utilizadas actualmente se basan generalmente en estadísticas relacionadas con los descriptores utilizados para cortar en planos. Así, se puede utilizar la imagen mediana, las imágenes extremas...

Por último, algunas técnicas de detección, reconocimiento, o identificación se utilizan para realizar tareas específicas para ayudar a la indexación. Entre ellos:

* La detección y la persecución de objetos móviles.

* La detección de objetos particulares: rostros, vehículos, texto incorporado, etc, para identificar el tipo de escena

* Identificación: el rostro de un personaje, un cierto tipo de vehículo,...

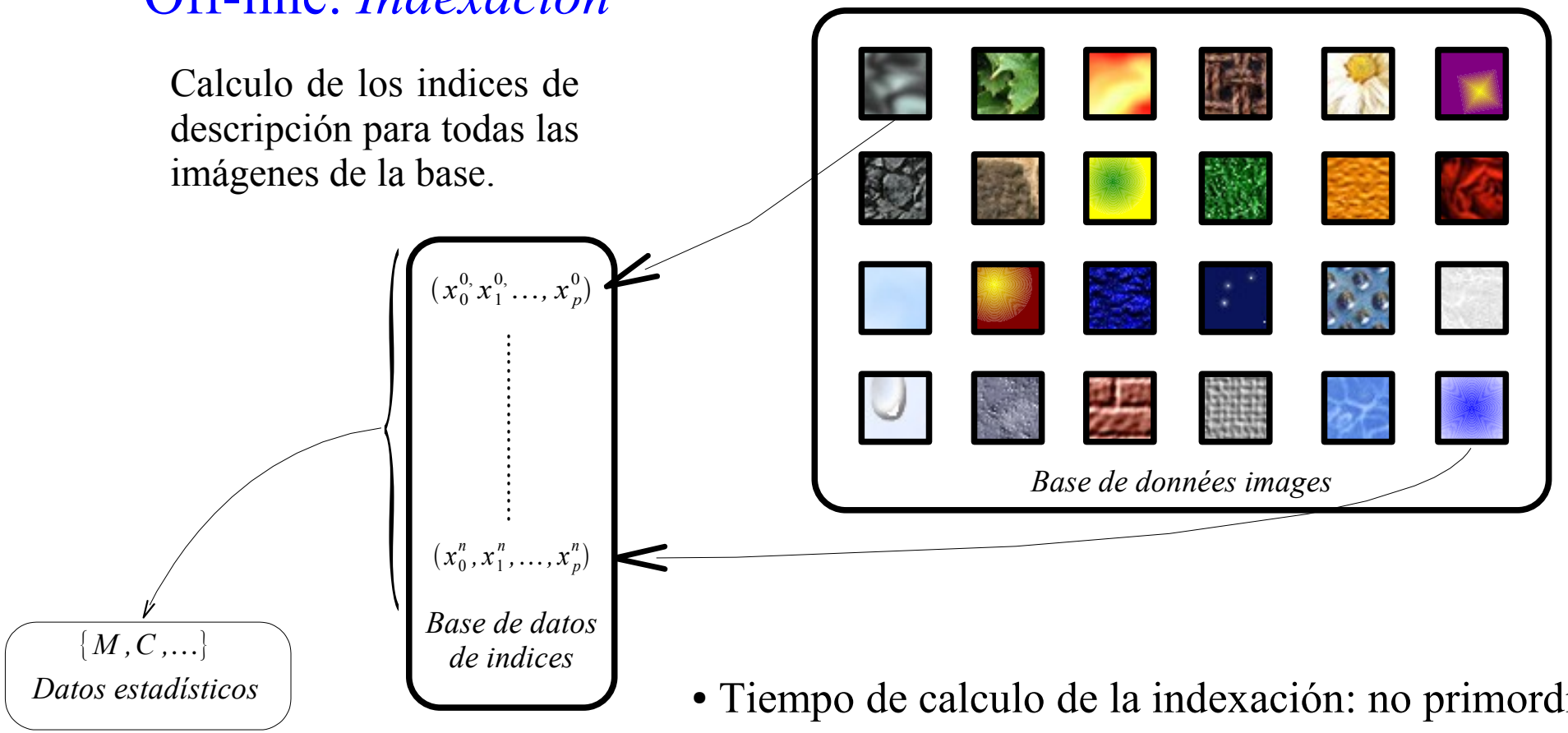
# Ej: video “clicable” (INRIA)


Interface de la herramienta de segmentación video desarrollada en la INRIA Rhône-Alpes – projet MOVI

# Indexación automática de imágenes

## Off-line: *Indexación*

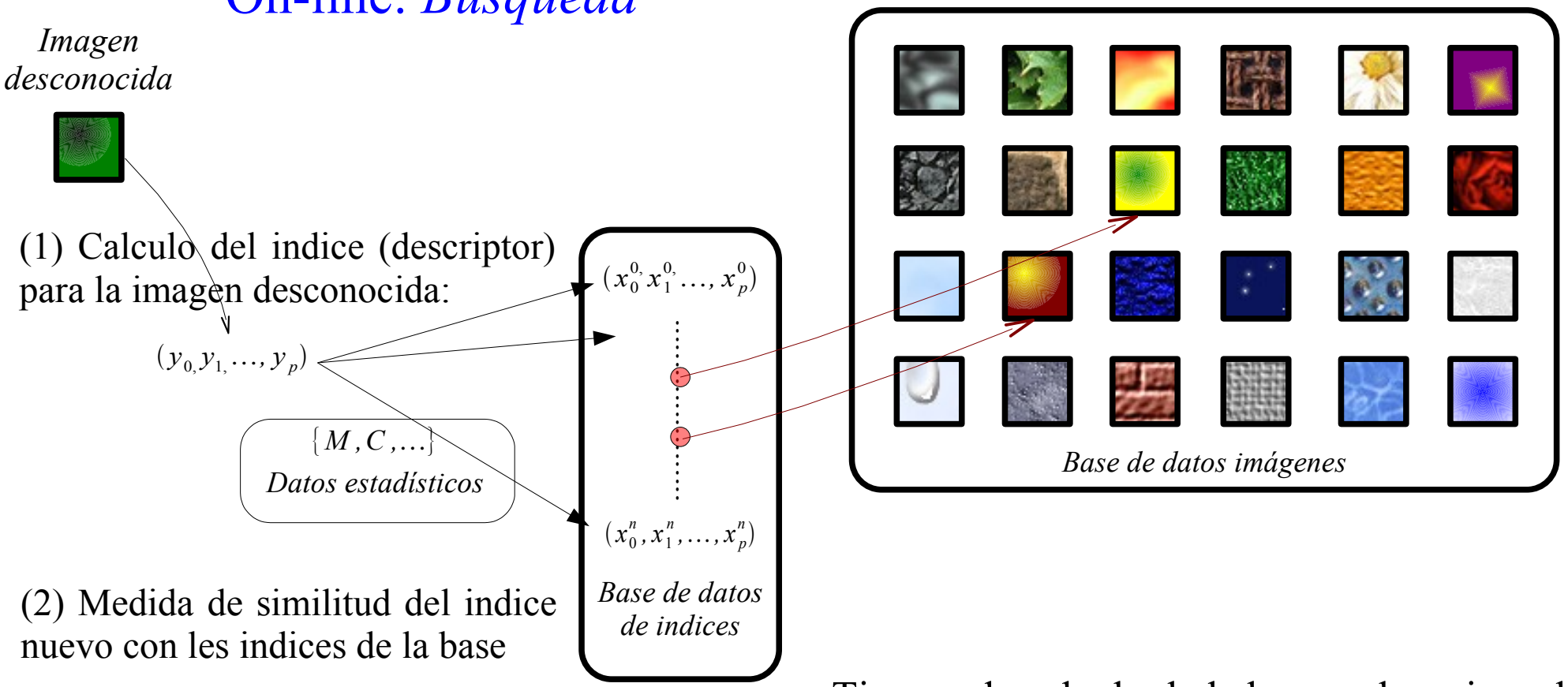
Calculo de los indices de descripción para todas las imágenes de la base.




- Tiempo de calculo de la indexación: no primordial
- Indices + datos estadísticos: calculo incremental
- Almacenamiento: bases de datos imágenes y indices
- Representación de los indices: primordial

# Indexación automática de imágenes

## On-line: *Búsqueda*



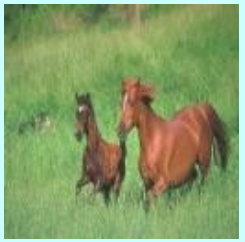

(2) Medida de similitud del indice nuevo con les indices de la base

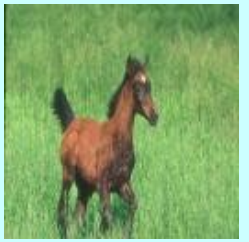
(3) Resultado: referencia de las mejores imágenes en el sentido de la medida de similitud

- Tiempo de calculo de la busqueda: primordial
- Medida de similitud: indice de confianza
- ¿ Cuales descriptores ?
- ¿ Cuales medidas de similitud ?

# Indexación automática de imágenes

• Consulta por el ejemplo: búsqueda de imágenes semejantes

• Búsqueda de un objeto, o de un cierto tipo de objetos




•*Dificultades:*

• Variabilidad: rotación, traslación, homotecia,...
• Reconocimiento 2d o 3d
• Visibilidad parcial
• Cambio de luminosidad
•.../...




# Apareamiento de imagenes estructuradas

Imagenes con estructuras geométricas “simples”:
contornos rectos, redondos...

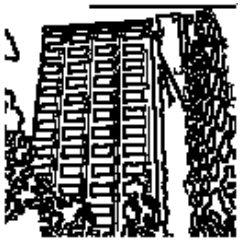
Correspondencia entre estructuras 2d

↕

Reconstruccion 3d





- Técnicas de processamiento:
- Deteccion + cerradura de contornos
- Deteccion de formas parametradas (transformada de Hough)

*Indices*: listas de coordenadas de las estructuras (segmentos, elipses,...)

Metrica de apareamiento: basada sobre la correspondencia de las estructuras
Ej.: Calculo de la transformacion + Distancia de Hausdorff

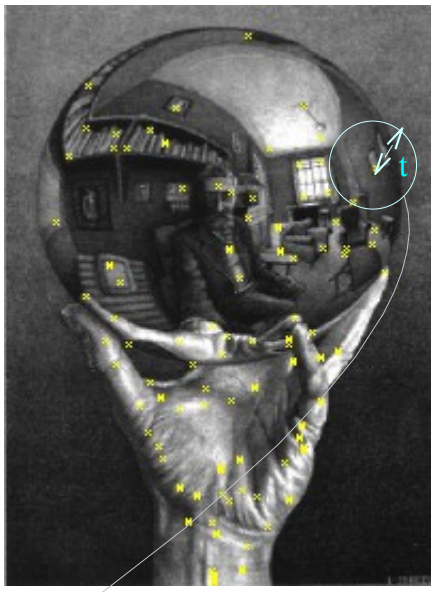
*distancia de Haussdorff entre dos conjuntos P y Q :* $H(P,Q)=max\{h(P,Q),h(Q,P)\}$ *con:* $h(X,Y)=\max_{x\in X}\min_{y\in Y} d(x,y)$

*Relacion con morfologia matematica:* $H(P,Q)=min\{\lambda\in\mathbb{R}\,;\delta_{B_\lambda}(P)\subset Q \;\; et \;\; \delta_{B_\lambda}(Q)\subset P\}$ $\delta_{B_\lambda}$: *dilatacion por una bola de radio* $\lambda$

# Apareamiento de imágenes texturadas

•En este caso, no se busca estructuras particulares, sino similitudes *globales* (histogramas, espectros de Fourier), o *locales*...
•Si se busca similitudes locales, es esencial reducir el espacio de representación, por dos razones:
• reducción del tiempo de calculo
• aumento de la robusteza

⟶Utilización de los puntos de interés:




*Puntos de interés (método de Harris)*

Se computa *descriptores locales* únicamente en la vecindad de los puntos mas "interesantes".
Después, se representa el comportamiento local en la vecindad de los puntos por los *descriptores diferenciales*:

*"Local* jet": $L^t_{ij} = G^t_{ij} * I$ con: $G^t_{ij} = \frac{\partial^{i+j}}{\partial x^i \partial y^j} G^t$

*t : factor de escala*

y: $G^t(x,y) = \frac{1}{2\pi t^2} \exp\left(\frac{-(x^2+y^2)}{2t^2}\right)$

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

(derivadas hasta el 3o orden)

# Invariantes geométricos y fotométricos

El principio del calculo de los invariantes es *combinar* las diferentes componentes del "local jet" de manera a tener valores que sean invariantes a varios cambios de aspecto, incluso transformacion afines y cambio de iluminacion.

1 - *Invariancia por desplazamiento*

Invariantes diferenciales de Hilbert:

⟶ cuantidades invariantes por rotacion (NB: invariancia por rotacion del nucleo gausiano)

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

Con: $\varepsilon_{xx} = \varepsilon_{yy} = 0$
$\varepsilon_{xy} = -\varepsilon_{yx} = 1$

Notationes de Einstein: sumacion sobre los indices

Por ej.:

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

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

# Invariantes geométricos y fotométricos

2 - *Invariancia fotométrica*

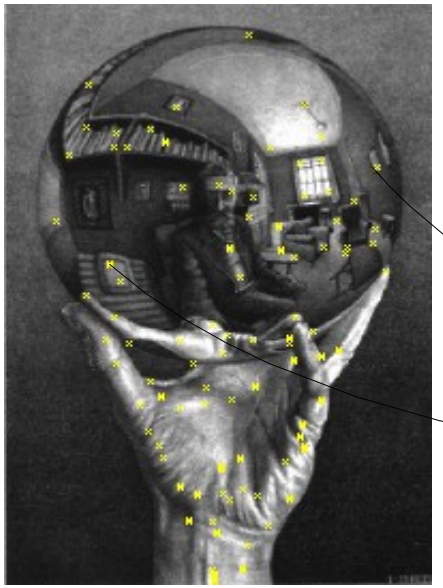
El objetivo es ser invariante a una modificacion afin de la funcion de iluminacion: $f(I) = aI + b$

→ Normalizar por uno de los invariantes (por ej. $\Psi_1$).

3 - *Invariancia por cambio de escala*

→ Usar invariantes en varias escalas.

Entonces un vector de invariantes esta calculado para cada punto de interés en todas las imagenes. Son estos vectores que seran comparados posteriormente.



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

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

.../...

# Métricas de apareamiento

El problema consiste en comparar descriptores que son vectores poco precisos:

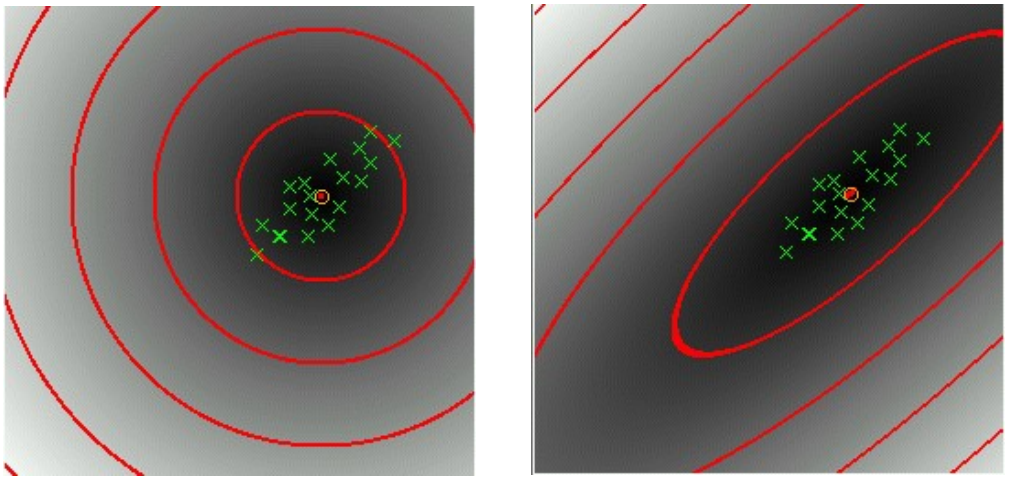
$$x=\{x_1,\ldots,x_n\}$$

## *Métricas de aparaemiento:*

### *Distancia euclidiana*

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

La distancia euclidiana no toma en cuento las *diferencias de amplitud* ni de las *correlaciones* entre las diferentes componentes del vector de descripcion.
Ma bien se utiliza la distancia siguiente:



*Distancia de los puntos del espacio al centro de masas (media) de una nube de puntos, en distancia euclidiana (izq.) y en distancia de Mahalanobis (der.).*

### *Distancia de Mahalanobis*

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

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

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

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

...donde < . > denota la media.

# Métricas de apareamiento

La matriz de covarianza $C$ es calculada y actualizada "off-line".
Diagonalizando $C^{-1}$, se puede volver a un calculo de distancia euclidiana por respecto a los vectores descriptores "normalizados":

$$C^{-1} = {}^{t}P\,D\,P \longrightarrow \sqrt{{}^{t}(x-x')\,C^{-1}\,(x-x')} = \left\| \underbrace{\sqrt{D}}_{\text{normalizacion}}\,\underbrace{Px}_{\text{distancia elipsoidal}} - \sqrt{D}\,Px' \right\|$$

A cada actualizacion de la base se debe, entonces:

- actualizar la matriz de covariancia $C$.
- calcular y diagonalizar $C^{-1}$.
- normalizar todos los vectores: $x \to \sqrt{D}\,Px$

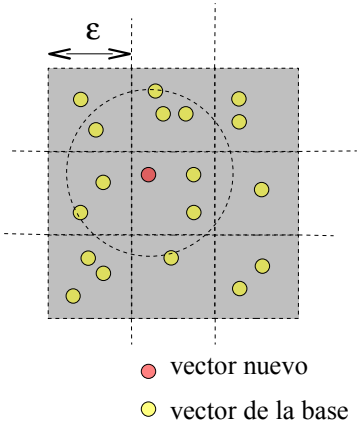
El problema de la busqueda puede ahora definirse asi: considerando un dato desconocido de vector descriptivo $x$, y un umbral $\varepsilon$, hallar todos los datos de la base cuyos vectores descriptivos $y$ son tales como:

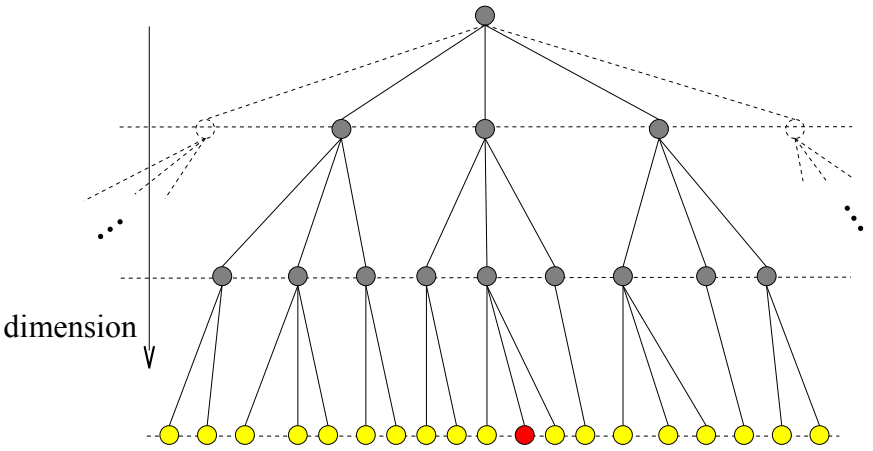
$$\delta_m(x, y) = \delta_e(\sqrt{D}\,P\,x, \sqrt{D}\,P\,y) \leqslant \varepsilon$$

# Exploracion del espacio de busqueda

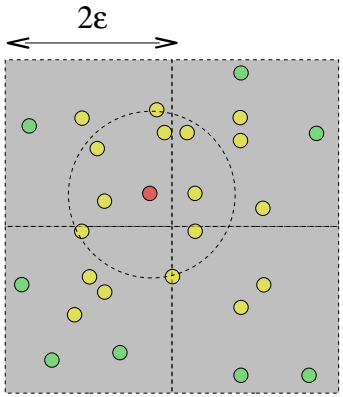
Para limitar el tiempo de busqueda en una base grande de indices, se trata limitar la busqueda a una cierta “vecindad” del indice desconocido. Este problema es intimamente relacionado al almacenamiento de los vectores descriptivos de la base.

*Cortada de la base de indices en hipercubos:* ⟷ *Representacion de la base de indices con un* Kd-tree:




*representacion de la vecindad del vector nuevo en un Kd-tree*



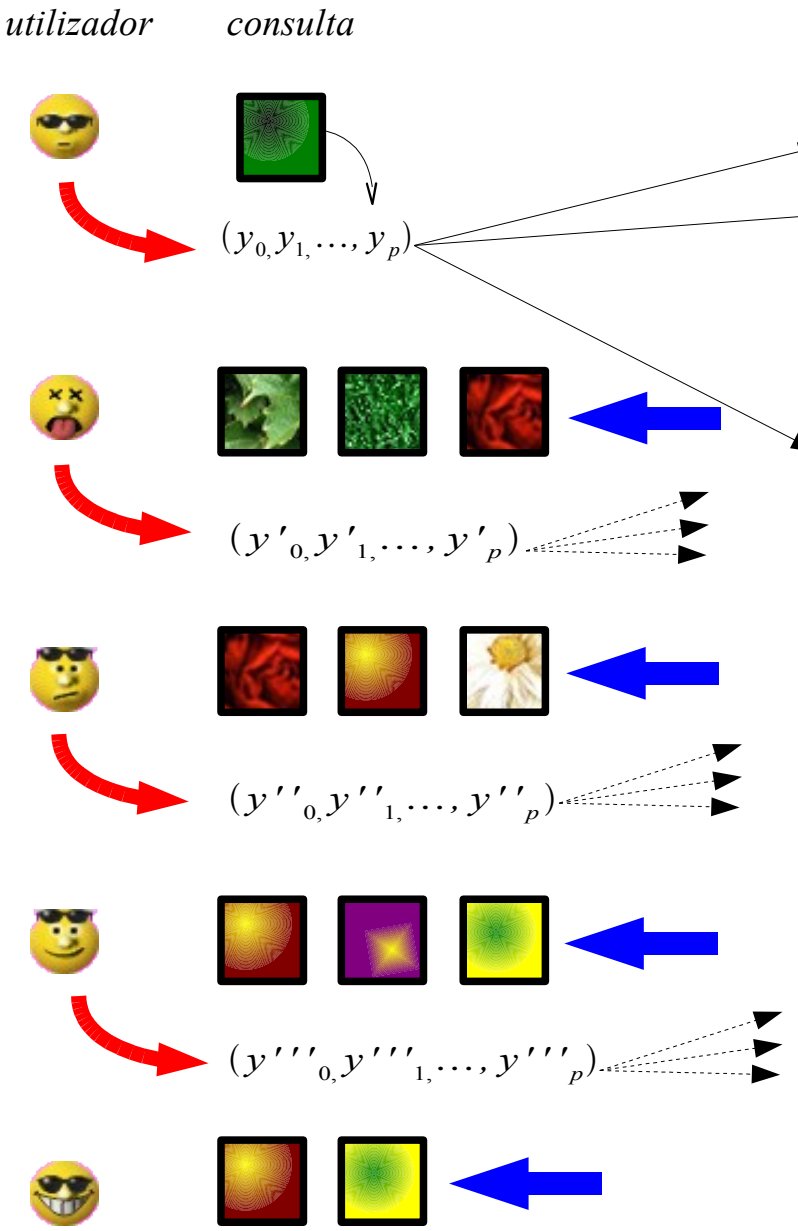

*Otra cortada posible de la base de indices*

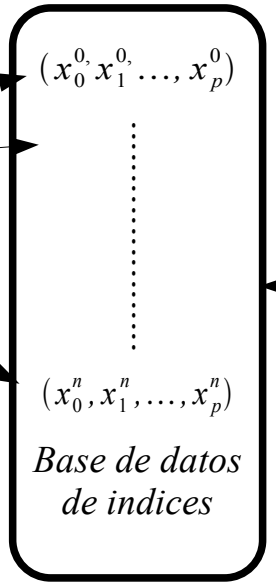
*Complexidad de la busqueda:*

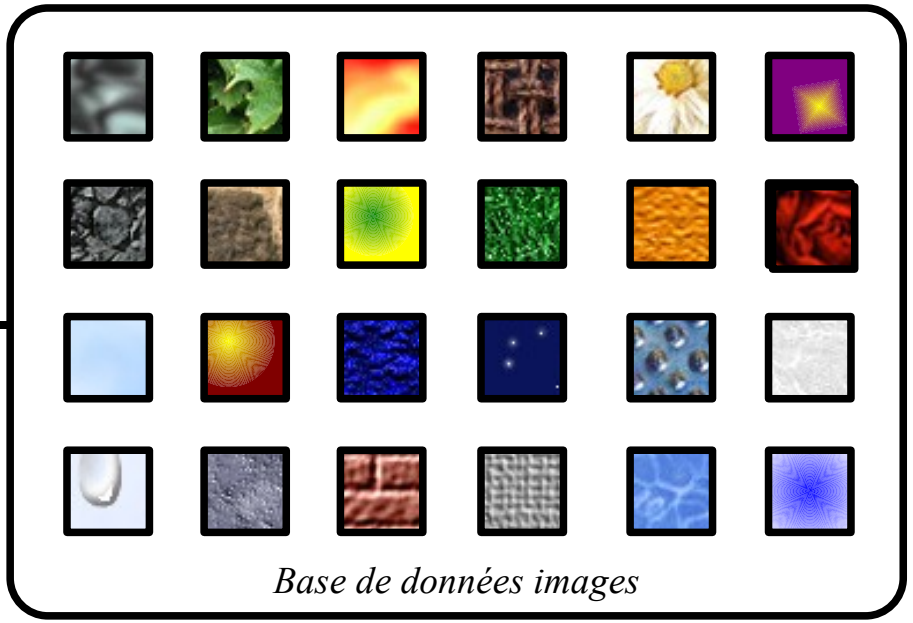
$$\underbrace{\frac{m^2 N 3^d}{k^d}}_{\text{costo del apareamiento}} + \underbrace{m 3^d}_{\text{Costo de la exploracion del Kd-tree}}$$

$N$ = numero de imagenes de la base
$m$ = numero de invariantes por imagen
$k$ = numero de hipercubos por dimension
$d$ = dimension de los invariantes

# “Relevance feedback” y aprendizaje




El bucle de busqueda (relevance feedback) es una técnica utilizando una evaluacion interactiva con el utilizador para modificar el descriptor de la consulta o la métrica de apareamiento, por ejemplo modificando los pesos de las componentes, o guardando solamente los invariantes juzgados mas pertinentes respecto a las evaluaciones.

# Bibliografia y fuentes


• P. Gros : *Traitement des images par le contenu* - documento de curso - IRISA 1999.

• C. Schmid : *Appariement d'images par invariants locaux de niveaux de gris* - Tésis de doctorado - 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
- .../...