Recherche d'images par le contenu Application au monitoring Télévisuel à l'institut national de l'audiovisuel

Alexis Joly
alexis.joly@inria.fr

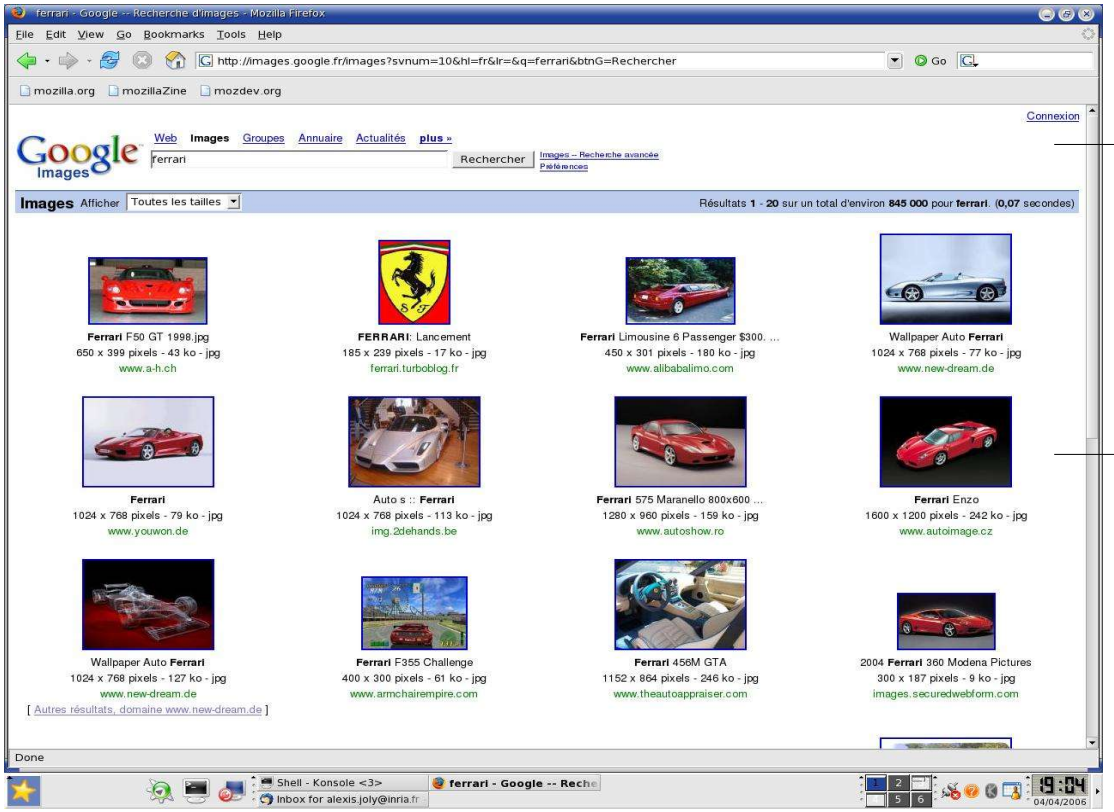
INRIA - IMEDIA

Plan de l'exposé

- Recherche d'images par le contenu
  - Principe des CBIR
  - Descripteurs et similarité
  - Démo et applications
- Monitoring Télévisuel
  - L'INA et les bases gérées à l'INA
  - Principes et architecture
  - Résultats et exploitation

CBIR: Intro

- Recherche d'images par requête textuelle

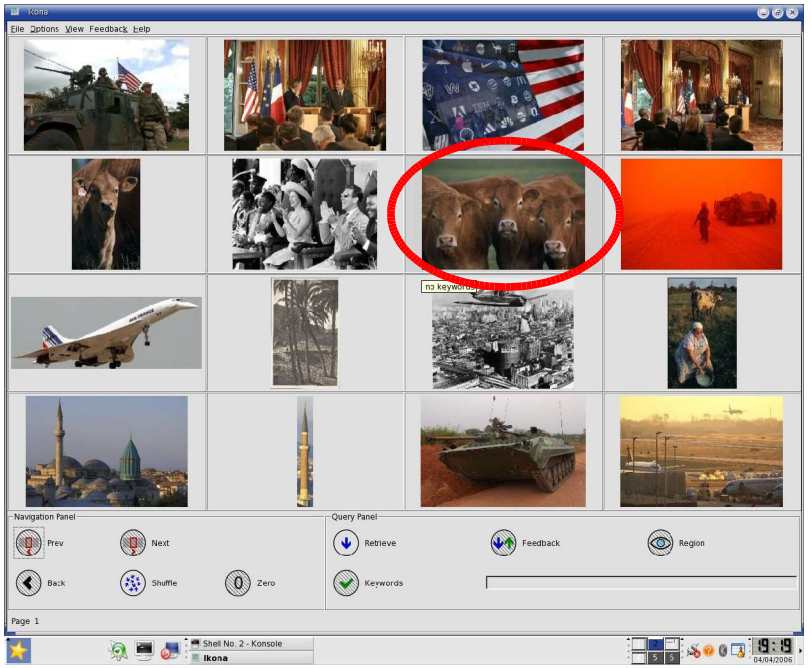


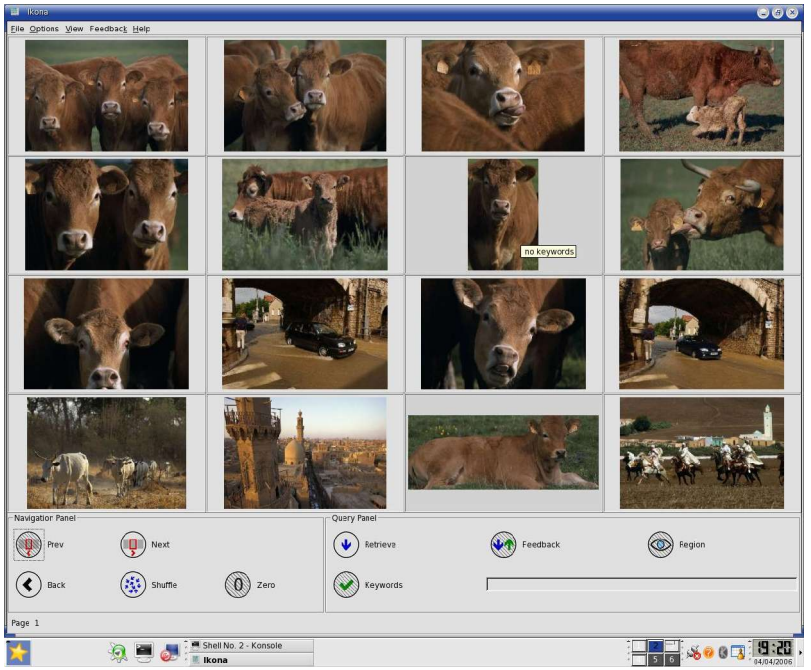
Ferrari Enzo
1600 x 1200 pixels - 242 ko - jpg
www.autoimage.cz

- Annotations textuelles

CBIR: Intro

- Recherche d'images par similarité visuelle





- Utilisation du contenu de l'image uniquement

CBIR: Principe

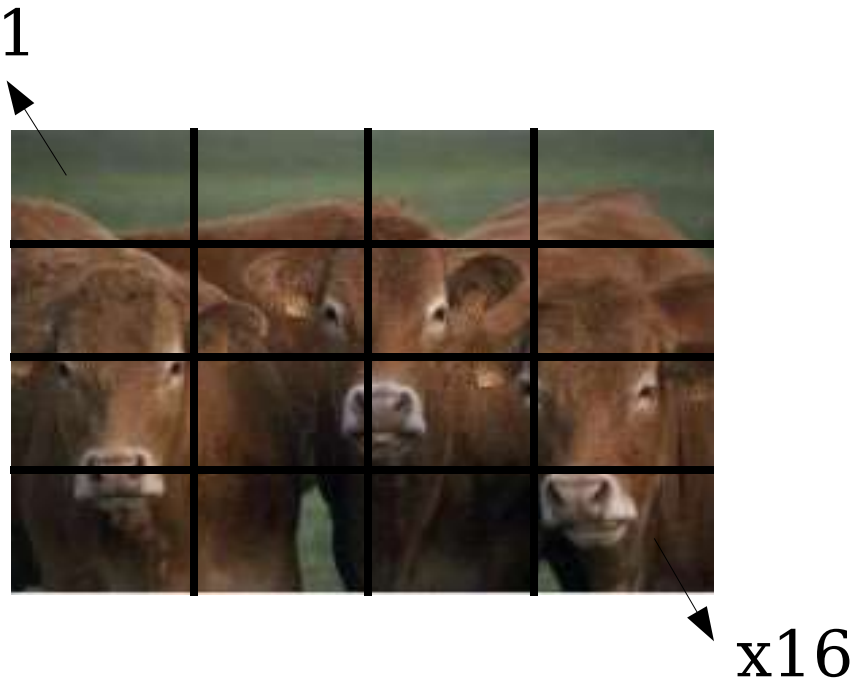
- Descripteurs d'image
- Vecteur de dimension D (signature)
- Caractéristiques spécifiques de l'image

$\longleftrightarrow \quad X(I)=(x_1, x_2, \ldots, x_D)$

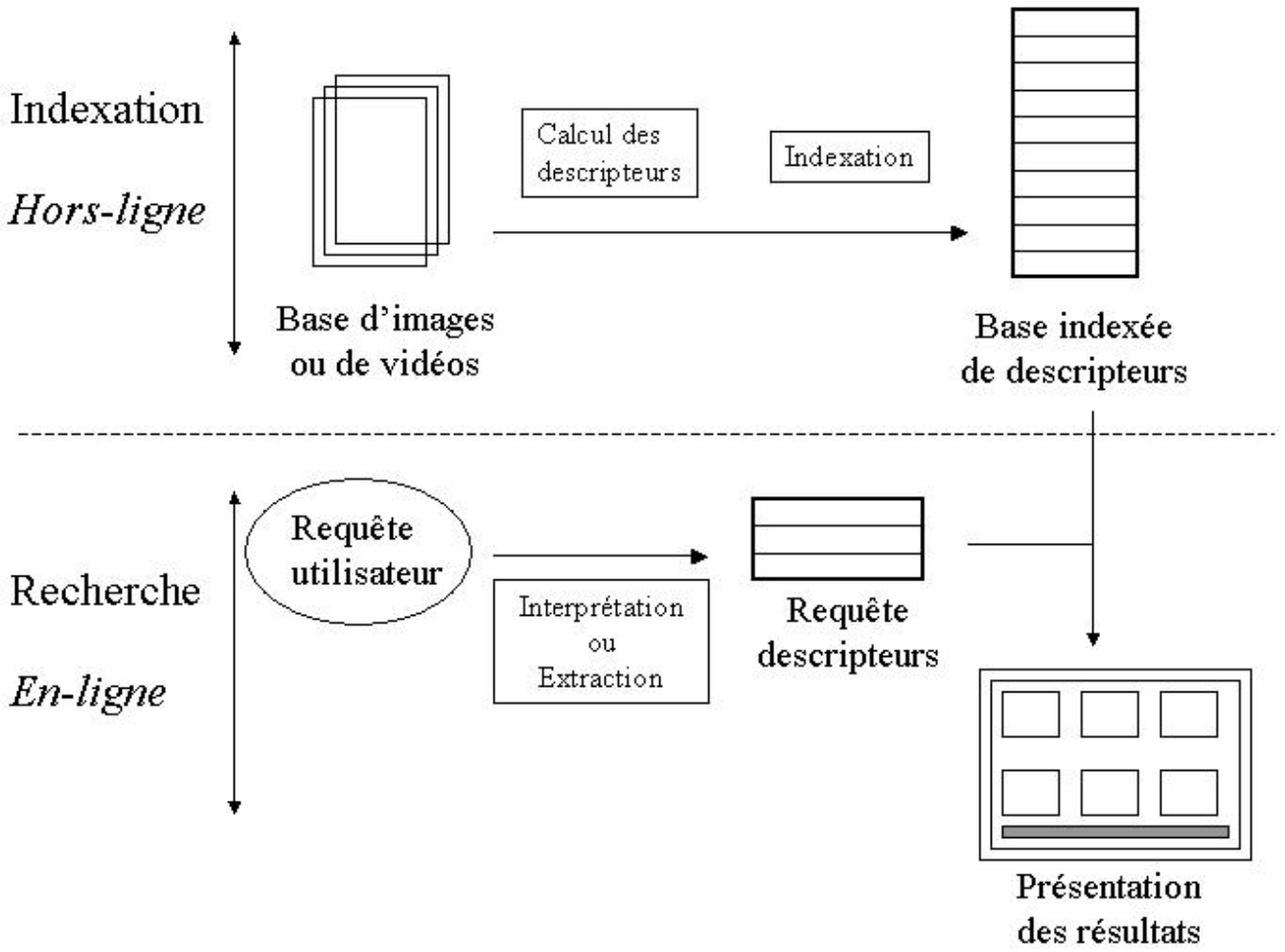
- Compression

CBIR: Principe

- Exemple d'un descripteur simple
  - Moyenne de l'intensité sur les blocs d'une grille D=16



CBIR: Principe

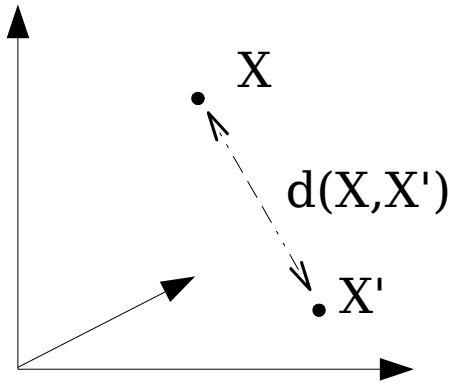


CBIR: Principe

- Similarité visuelle entre deux images
  - Soit X=X(I) et X'=X(I') les descripteurs de deux images I et I'
  - Mesure de similarité entre descripteurs

$$d(X, X')$$

- Cas où d(.,.) est la distance euclidienne



CBIR: Descripteurs

- 3 grandes familles de descripteur
  - Descripteurs de couleur
  - Descripteurs de texture
  - Descripteurs de forme



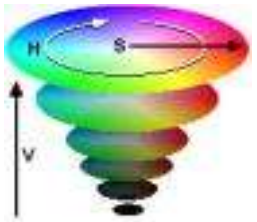
CBIR: Descripteurs de couleur

- Histogrammes couleurs
  - Espace couleur RGB $I(u,v)=\begin{pmatrix} I_r(u,v) \\ I_g(u,v) \\ I_b(u,v) \end{pmatrix}$

  - Histogramme D=64
    - Partition de l'espace RGB en 4x4x4=64 blocs (64 “intervalles de couleurs”)
    - Comptage du nombre de pixels pour chacun des 64 “intervalles de couleurs”
    - Normalisation: $X(I)=\frac{(x_1,x_2,\ldots,x_D)}{\|(x_1,x_2,\ldots,x_D)\|}$

CBIR: Descripteurs de couleur

- D'autres espaces couleurs:
  - (Y,U,V) = standard de télévision
  - **(H,S,V)** =
    - **H**ue = teinte (orange, bleu)
    - **S**aturation = saturation (vif, pastel)
    - **V**alue = intensité (sombre,clair)



- Histogramme dans HSV donne une meilleur similarité visuelle

CBIR: Descripteurs de couleur

- Histogrammes couleurs pondérés
  - Par des attributs de forme
    - Poids maxi sur les contours (gradient de l'image)

Pondérée par

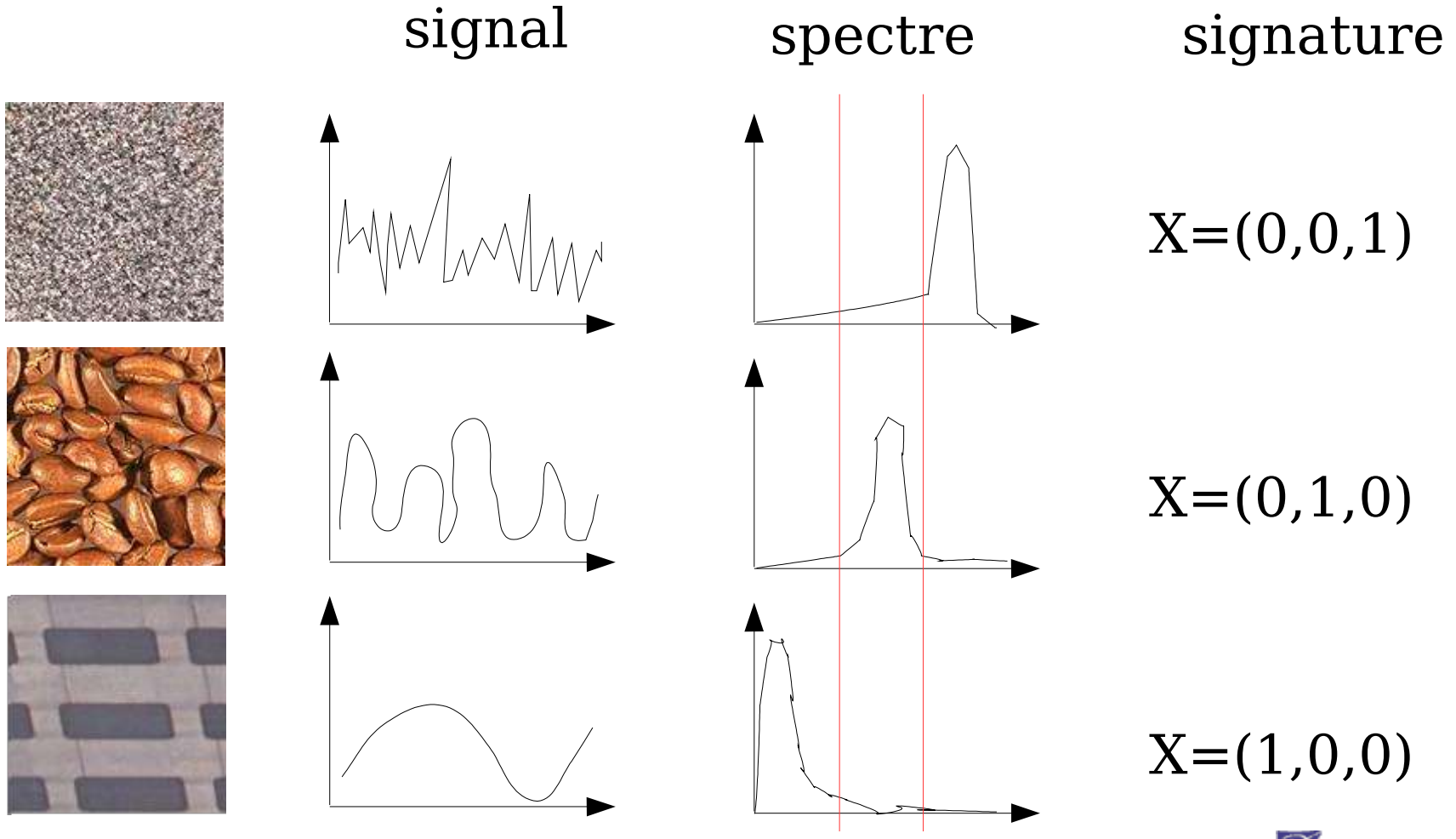
$$\|\vec{\nabla} I\| = \sqrt{\left(\frac{dI}{du}\right)^2 + \left(\frac{dI}{dv}\right)^2}$$

  - Par des attributs de texture

Pondérée par

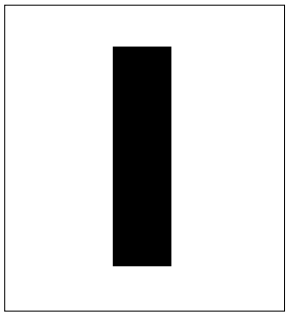
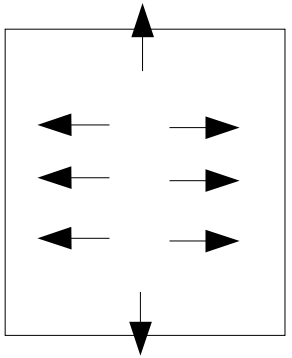
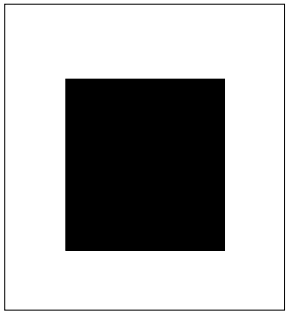
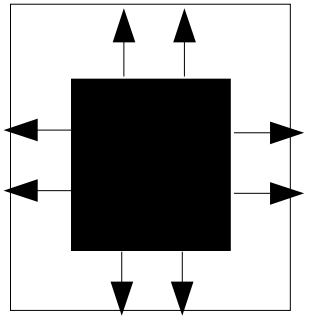
CBIR: Descripteurs de texture

- Contenu fréquentiel de l'image (Fourier,etc.)



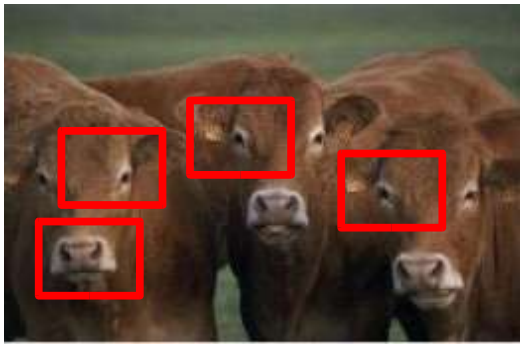
CBIR: Descripteurs de forme

- Histogramme de l'orientation du gradient

image	gradient	signature
		X=(1,3,3,1)
		X=(2,2,2,2)

CBIR: Descripteurs locaux

- Image caractérisée par un ensemble de M descripteurs locaux (dimension D)
- Un descripteur local caractérise le contenu local d'une image



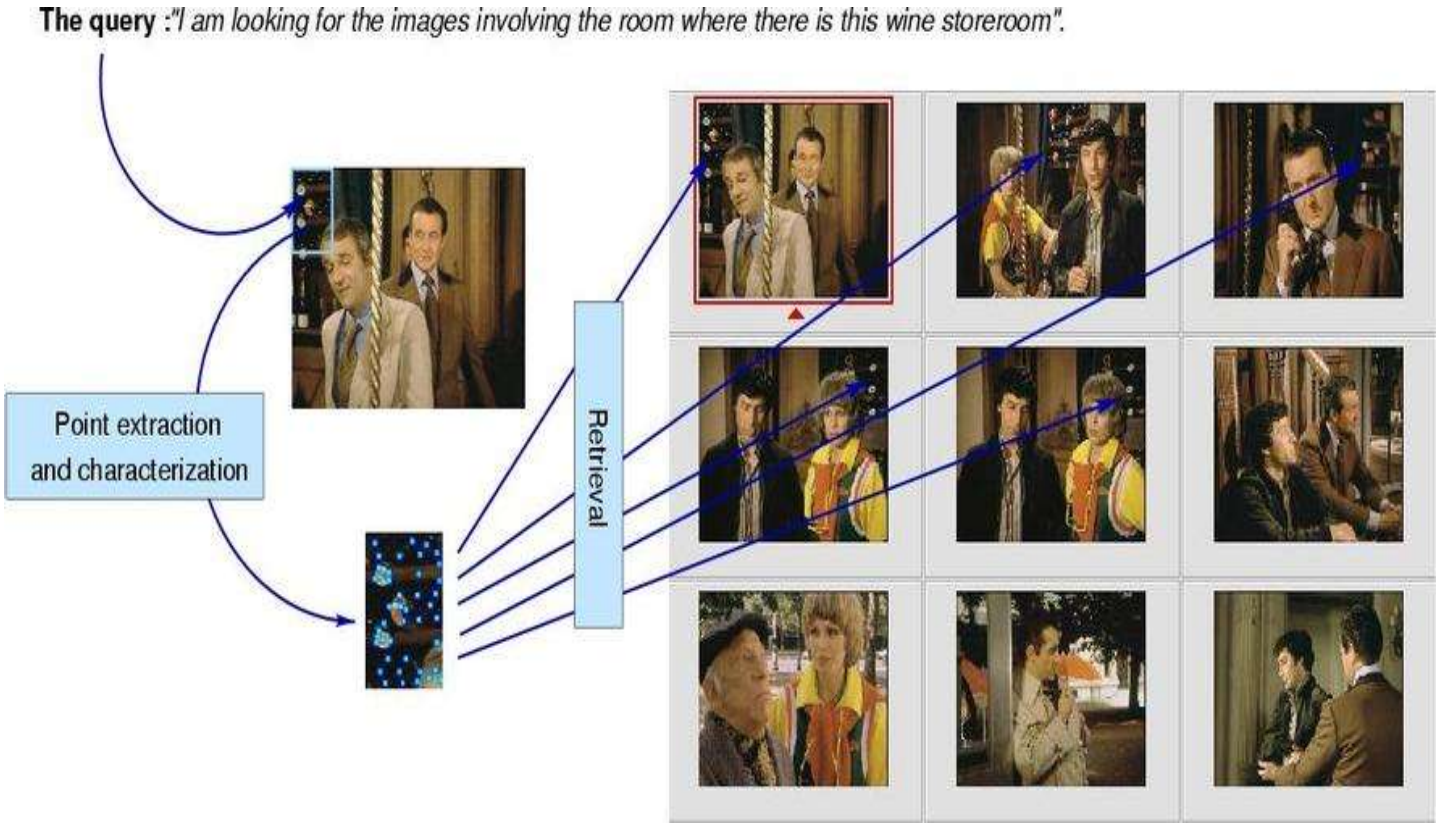
M=4
D=3

X1=(1,3,3,9)
X2=(7,7,2,6)
X3=(7,7,1,6)
X4=(6,6,2,7)

- Scénarios de recherche différents

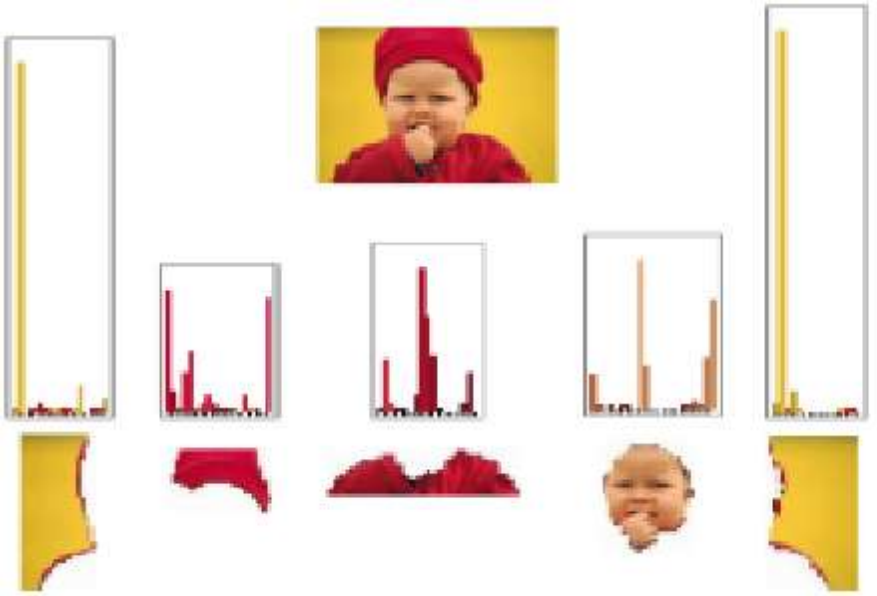
CBIR: Descripteurs locaux

- Requêtes partielles
- Recherche de parties d'image



CBIR: Descripteurs locaux

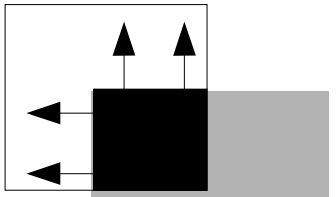
- Régions
  - Détecteur de régions grossières
  - Descripteur fin de chaque région par un histogramme adapatatif



CBIR: Descripteurs locaux

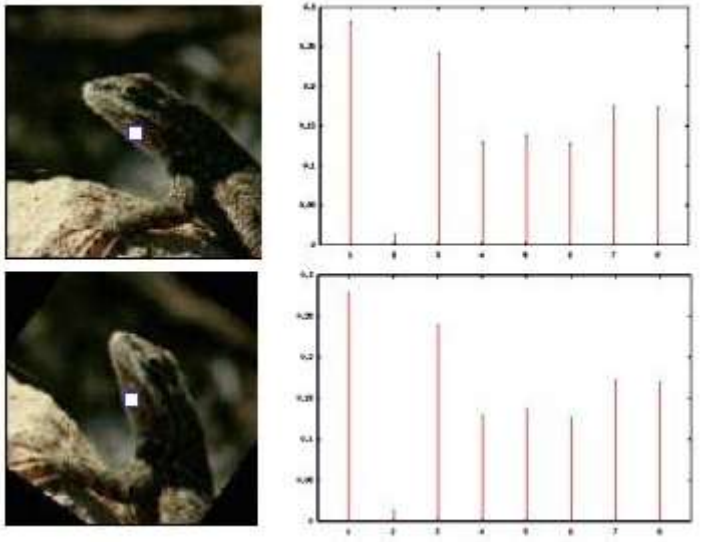
- Points d'intérêt
  - Détecteurs (Harris, DoG)

  - Harris = filtre maximum sur les coins (fortes dérivées en x et en y)



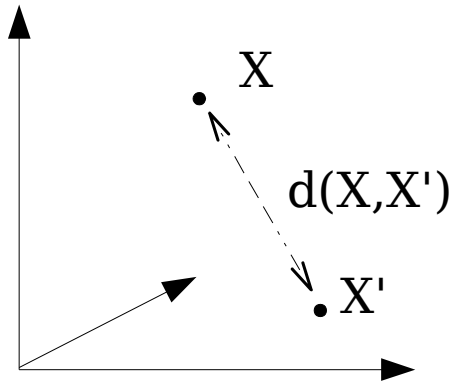
CBIR: Descripteurs locaux

- Points d'intérêt
  - Decripteurs (différentiels, SIFT, moments invariants, filtres orientés,etc.)
  - Ex: $X_i = \left(R, \|\nabla R\|^2, G, \|\nabla G\|^2, B, \|\nabla B\|^2, \nabla R.\nabla G, \nabla R.\nabla B\right)$



CBIR: Mesures de similarité

- Similarité entre deux images = une mesure de similarité entre descripteurs
- L2: $d(X,X') = \sqrt{(x_1 - x'_1)^2 + (x_2 - x'_2)^2 + \ldots + (x_D - x'_D)^2}$

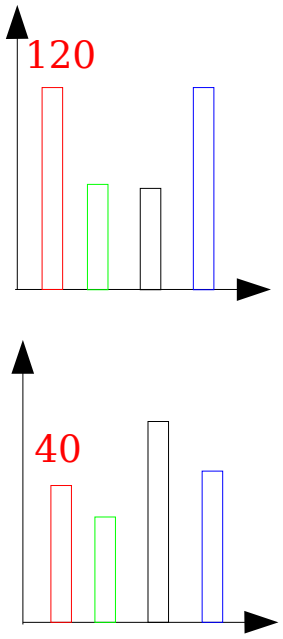


- L1: $d(X,X') = |x_1 - x'_1| + |x_2 - x'_2| + \ldots + |x_D - x'_D|$
- Lmax: $d(X,X') = \max(|x_1 - x'_1|, |x_2 - x'_2|, \ldots, |x_D - x'_D|)$

CBIR: Mesures de similarité

- L'intersection d'histogrammes

$$d(X,X') = 1 - \frac{\min(x_1, x'_1) + \min(x_2, x'_2) + \ldots + \min(x_D, x'_D)}{\min(x_1 + x_2 + \ldots + x_D, x'_1 + x'_2 + \ldots + x'_D)}$$



Min(120,40)=40 → Au moins 40 pixels rouges dans chacune des deux images

\+ Au moins 35 pixels verts dans chacune des deux images

\+ Au moins 34 pixels noirs dans chacune des deux images

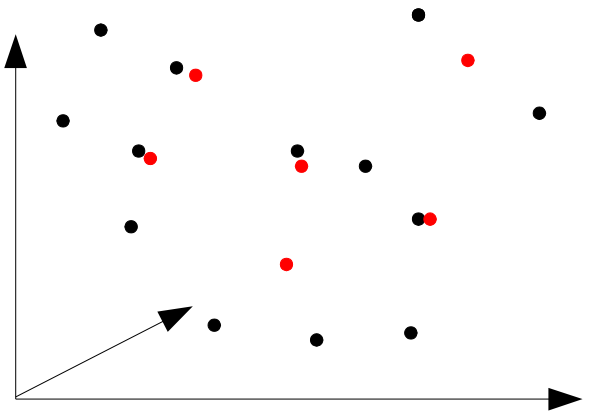
\+ Au moins 50 pixels noirs dans chacune des deux images

---

159 pixels ayant une couleur commune dans les deux images

CBIR: Mesures de similarité

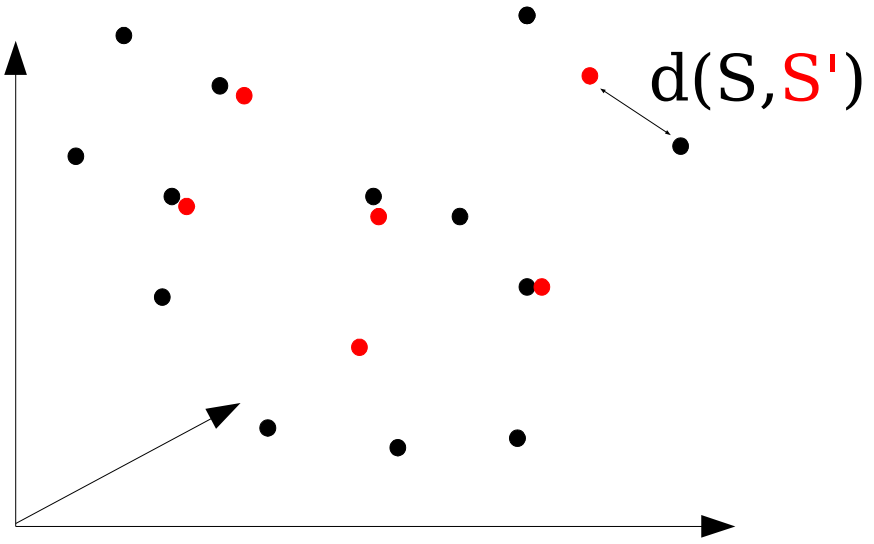
- Cas de plusieurs descripteurs locaux
  - Une mesure de similarité entre deux descripteurs locaux d(Xi,X'j)
  - Une mesure de similarité globale entre deux ensembles de descripteurs locaux d(S,S')



d(S,S') ?

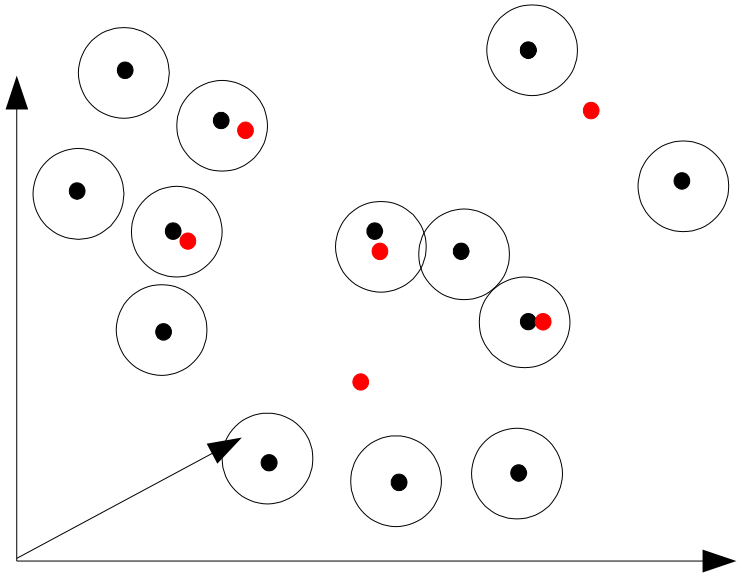
CBIR: Mesures de similarité

- Distance de Hausdorff
  - $d(S,S') = \max_j (\min_i (d(Xi,X'j)))$



CBIR: Mesures de similarité

- Vote
  - Comptage du nombre de matchs



$n(S,S')=4$

$d(S,S')=1/(1+n(S,S'))$

CBIR: Recherche par similarité

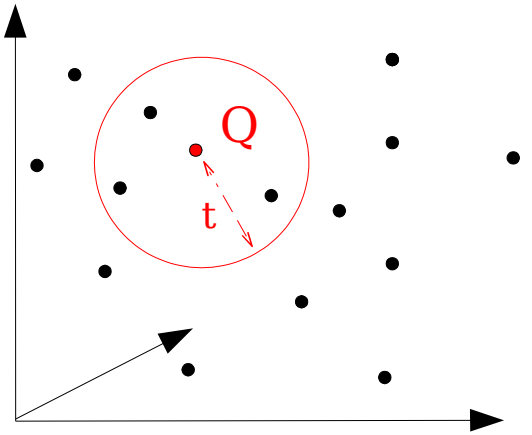
- Requête Q (descripteur d'une image candidate Iq)
- Base de descripteurs B={Xi} (1<i<N)
- Quelle est l'image de la base la plus similaire à la requête ?

  Quels est le descripteur ayant la plus petite distance avec la requête Q ?

  Quel j tel que d(Q,Xj)=min(d(Q,Xi)) (1<i<N)

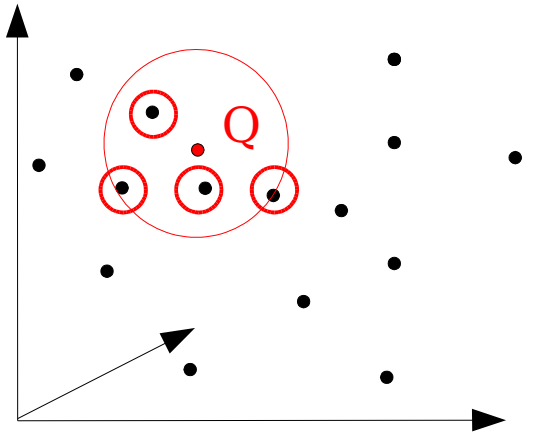
CBIR: Recherche par similarité

- Requête à un rayon près
  - Donne moi toutes les images de la base ayant une certaine similarité avec mon image requête
  - {Xj} tel que d(Xj,Q)<t quel que soit j
  - Dans le cas de la distance L2:



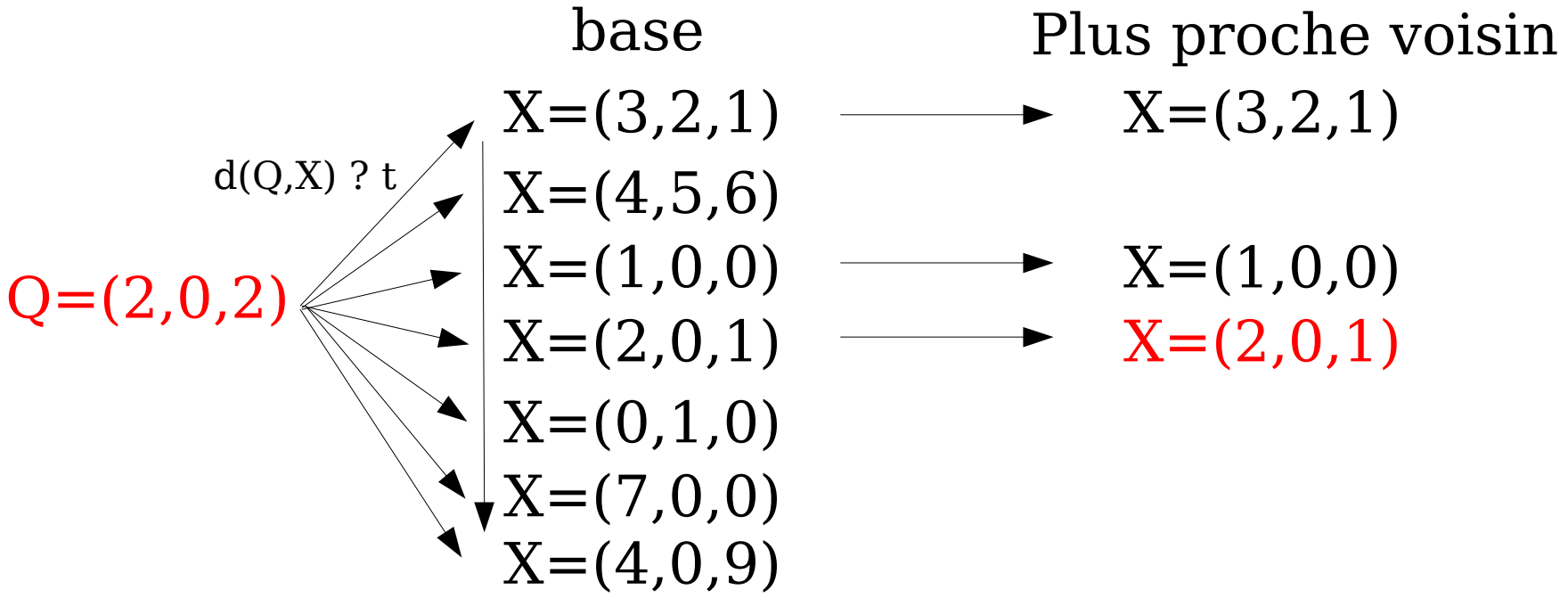
CBIR: Recherche par similarité

- Requête des K plus proches voisins
  - Donne moi les K images de la base ayant la plus grande similarité avec mon image requête
  - {Xj} tel que d(Xj,Q)<max(d(Xj,Q))
  - Dans le cas de la distance L2 et K=4:



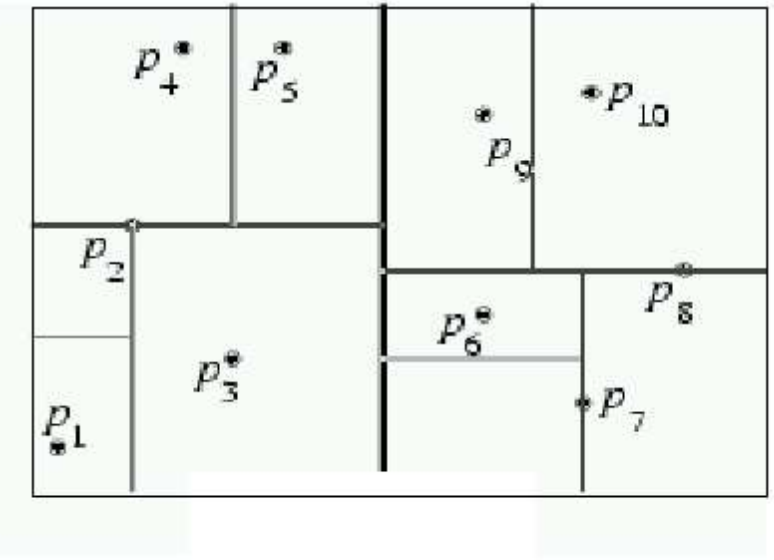
CBIR: Recherche par similarité

- **Scan séquentiel** et exhaustif de la base
  - Méthode de recherche la plus simple
  - Coût **linéaire** en fonction de la taille de la base O(N)

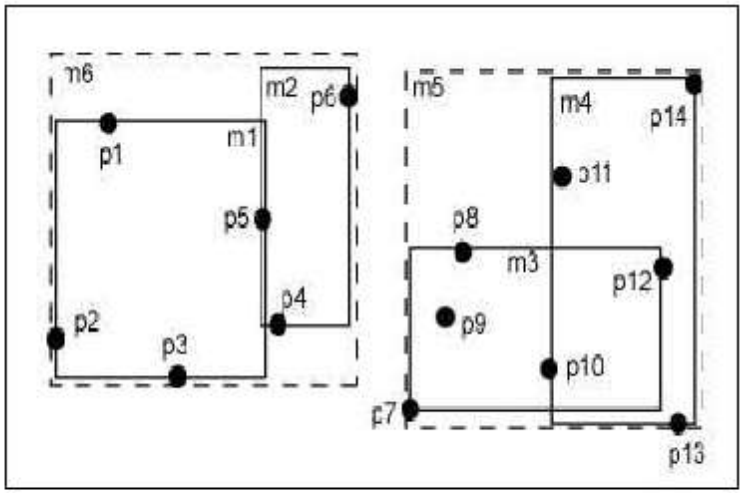


CBIR: Structures d'indexation

- Structuration des données en mémoire (ou sur disque)
- Partitionnement de l'espace ou partionnement des données



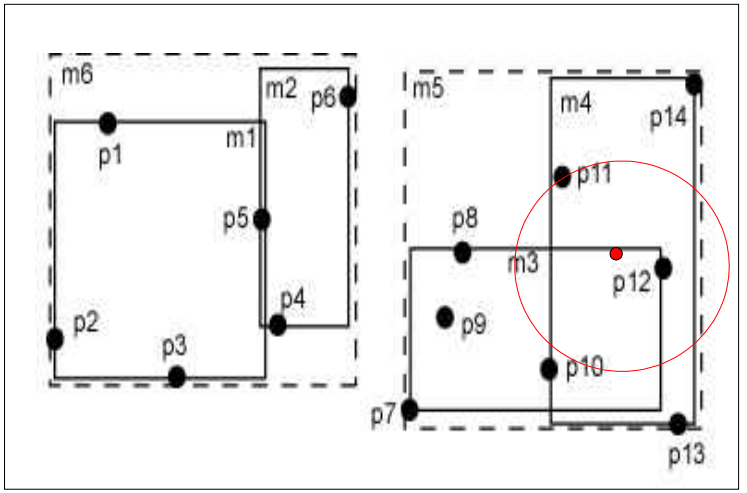
KD-tree

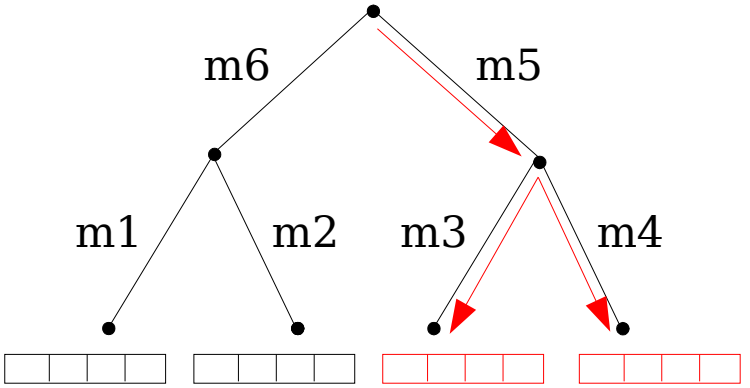


R-tree

CBIR: Structures d'indexation

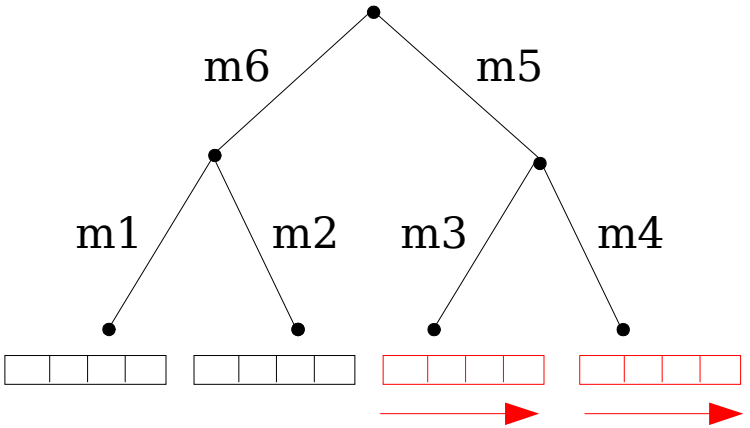
- Recherche en deux étape
  - Etape de filtrage = sélection des formes englobantes susceptibles de contenir des réponses à la requête
  - Requête à un rayon près dans un R-tree

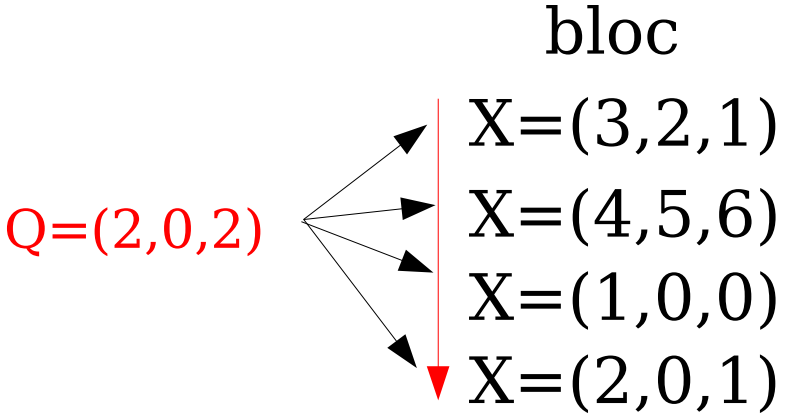




CBIR: Structures d'indexation

- Etape de raffinement: scan exhaustif des blocs sélectionnés

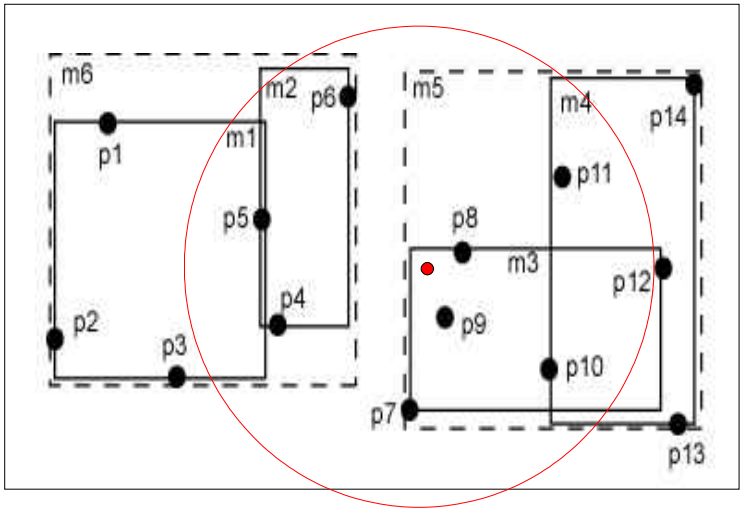




- Temps total de recherche:
  - $T = T_f + T_r = N_{blocs} . t_f + \frac{N_r}{N} T_s$

CBIR: Structures d'indexation

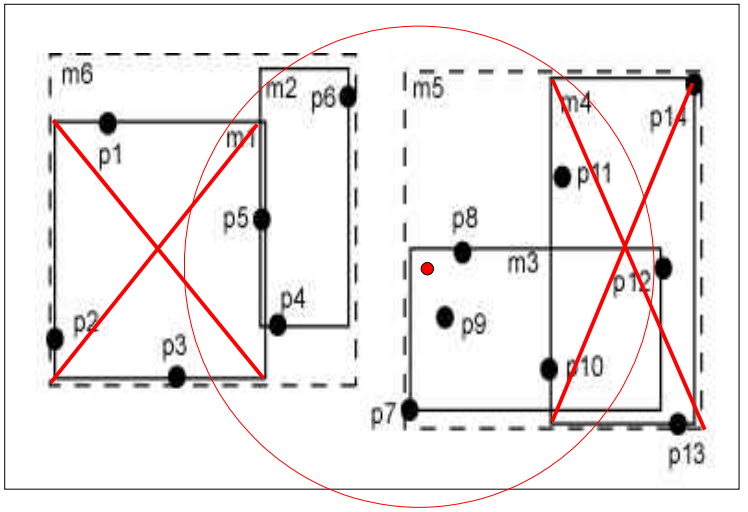
- Malédiction de la dimension
  - Lorsque D est très grand
  - Nblocs est très grand



  - $T = T_f + T_r = N_{blocs} . t_f + \frac{N_r}{N} T_s \gg T_s$

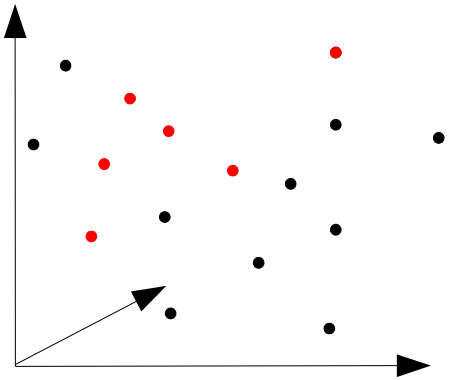
CBIR: Structures d'indexation

- Méthodes approximatives
  - Faible perte de qualité des résultats
  - Gain de temps de recherche très important

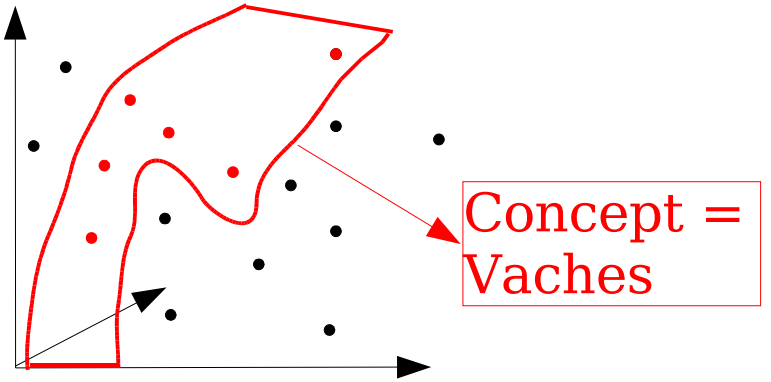


CBIR: Apprentissage

- Principe
  - Base de connaissance
    - Non “vache”
    - Vaches

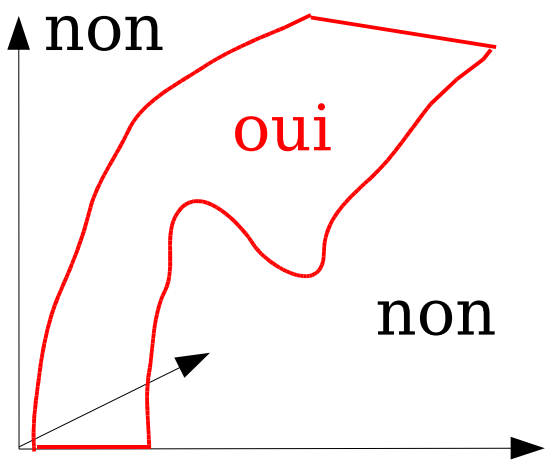


  - Apprentissage:



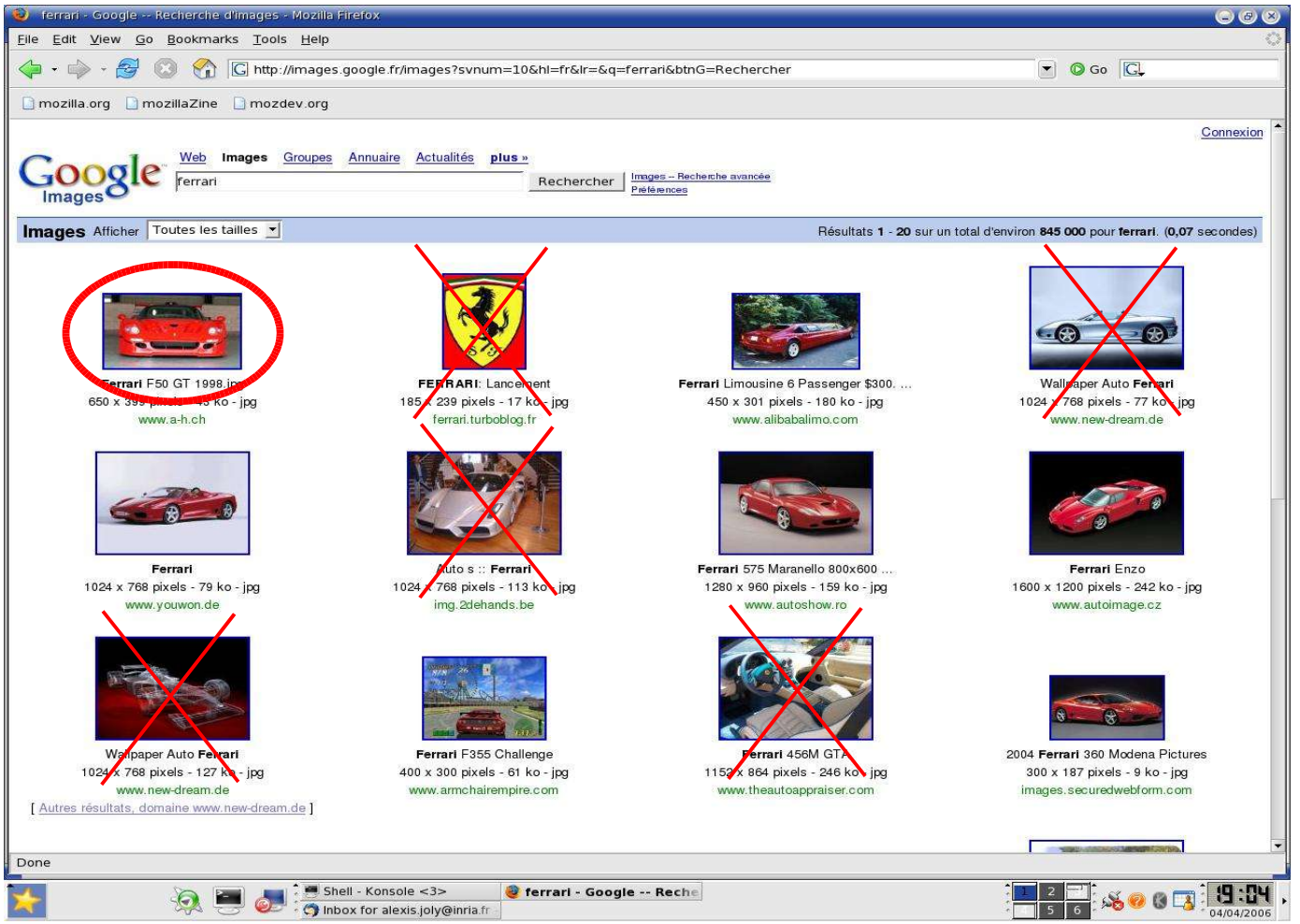
CBIR: Apprentissage

- Requête Q (descripteur image Iq)
- Est-ce une vache ?



CBIR: Applications

- “Google filtering”



CBIR: Applications

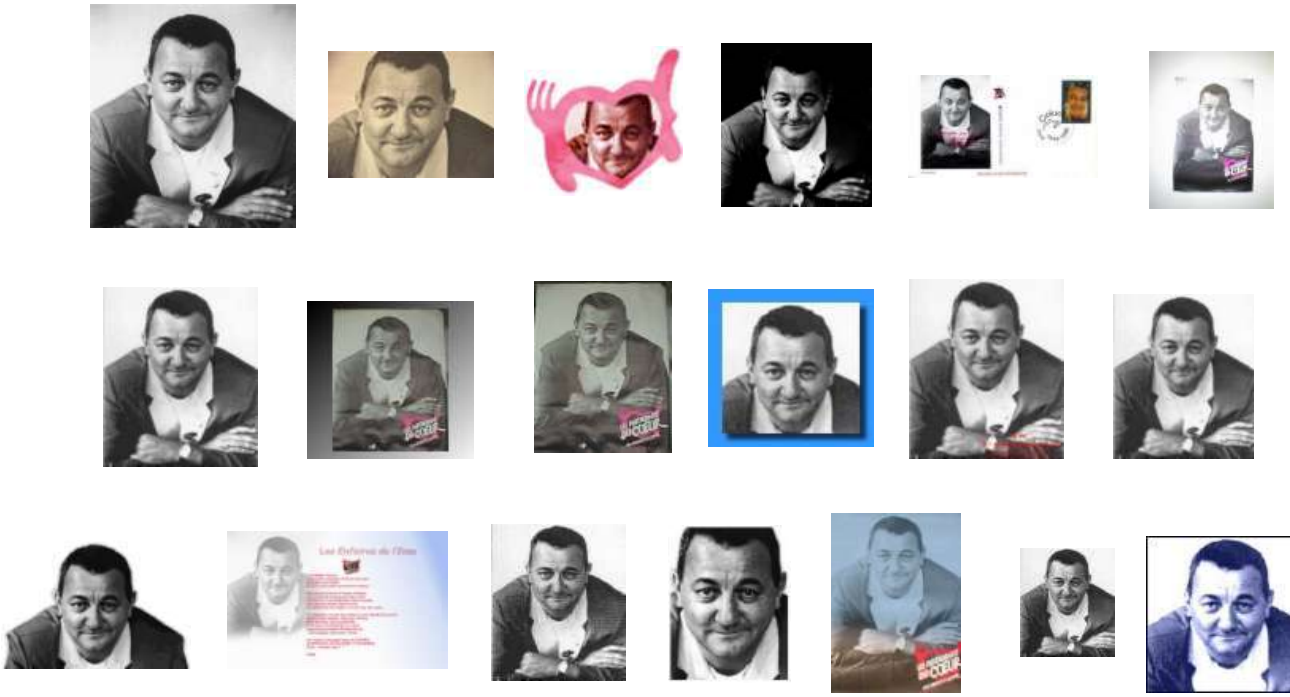
- Satellite
  - Requêtes = plages, stades, parcs

CBIR: Applications

- Annotation semi-automatique
  - Annotations manuelles très coûteuses
  - Reconnaissance de personnes
  - Reconnaissance d'objets
- Indexation de vos photos
  - Recherche des photos de mon frère
  - Recherche des photos de mon chat
  - Recherche des photos où je suis au bord de la mer

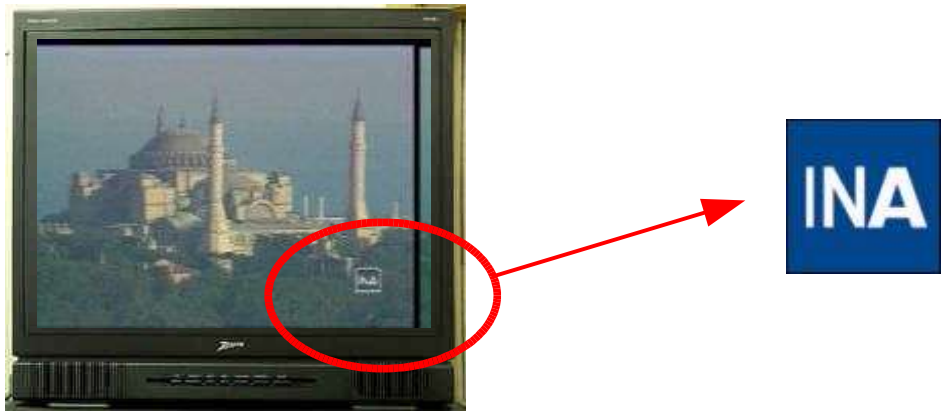
CBIR: Applications

- Détection de copies
  - Copyright
  - Quel est l'image la plus présente sur le Web ?



TV Monit: Intro

- Monitoring de télévision = détection de copies
- Détection d'images diffusées à la TV appartenant à une très grande base d'archive
- Système développé à l'INA



TV Monit: Intro

- Plan
  - L'INA et les bases gérées à l'INA
  - Principes et architecture
  - Résultats et exploitation

TV Monit: INA

- Conservation et exploitation du patrimoine TV français (70 ans)
- 1,000,000 d'heures de vidéo (Archives, dépôt légal, Inathèque, SNC)
- 300,000 heures de vidéo numérisées

TV Monit: INA

- Les supports de stockage
  - Le film (analogique)
    - 35mm, 1914-1969, cinéma et actualité dans les salles
    - 16 mm, 1949-1974, + rapide à traiter
    - Kinéscope, pas de magnétoscope !

TV Monit: INA

- Les supports de stockage
  - Les bandes analogiques
    - 2 pouces, 1962-1986, support magnétique, 90 minutes, conserve la qualité d'origiene
    - 1 pouce, miniaturisation (½)
    - Sony ¾ pouce U-matic, ¾ pouce BVU
    - Beta analogique ½ pouce, caméscope, reporters




INRIA

TV Monit: INA

- Les supports de stockage
  - Les bandes numériques
    - La betacam SP, 1990, premier standard de production sur support numérique
    - La betanum, 1993, standard actuel du haut de gamme, sauvegarde des fonds INA, 124 minutes, 100 Mbits/sec.
    - La Sony betacam SX, programmes France 2 depuis 1999, “MPEG Sony”, compressés à 20 Mbits/sec.

TV Monit: INA

- Les supports de stockage
  - Les nouveaux supports numériques
    - MPEG1 1 Mbit/s, consultation des extraits, consultation en ligne
    - MPEG2 2 à 8 Mbit/s, format actuel de livraison des extraits
    - DVD, MPEG1 1 Mbit/s ou MPEG2 2 à 8 Mbit/s, robots de gravage, dépôt légal

TV Monit: INA

- Les bases
  - Le fond d'archives
    - 700,000 heures
    - déposées par les chaines de TV
    - Parallèles antennes
    - Supports: du film à la betanum
    - **1,600,000 notices, 150 corpus thématiques**
      - Méta-data: émission, date, etc.
      - Sous-découpages
      - Descriptions (noms des artistes, évènement,etc.)

TV Monit: INA

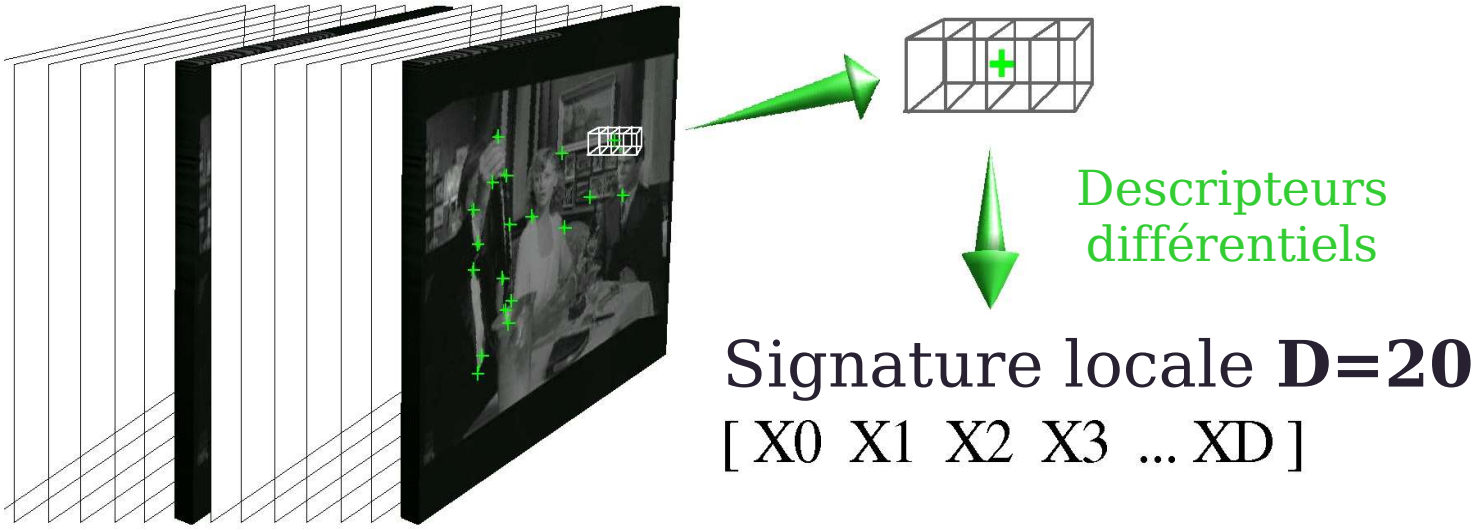
- Les bases
  - Numérisation du fond d'archives
    - 250,000 heures
    - Numérisation des notices dans l'**INAThèque** (Base Oracle)
    - Calcul et indexation des fichiers MPEG1 de visualisation dans la base **SNC** (serveurs linux)
    - Pointeurs depuis l'INAThèque vers SNC
  - Moteurs de recherche INAThéque
    - Par mots-clés sur les notices, par date, par émission, etc.

TV Monit: INA

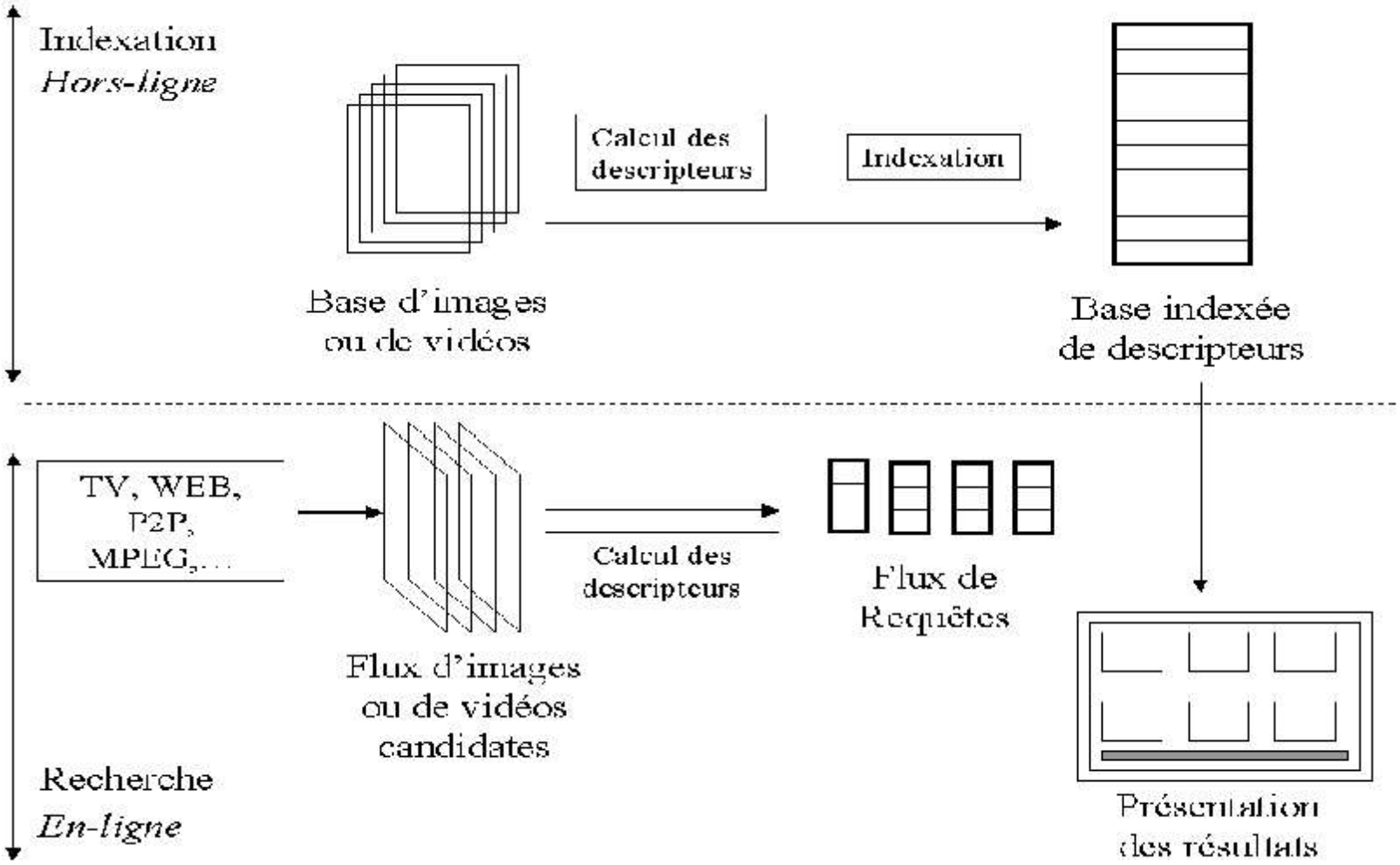
- Les bases
  - Le dépôt légal
    - Base indépendante
    - Loi du 20 juin 1992
    - 40 chaîne de TV en continu
    - MPEG 1 et MPEG 2 sur DVD
    - Indexés uniquement par date en tertiaire
    - Buffer disque de 10 jours accessibles par le CSA, etc.

TV Monit: Principe global

- Recherche de copies par le contenu
- A l'aide de descripteurs locaux basés sur des points d'intérêt dans des images clés

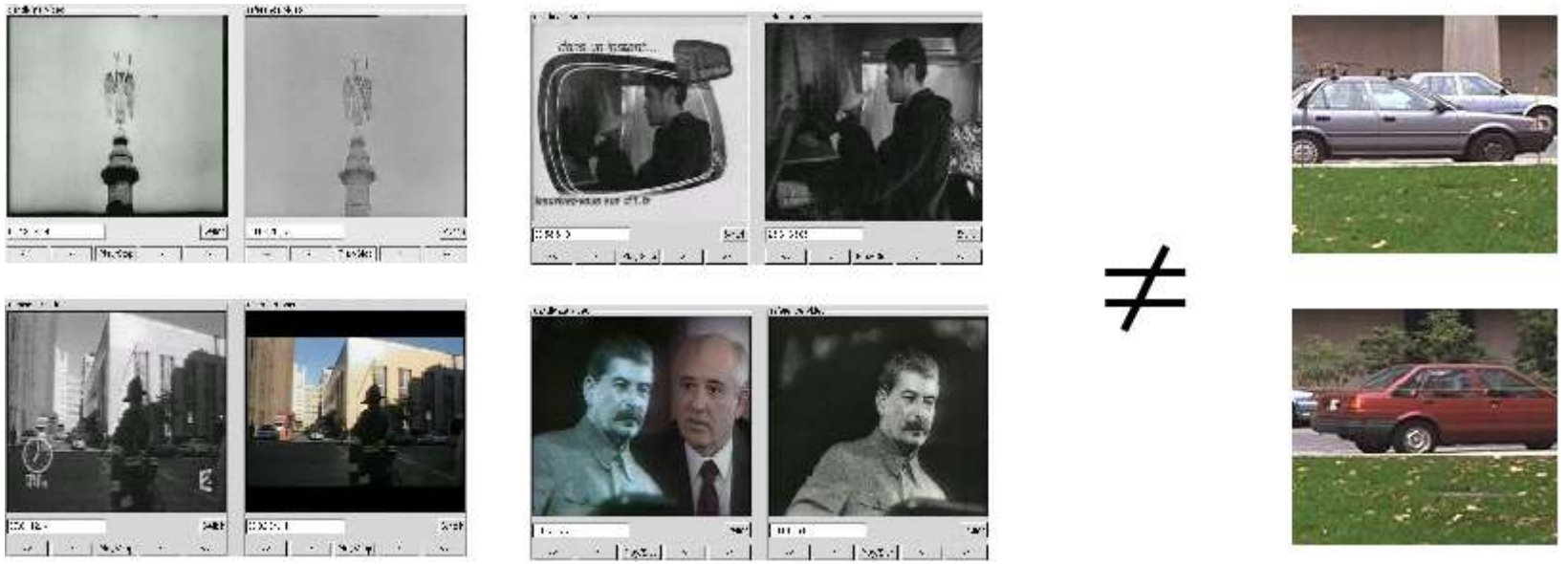


TV Monit: Principe global

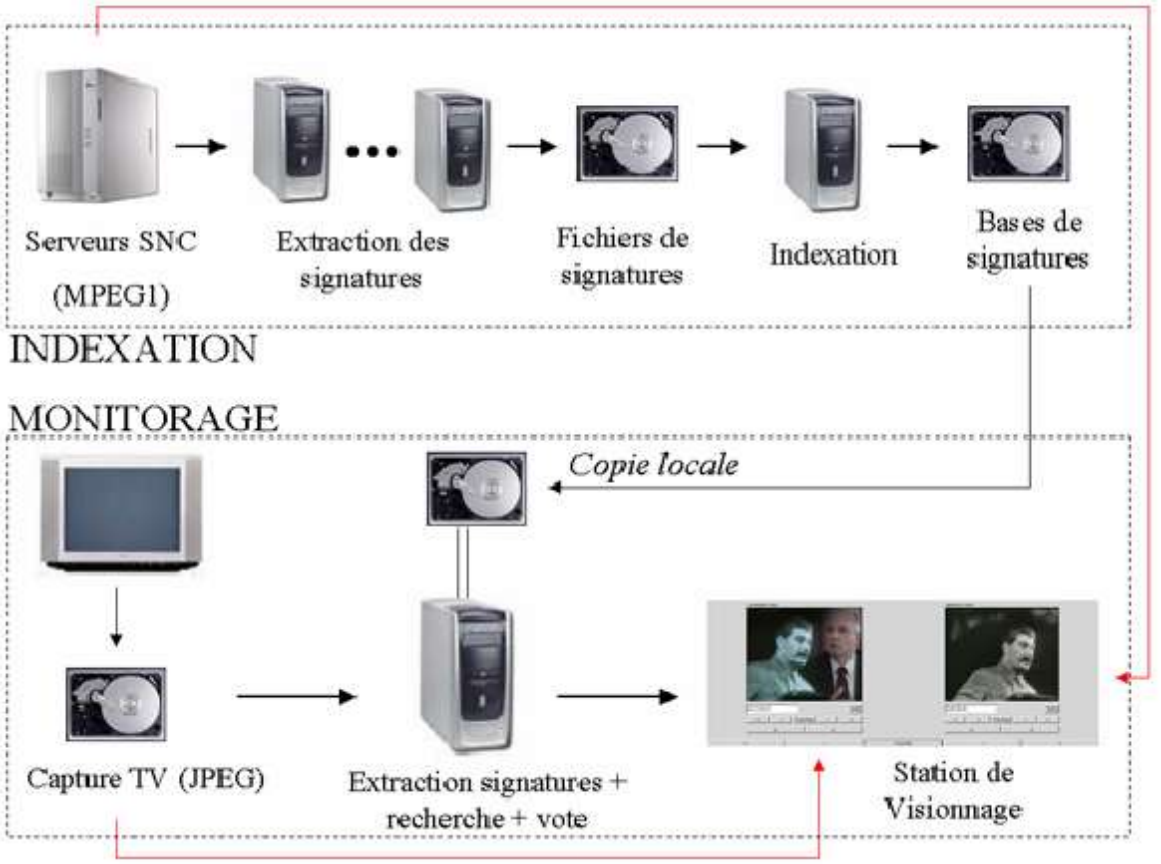


TV Monit: Principe global

- Spécificités
  - Requêtes soumises automatiquement par la capture télé
  - Similarités + transformation

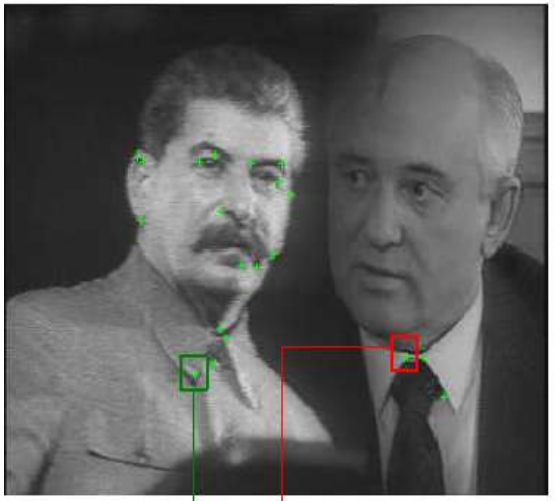


TV Monit: Architecture



100,000 heures en 1 an

TV Monit: Principe de la recherche



Base: Extrait **123**

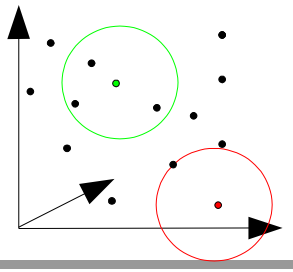
Q1

Q2

Recherche dans la base

8:S_{13} **84**:S_2 **123**:S_{12} **67**:S_3 **849**:S_9

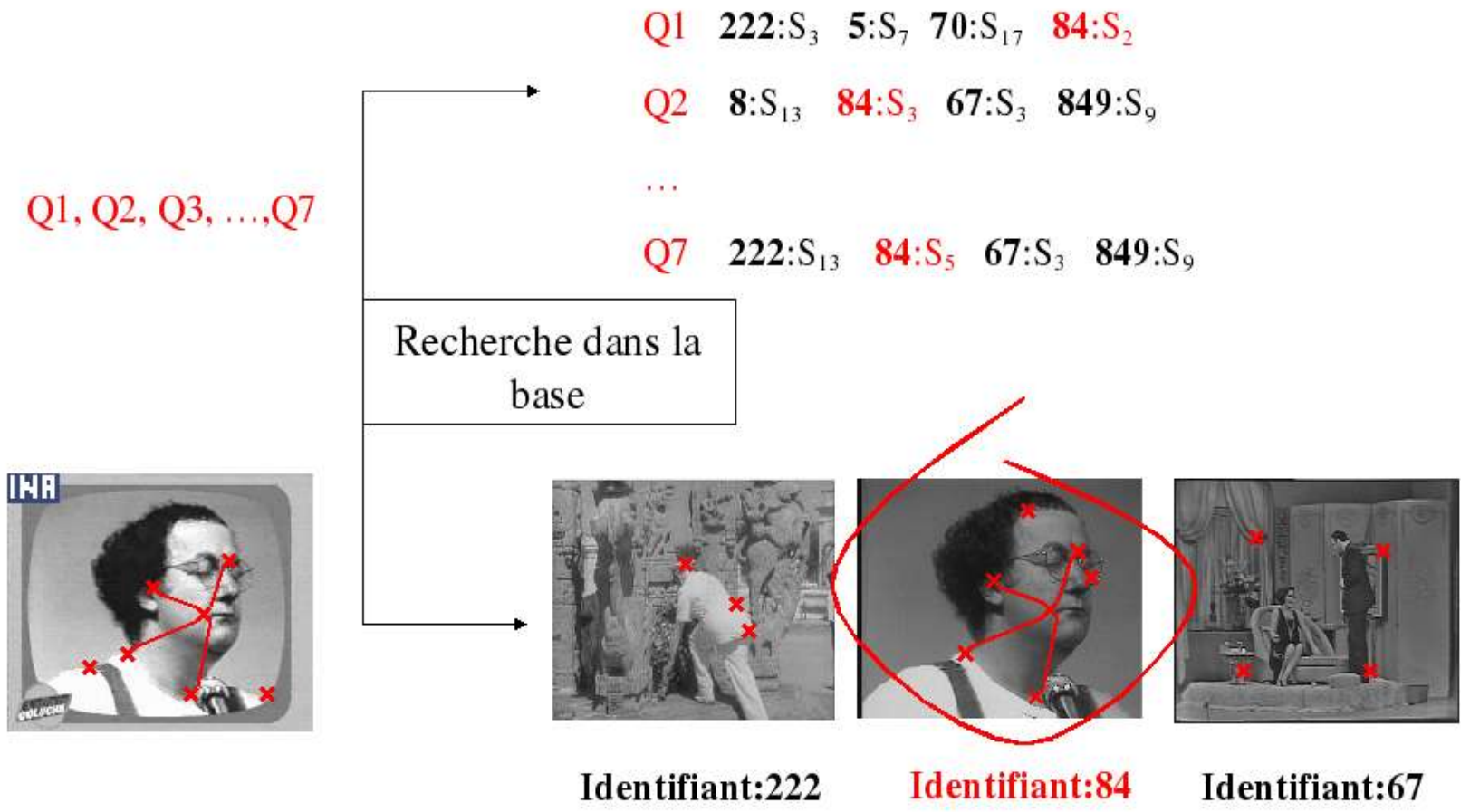
222:S_3 **5**:S_7 **70**:S_{17} **146**:S_{13}



Recherche à un rayon près approximative

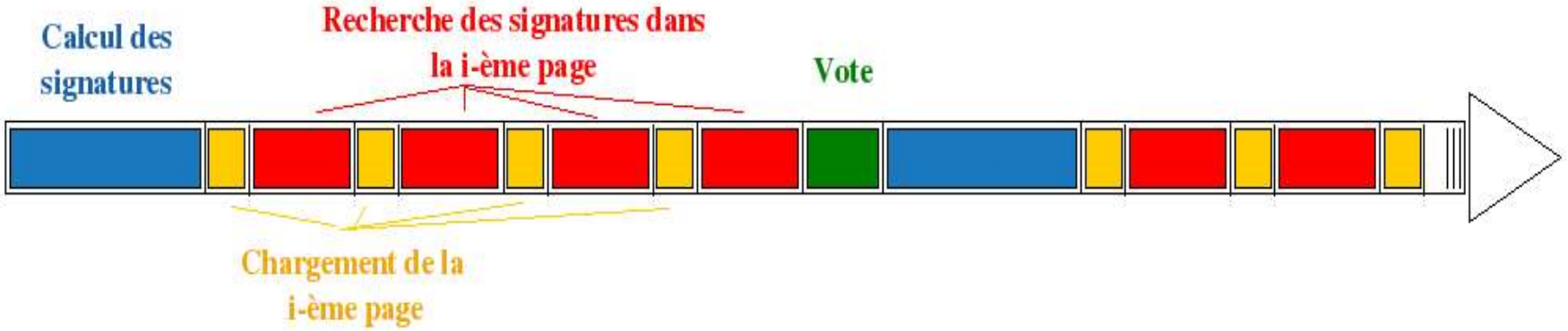
TV Monit: Principe de la recherche

- Contraintes géométriques



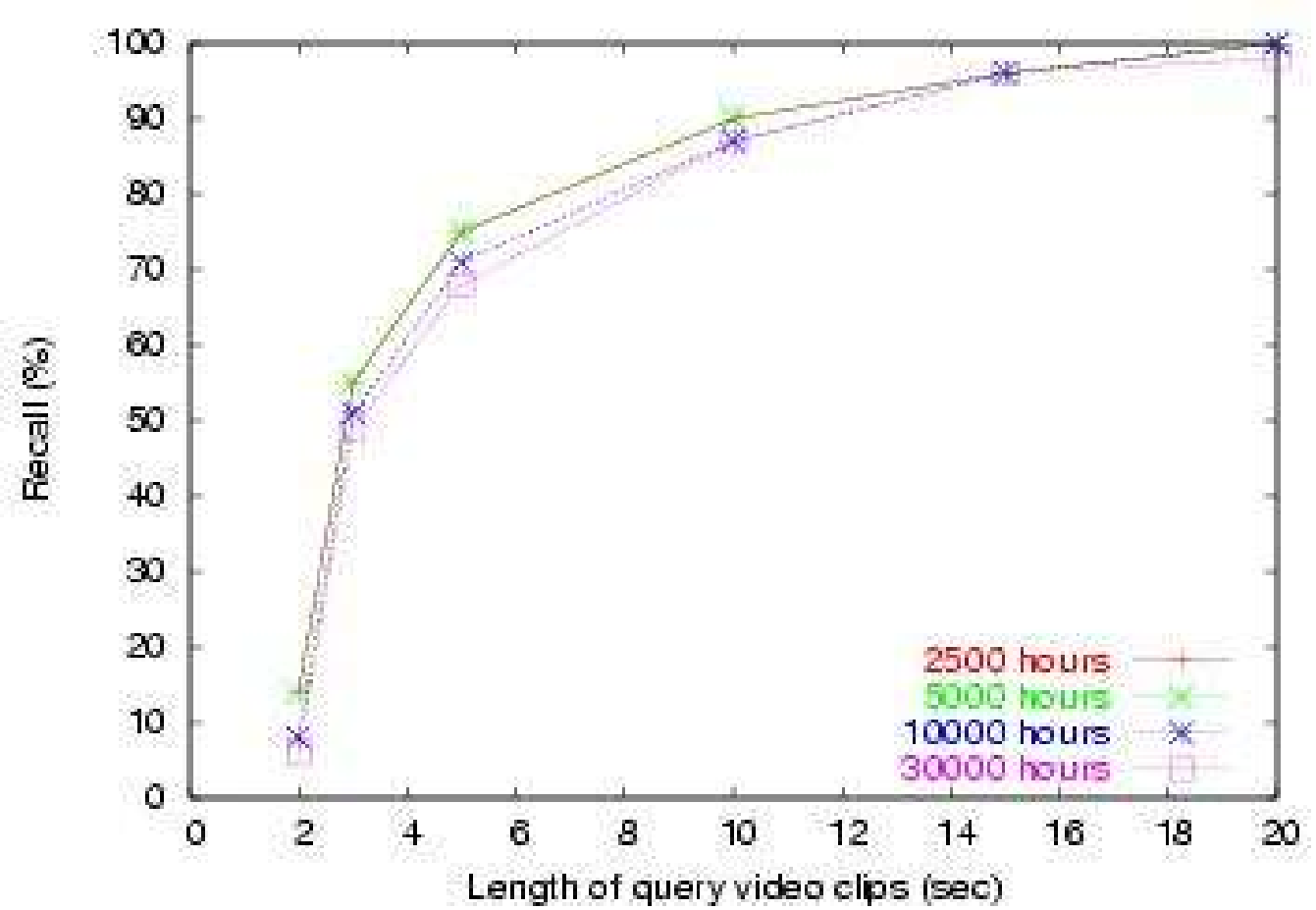
TV Monit: Principe de la recherche

- Très grandes bases de signatures
  - Taille > taille mémoire
  - Cumul des requêtes pendant plusieurs heures de signatures (résultats plusieurs après diffusion)
  - Load de la base par pages successives



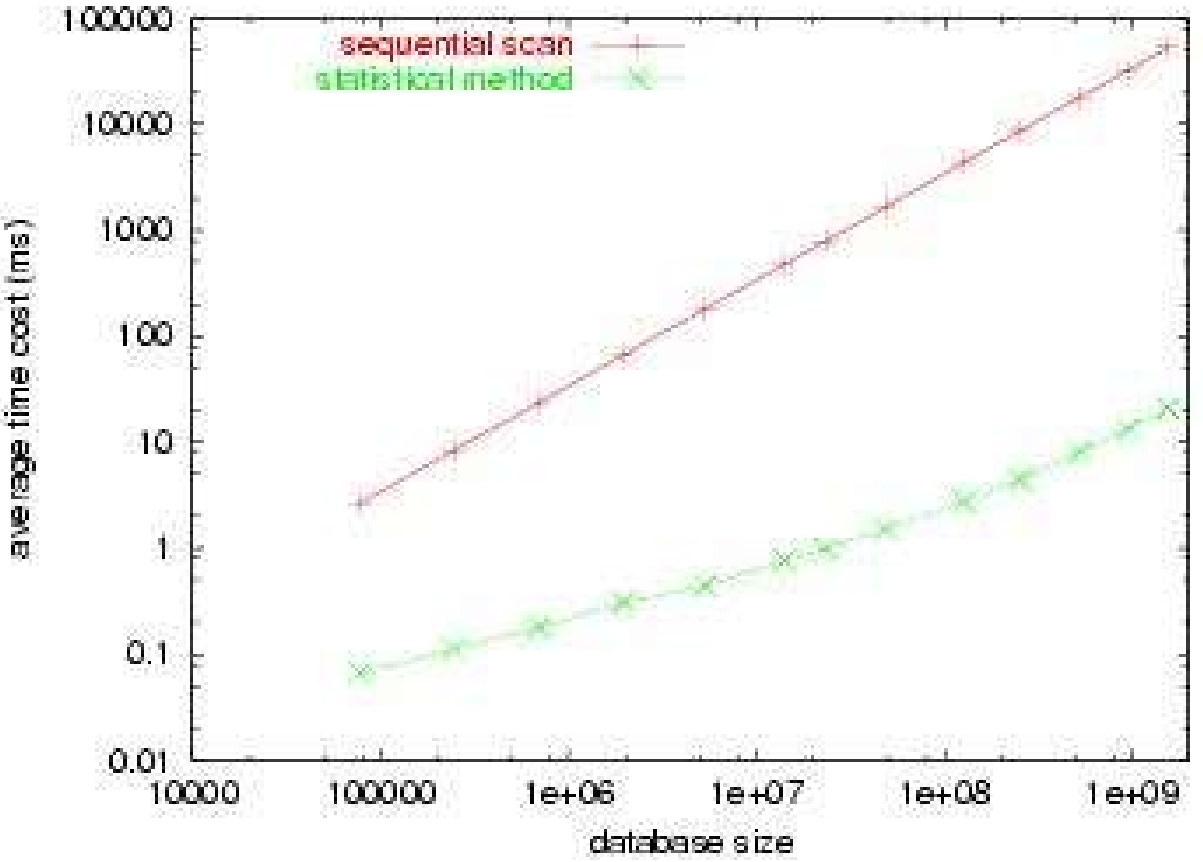
TV Monit: performances

- Qualité de la recherche



TV Monit: performances

- Temps de recherche



TV Monit: Résultats

- Statistiques
  - Sans filtrage des parallèles antennes
    - Base 30,000 heures
    - 300 détections / jour / chaine (90% de pubs)
    - 20 fausses alarmes / jour / chaine
    - Précision = 93 %
  - Avec filtrage des parallèles antennes
    - Base 30,000 heures
    - 30 détections / jour / chaine
    - 20 fausses alarmes / jour / chaine
    - Précision = 60 %

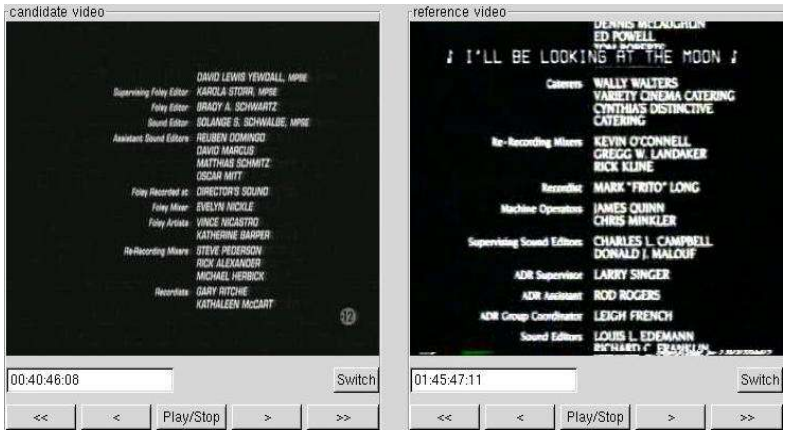
INRIA

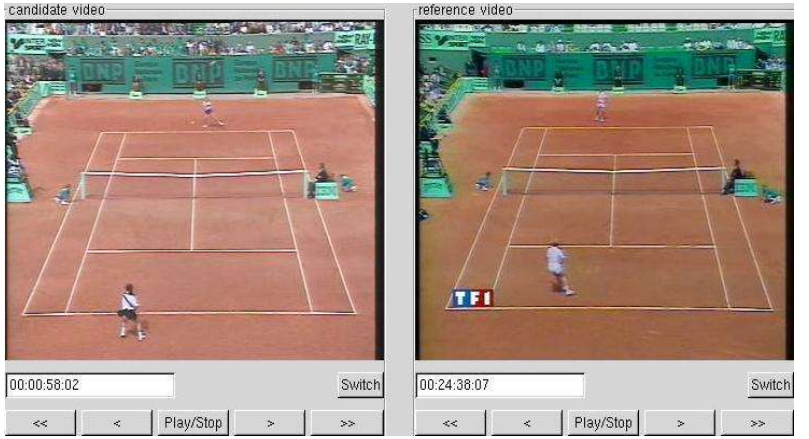
TV Monit: Résultats

- Bonnes détections

TV Monit: Résultats

- Fausses alarmes





TV Monit: Autres applications

- Détections inter-chaînes
  - Sujet international ou national

TV Monit: Autres applications

- Détections intra-chaînes
  - Fréquence de diffusion
  - Elimination des publicités

Conclusion et perspective

- CBIR en plein expension
  - Volumes multimédia tjs + en + grands
  - Diversification des applis
    - Pro
    - Grand public
- Détection de copies vidéo
  - 100,000 heures d'archives + temps réel
  - Applis diverses, extraction de connaissance haut niveau grâce au contexte

Conclusion et perspective

- Limitations des CBIR
  - “Semantic Gap” = Gap sémantique
    - Requête utilisateurs:
      - “La première télé de Chirac”
      - Retrouver des images de “chute”
    - Construction de bases d'apprentissage réalistes
      - Trop cher
      - Trop lent
    - On est collé au signal
      - Les sémantiques de haut niveau ne correspondent pas à des primitives visuels
      - Utilisation du contexte