

# INDEXATION CONTENU MULTIMEDIA INFORMATION SPOTTING



Nicolas Sidere

**Laboratoire Informatique Image Interaction (L3i)**

Université de La Rochelle - Pôle Sciences et Technologie - Avenue Michel Crépeau - 17042 LA ROCHELLE CEDEX 1 France

Tél : +33 (0)5 46 45 82 62 – Fax : 05.46.45.82.42 – Site internet : <http://l3i.univ-larochelle.fr/>

# OBJECTIFS ET PROGRAMME DE CE PROJET

## Indexation visuelle de document

- Décrire le contenu
- L'indexer
- Le retrouver

# UN MONDE 100% NUMÉRIQUE ?

273 milliards d'emails sont échangés chaque jour

- 1,4 milliards pour la France (hors spam)
- Entre 55% et 95% des emails sont des spams

Des données de plus en plus nombreuses

- 90% des données numériques créées ces deux dernières années

Des contenus de plus en plus hybrides

- Existent à la fois en numérique et en papier
- Besoin de retrouver de l'information dans les « deux mondes »

# UN MONDE SANS PAPIER ?

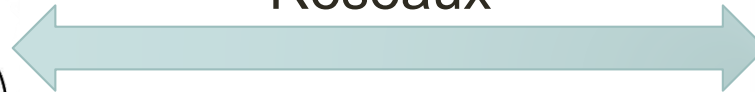
Besoin accru d'accéder et de manipuler l'information



Document papiers



Réseaux



Document numériques



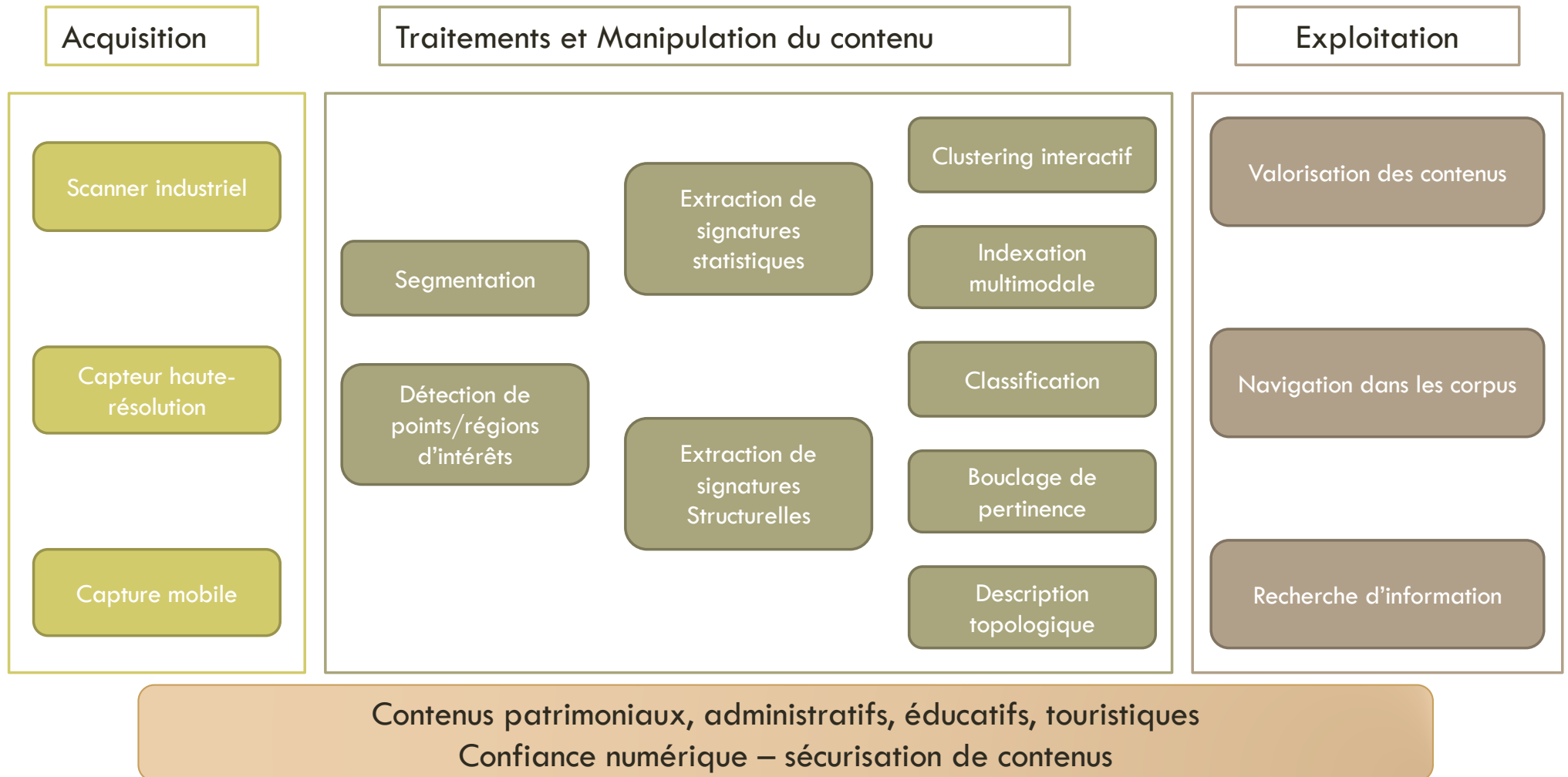
Stockage et indexation



Numérisation



# DE LA CAPTURE À L'EXPLOITATION DES DOCUMENTS



# LIMITES ET IMPACTS DE LA NUMÉRISATION

Questions fondamentales pour votre projet

- Savez-vous comment numériser vos documents ?
- Savez-vous comment décrire vos documents ?
- Etes-vous conscient des limites de votre projet ?
- Savez-vous comment permettre un accès adapté à vos données, pour des publics divers ?

# EXEMPLE DE COUTS

Exemple d'une Université de 1 000 salariés et 9 000 étudiants

- Tous les jours, un salarié imprime 1 page ( $\sim 250$  pages / an)
  - Si chaque page coute 0,10 € (encre + papier + machine)
- Cout annuel = 25 000 €
  - Et les étudiants ne sont pas inclus
- Si chaque salarié passe 20 min par jour à chercher de l'information
  - 250 jours travaillés / an ; 1h coute 20 € (estimation basse)
- Cout annuel = 1 000 000 €

# COÛT D'UNE GESTION MANUEL DES DOCUMENTS

**7.5%**  
**OF ALL DOCUMENTS GET  
LOST**

**\$30**

THE AVERAGE COST  
TO FILE  
ONE DOCUMENT

[www.imagenetconsulting.com](http://www.imagenetconsulting.com)



**\$120**  
TO FIND A MISFILED DOCUMENT

**\$220**

THE AVERAGE COST TO REPRODUCE A LOST DOCUMENT



# INFORMATION RETRIEVAL TO DOCUMENT IMAGE RETRIEVAL

Information

Audio

Visual



Text

Graphic

Type-written

Hand-written

Picture

Logo, seal,  
etc.

# DOCUMENT IMAGE RETRIEVAL APPROACHES

1

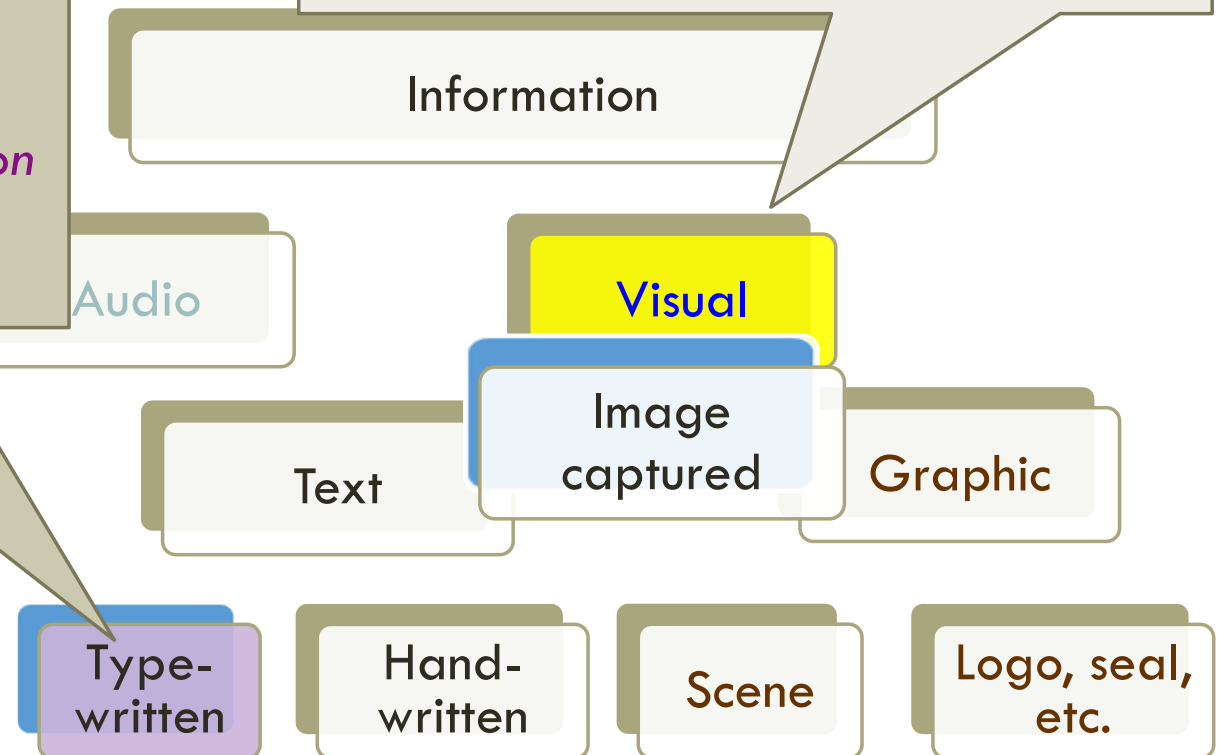
## ➤ Recognition-based approaches

- Convert documents to text using OCR
- Text retrieval techniques
- **Drawbacks:** *high computational cost, sensitive to image resolution or image noise, etc.*

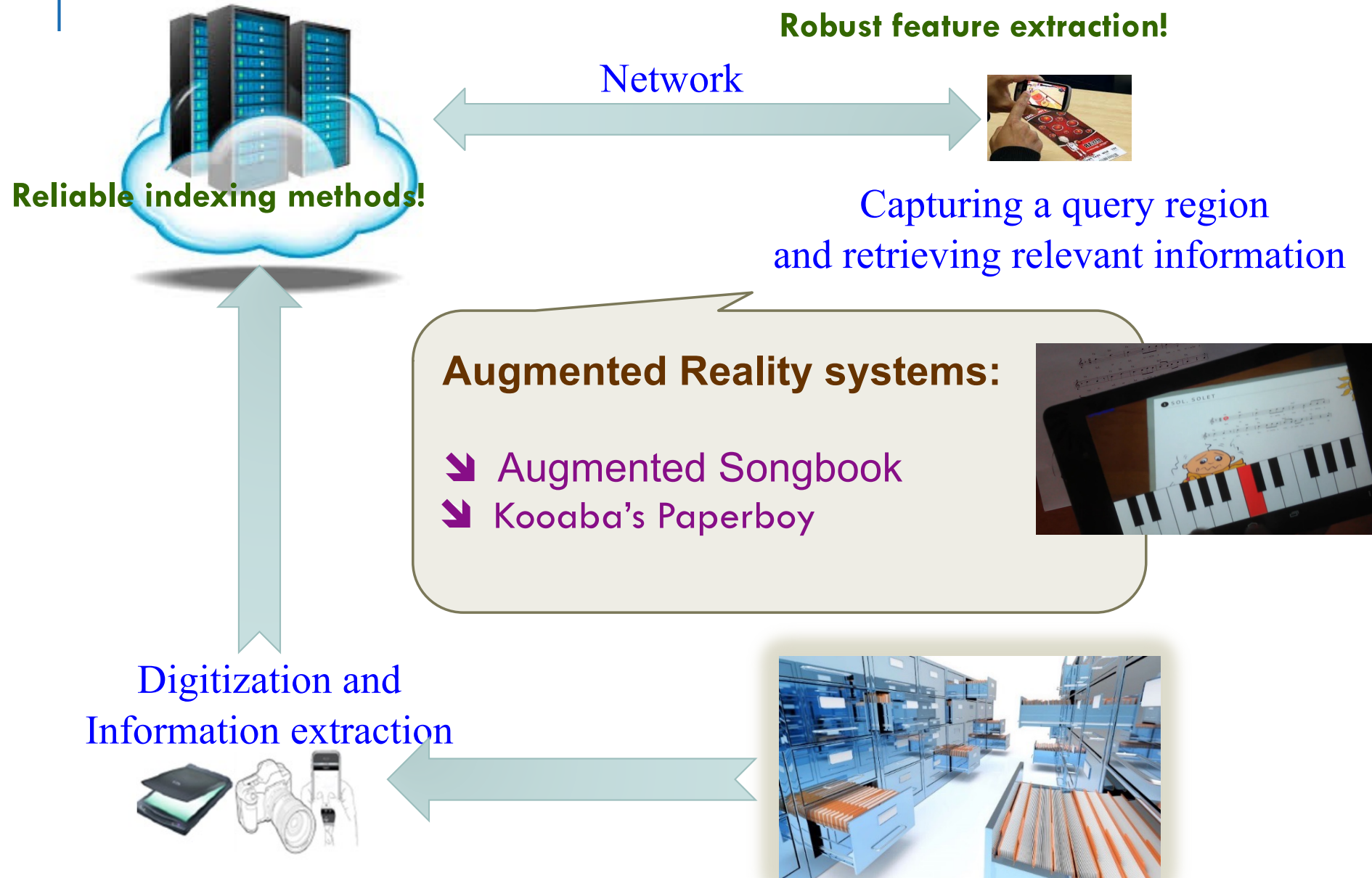
2

## ➤ Recognition-free approaches

- Computation of features
- **Be able to deal with** *image noise, low resolution, viewing transformations, etc.*



# CAMERA-BASED DOCUMENT IMAGE RETRIEVAL



# DEMO DU SYSTÈME À DÉVELOPPER

The image displays a desktop environment with a terminal window on the left and two main application windows. The terminal window shows a command prompt with the text "qiang01@indus: ~" and "Shortcut already assigned." Below the terminal, a "Query" window is visible, showing a video player with a document image. The main application window is titled "Map Retrieval" and displays a document titled "social skills" with a red "Document Retrieval" label. A blue rectangle highlights a section of the document, and a red "0075" label is overlaid. To the right, a "Region Spotting" window shows a similar document with a blue rectangle highlighting a section, labeled "Region spotted" in red. A large grey arrow points from the "Map Retrieval" window to the "Region Spotting" window.

Terminal

Map Retrieval

social skills

Document Retrieval

Query

video

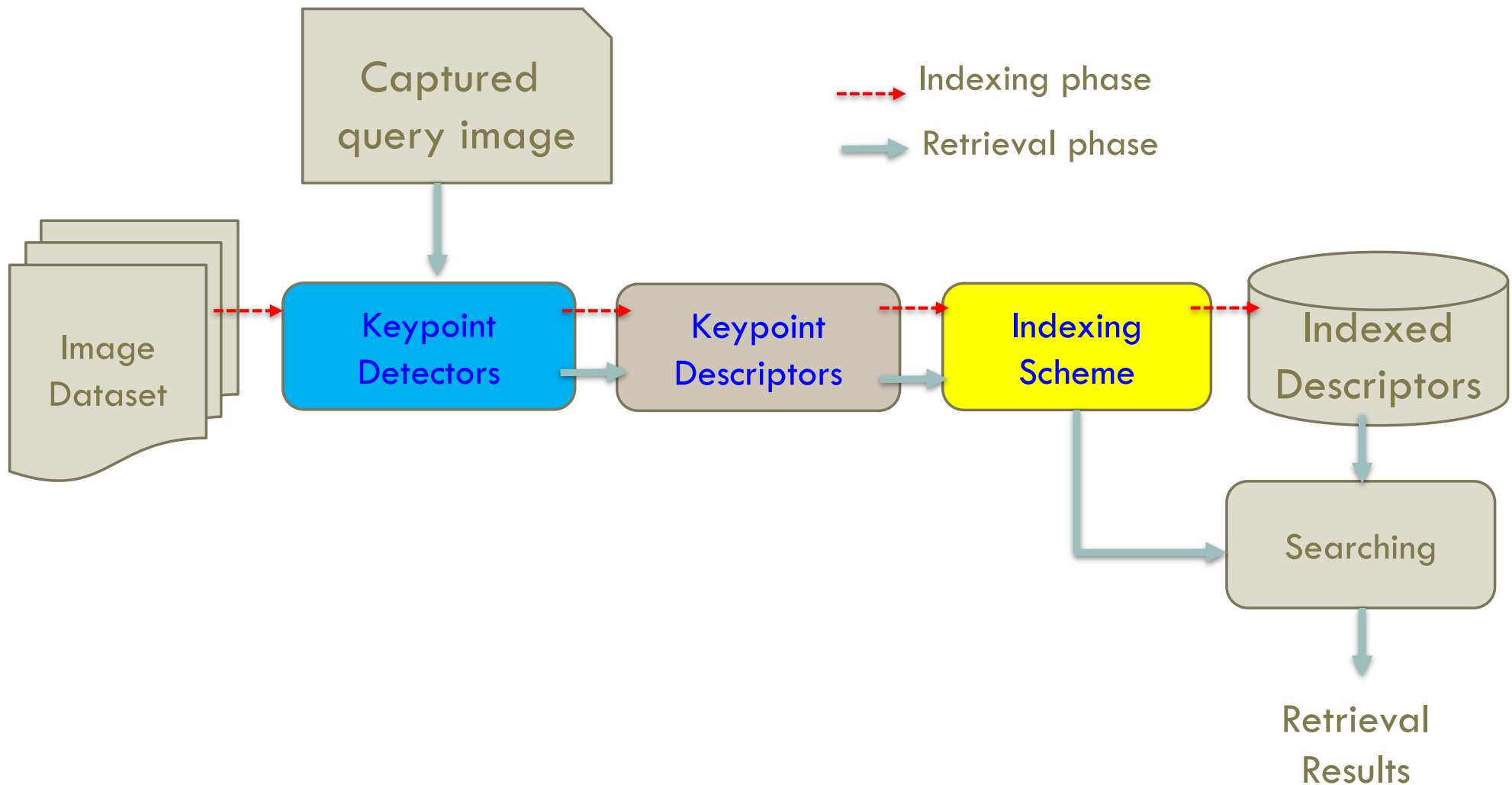
Region Spotting

Region spotted

0075

qiang01@indus: ~  
Shortcut already assigned.

# CAMERA-BASED DOCUMENT IMAGE RETRIEVAL SYSTEMS USING LOCAL FEATURES



# PARTIE 1

## DU CONTENU À SA DESCRIPTION



**Laboratoire Informatique Image Interaction (L3i)**

Université de La Rochelle - Pôle Sciences et Technologie - Avenue Michel Crépeau - 17042 LA ROCHELLE CEDEX 1 France

Tél : +33 (0)5 46 45 82 62 – Fax : 05.46.45.82.42 – Site internet : <http://l3i.univ-larochelle.fr/>

# DÉCRIRE UNE IMAGE... COMMENT ?

Détecter des points / des zones d'intérêts

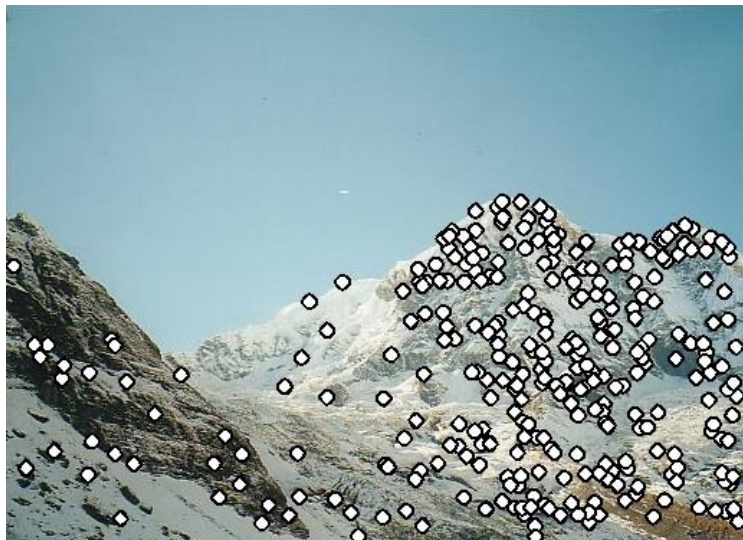
- Ce qu'il faut décrire
- Ce qui est important → Ce que l'on veut indexer

Les points / zones doivent être invariants ou robustes

- Aux changements affines
- Aux translations
- Aux rotations
- Aux changements d'échelle

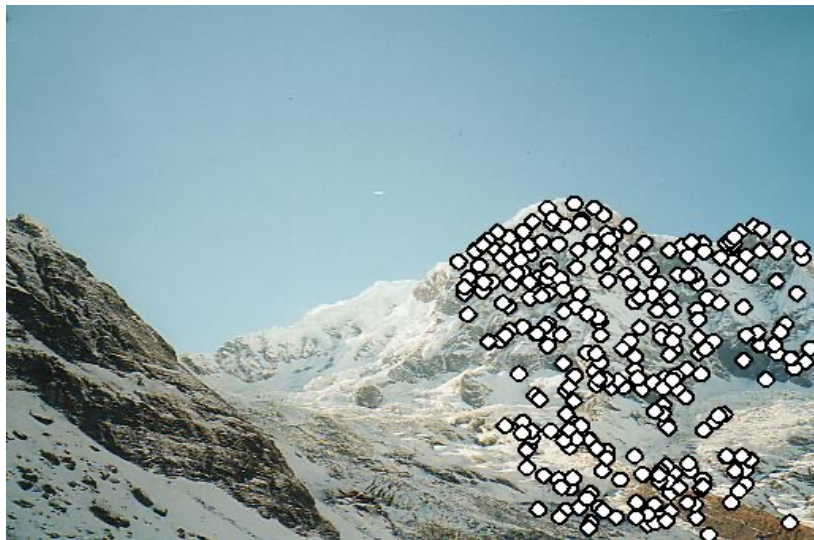


## EXEMPLE 1 : COLLER 2 PHOTOS D'UN MÊME OBJET (SIFT)

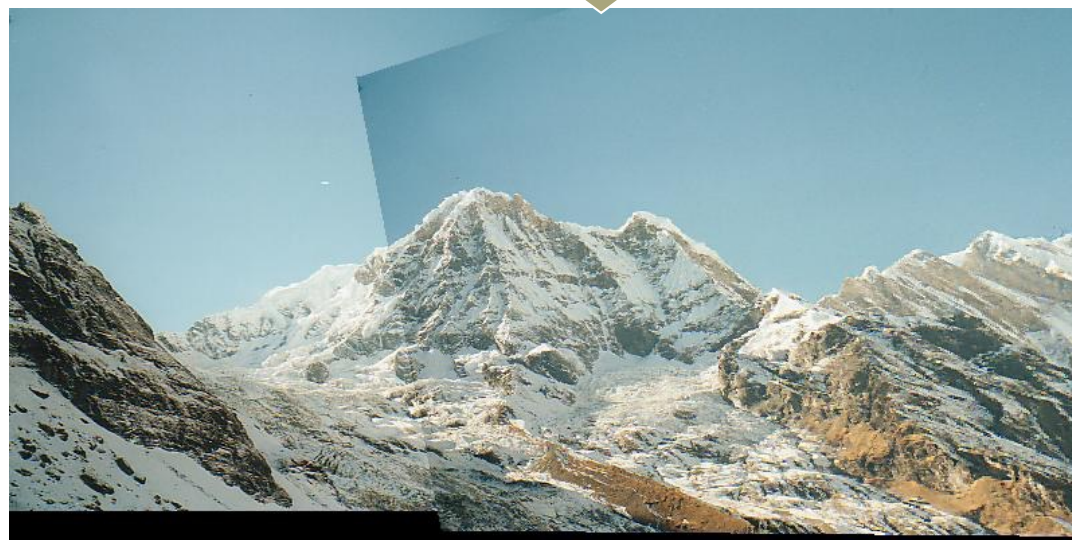




# ON UTILISE CES POINTS POUR CALCULER LA TRANSFORMATION À PARTIR DES POINTS COMMUNS

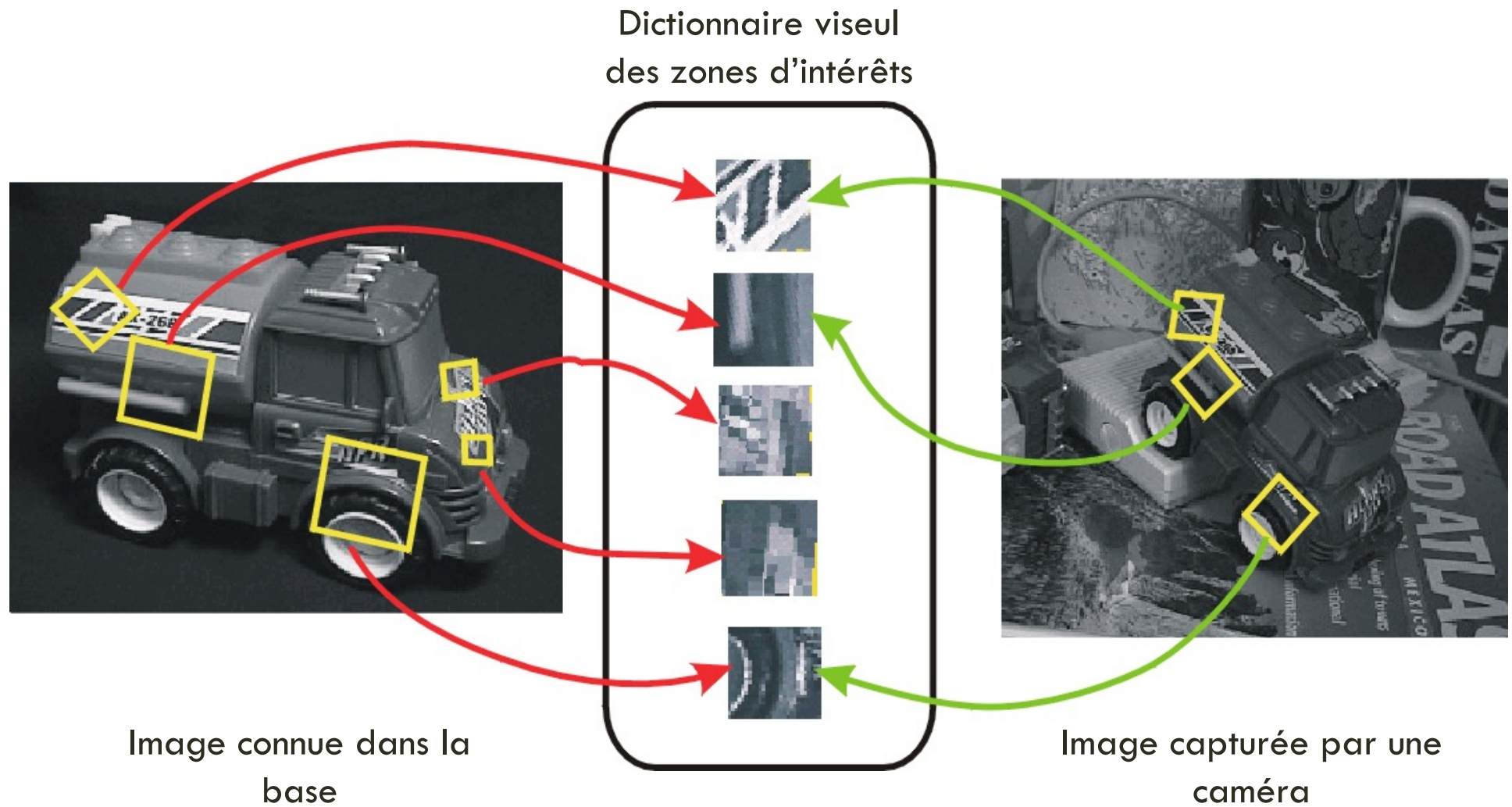


ET ON “COLLE” CES IMAGES

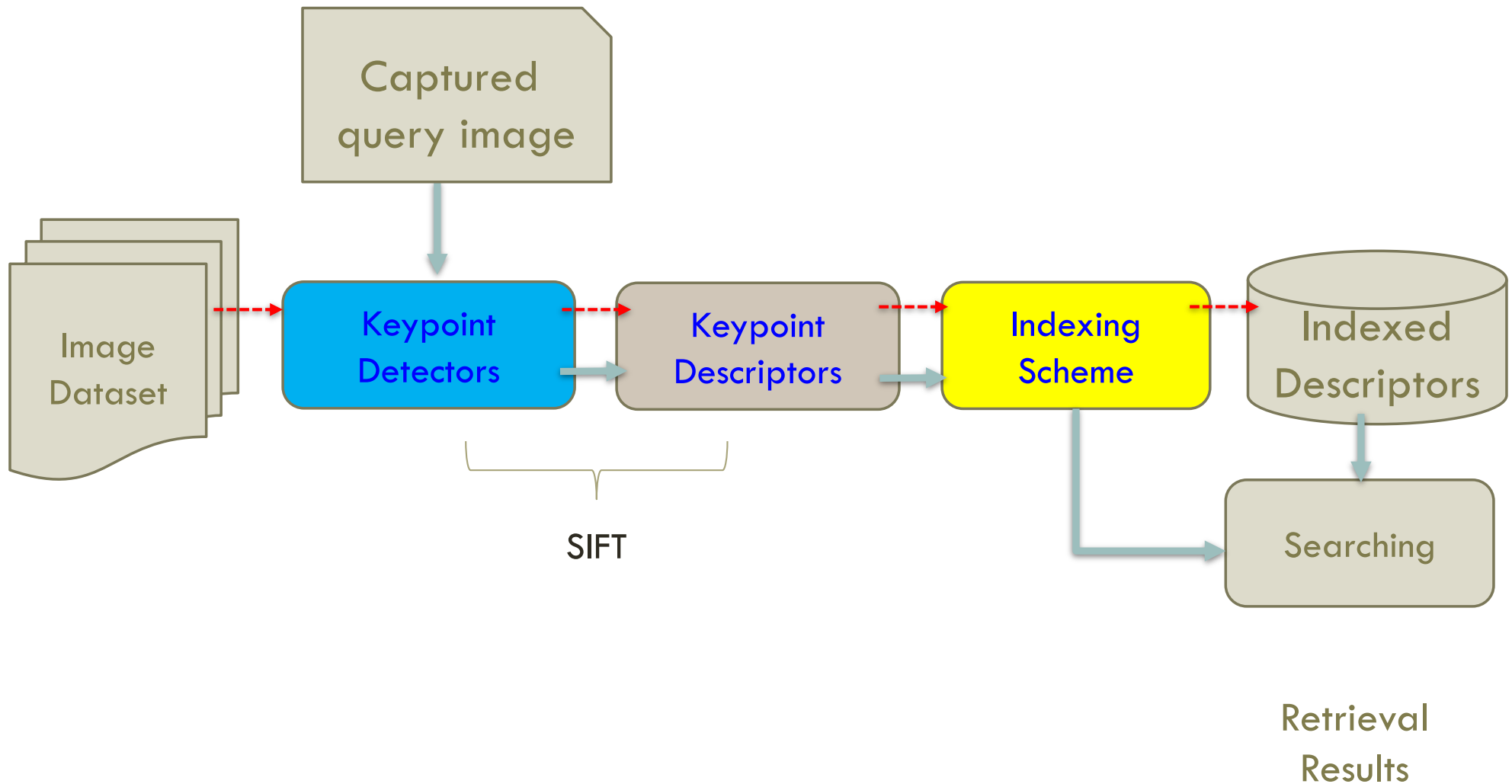




# PRINCIPE APPLIQUÉ À LA RECONNAISSANCE DES FORMES



# PREMIÈRE ÉTAPE : EXTRAIRE DES POINTS D'INTÉRÊTS



# SIFT — EN 3 ÉTAPES

## Detection

- Détecter les points qui sont intéressants à plusieurs échelles
- Extraire la region autour du point d'intérêt (patch visuel)
- Un patch visuel est un élément du dictionnaire visuel

## Description

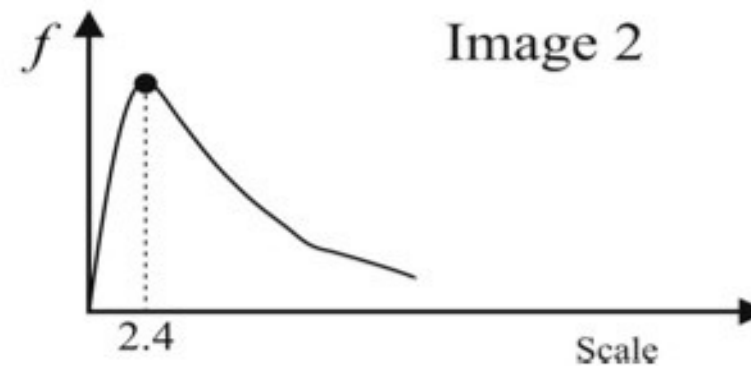
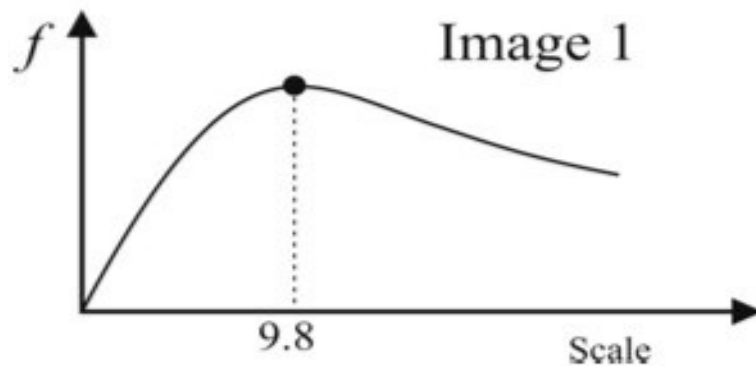
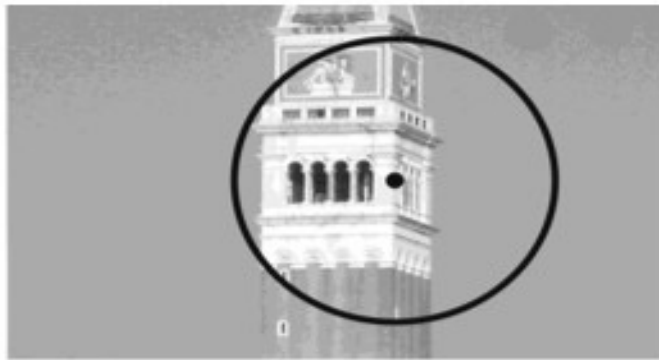
- Construire un descripteur pour la zone
- Assigner les orientations principales de la region détectée

## Matching

- Chercher les zones similaires pour indiquer qu'elle correspond à une zone de l'image dans la base de données

# 1. DÉTECTION DE KEYPOINTS

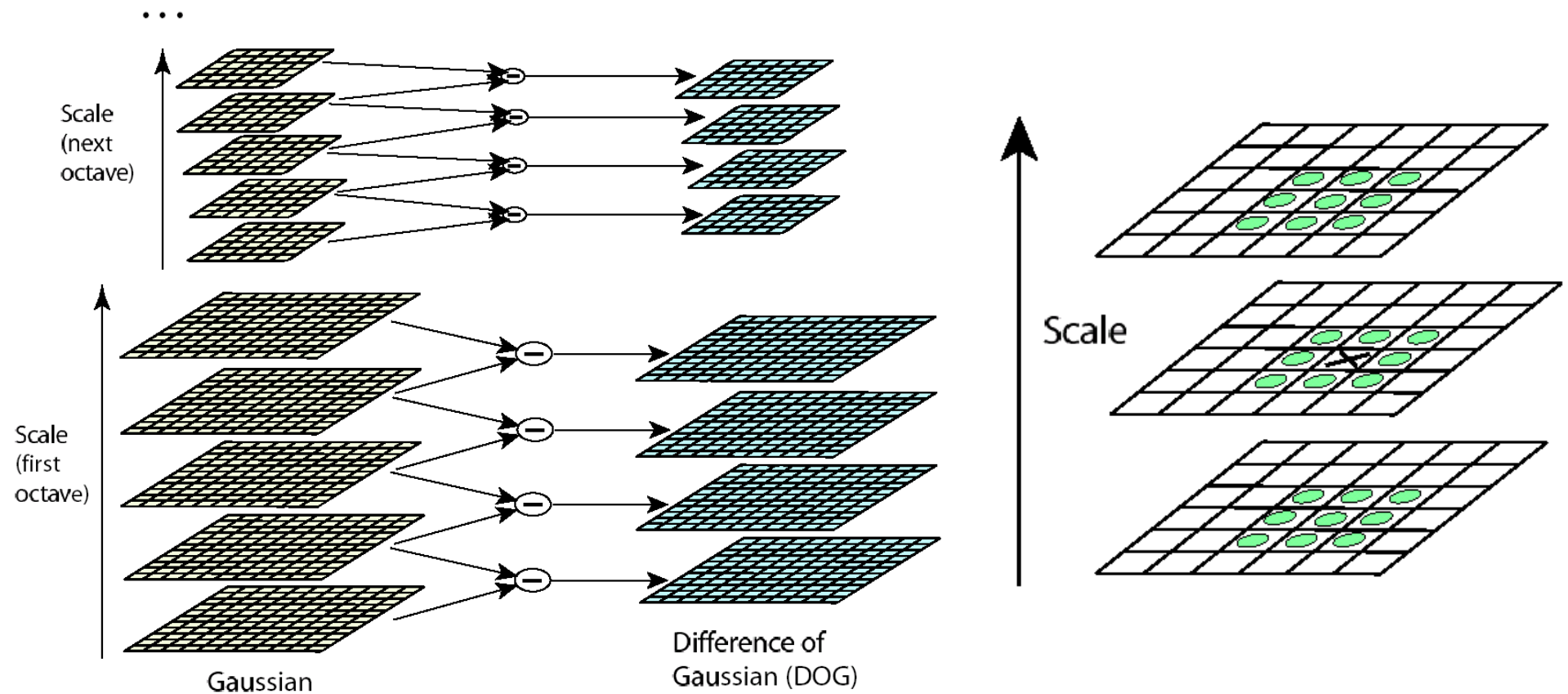
- Les changements brusques de niveaux de gris sont de bonnes fonctions
- Peut être calculé par un “Laplacian de Gaussienne” (LoG)



# 1. DÉTECTION DE KEYPOINTS

- L'image subit une convolution avec des filtres Gaussian à différentes échelles
  - On calcule des différences entre des filtres successifs (soustractions entre des résultats de filtres différents)
  - Les Keypoints correspondent aux maxima/minima dans ces soustractions appelées "Difference of Gaussians (DoG)"
- En pratique :
  - On compare chaque pixel de l'image "DoG" avec ses 8 voisins à la même échelle
  - On compare chaque pixel avec ses 9 voisins des échelles supérieure et inférieure
- Si le pixel est un maximum ou un minimum parmi tous ces voisins
  - c'est un keypoint

# 1. DÉTECTION DE KEYPOINTS





# 1. DÉTECTION DE KEYPOINTS

## Exemple

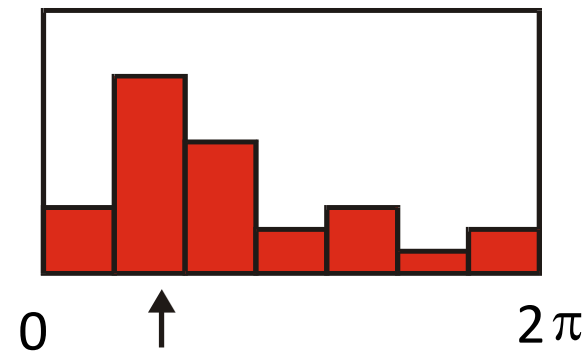
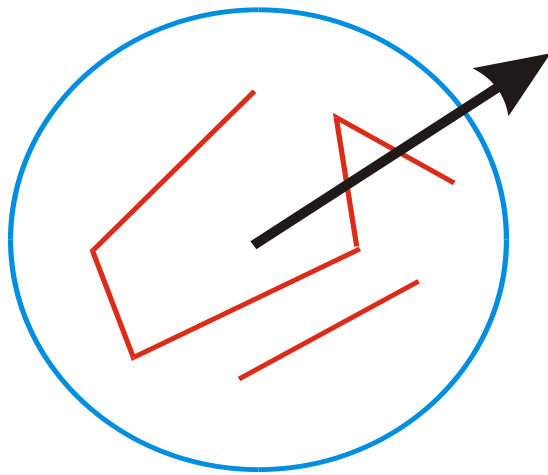


- (a) 233x189 image
- (b) 832 DOG extrema
- (c) 729 left after peak value threshold
- (d) 536 left after testing ratio of principle curvatures

## 2. DESCRIPTION DE LA ZONE

Attribuer les orientations présentes dans une region

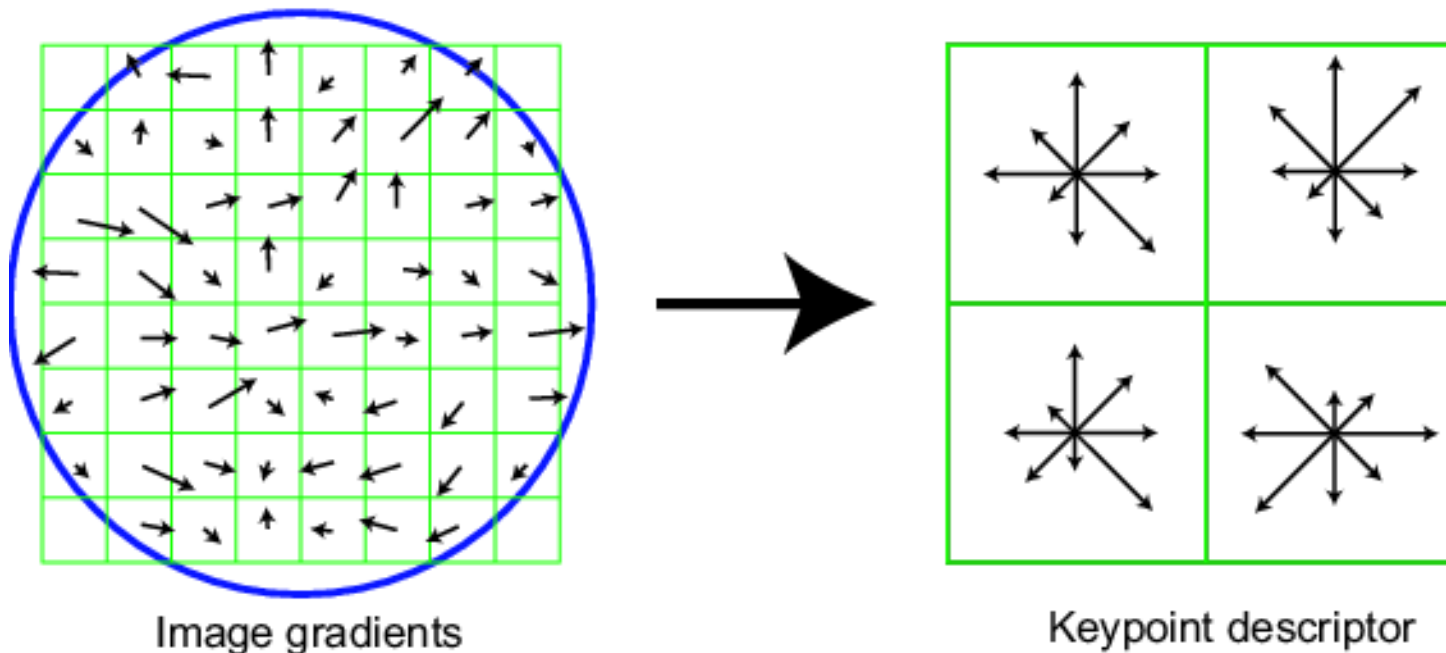
- Créer un histogramme des orientations des gradients de la zone
- Choisir l'orientation principale pour décrire la zone



## 2. DESCRIPTION DE LA ZONE

### Construction du descripteur SIFT

- Créer un tableau avec tous les orientations (zones)
- Au final, un vecteur de 128 valeurs
  - 8 orientations décrites par un tableau de 4x4 histogram



# PARTIE 2

## DE LA DESCRIPTION À SON INDEXATION



**Laboratoire Informatique Image Interaction (L3i)**

Université de La Rochelle - Pôle Sciences et Technologie - Avenue Michel Crépeau - 17042 LA ROCHELLE CEDEX 1 France

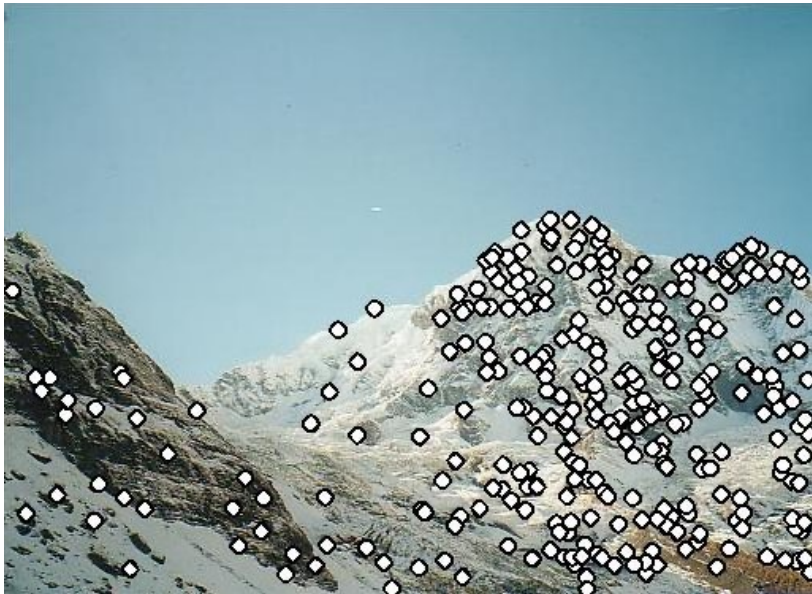
Tél : +33 (0)5 46 45 82 62 – Fax : 05.46.45.82.42 – Site internet : <http://l3i.univ-larochelle.fr/>

# RETROUVER DES POINTS SIMILAIRES

Pour chaque Keypoint de l'image A

- On recherche le point le plus similaire dans l'image B

**A**



**B**



# RETROUVER DES POINTS SIMILAIRES

Possibilité d'utiliser la recherche des plus proches voisins mais...

- Beaucoup de points d'intérêts dans une image / un document
- Chaque point a son propre vecteur de dimension 128

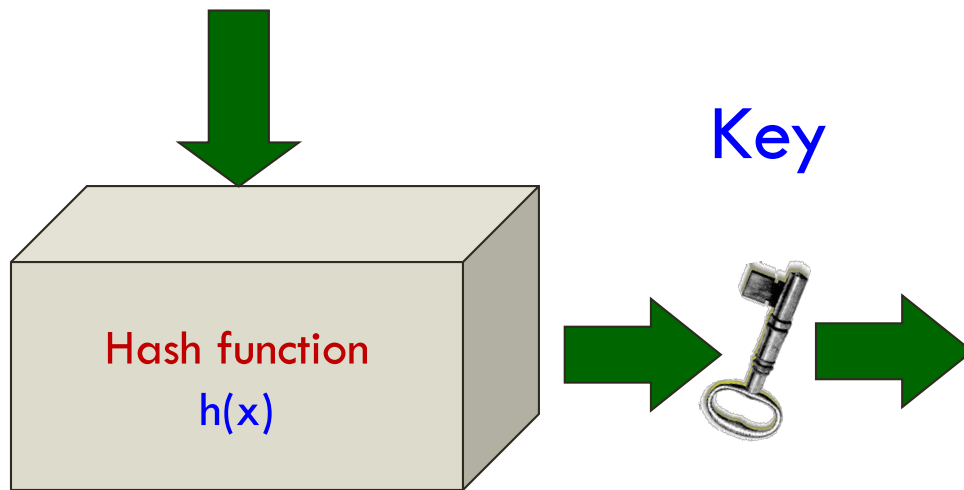
Si l'on a une base de 1000 documents avec chacun 1000 keypoints

- Base d'1 million de points d'intérêts
- Comment retrouver le plus proche ?
- Comment le retrouver vite ?

Besoin d'indexer les vecteurs !

# INDEXATION À PARTIR D'UNE TABLE DE HACHAGE

Data-item:  $x$



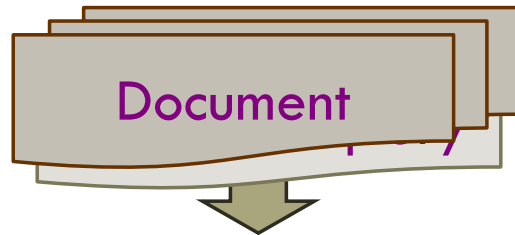
(Table de Hachage)



- Généralement, on voudrait une fonction de hachage  $h(x)$ 
  - Don't les éléments similaires soient stockés dans la même case
  - Eviter les colisions (zones visuellement différentes dans la même case)
  - Rapide à calculer
- Les fonctions de hachage rassemblent ces propriétés

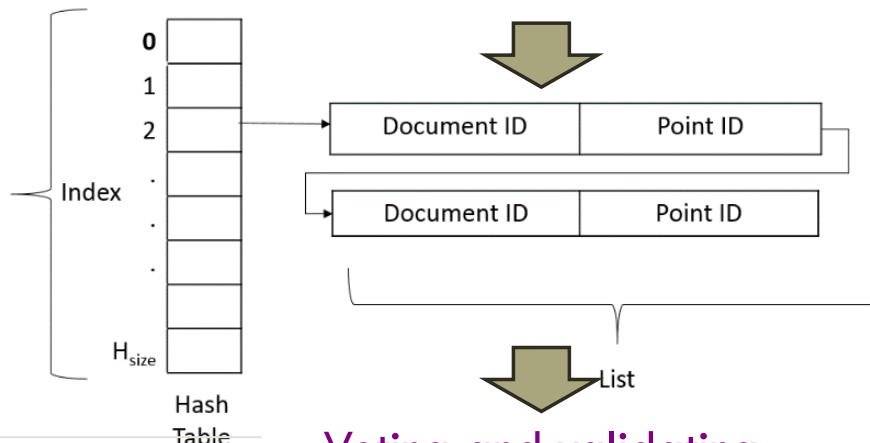


# INDEXATION À PARTIR D'UNE TABLE DE HACHAGE ?

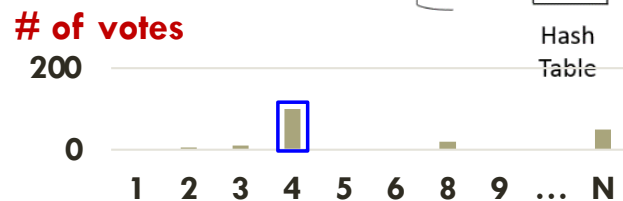


Extraire l'ensemble des vecteurs des points d'intérêt

$$H_{index} = \left( \sum_{i=0}^{d-1} f_i q^i \right) \bmod H_{size}$$



- $d$  est le nombre de valeurs du descripteur du point d'intérêt de l'image
- $q$  est le niveau constant de quantification
- $H_{size}$  est la taille de la table de hachage



Voting and validating

Result



# PARTIE 3

## DE L'INDEXATION À LA RECHERCHE D'INFORMATION



**Laboratoire Informatique Image Interaction (L3i)**

Université de La Rochelle - Pôle Sciences et Technologie - Avenue Michel Crépeau - 17042 LA ROCHELLE CEDEX 1 France

Tél : +33 (0)5 46 45 82 62 – Fax : 05.46.45.82.42 – Site internet : <http://l3i.univ-larochelle.fr/>

# DE LA TABLE DE HACHAGE À L'INFORMATION

Une fois que les points d'intérêt ont été indexés...

- Possibilité de recherche quel(s) document(s) dans la base possède le même keypoint
- Appliquer cela à l'ensemble des Keypoints

Il est alors nécessaire de faire une table de vote pour identifier le document dans la base de données qui a le plus de keypoints en commun avec la requête

- Simple histogramme
- Recherche de la valeur maximale

# MISE EN PRATIQUE



**Laboratoire Informatique Image Interaction (L3i)**

Université de La Rochelle - Pôle Sciences et Technologie - Avenue Michel Crépeau - 17042 LA ROCHELLE CEDEX 1 France

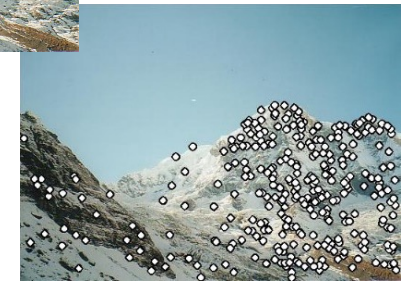
Tél : +33 (0)5 46 45 82 62 – Fax : 05.46.45.82.42 – Site internet : <http://l3i.univ-larochelle.fr/>

# INDEXATION À PARTIR D'UNE TABLE DE HACHAGE ?

Image à indexer / rechercher



Extraction des points d'intérêts



Calcul des descripteurs des keypoints



Indexation des keypoints pour faciliter leur comparaison



Vote du document le plus probable (avec le plus de votes)

# of votes

200

0

1 2 3 4 5 6 8 9 ... N



## POUR LE TP

- Deux groupes dans une même équipe
  - Chaque groupe correspond à un binôme (ou monome) qui devra travailler sur une des deux parties
    - Extraction de keypoints et de leurs descripteurs
    - Indexation et recherche du document le plus similaire

# POUR LE TP — QUELQUES RESSOURCES

- Introduction à SIFT avec OpenCV et Python
  - [https://docs.opencv.org/master/da/df5/tutorial\\_py\\_sift\\_intro.html](https://docs.opencv.org/master/da/df5/tutorial_py_sift_intro.html)
- Lire une vidéo à partir d'OpenCV en Python
  - [https://opencv-python-tutroals.readthedocs.io/en/latest/py\\_tutorials/py\\_gui/py\\_video\\_display/py\\_video\\_display.html](https://opencv-python-tutroals.readthedocs.io/en/latest/py_tutorials/py_gui/py_video_display/py_video_display.html)
- Un exemple d'utilisation d'une table de hachage
  - <https://www.pyimagesearch.com/2017/11/27/image-hashing-opencv-python/>
  - Attention, ce n'est pas la bonne fonction de hachage ;-)
- Jeu de données fournies sur Moodle
  - Issues d'un livre et d'une base de documents administratifs
- Composé de deux types de contenus
  - Les images des documents
  - Des vidéos qui simulent la capture du document avec un smartphone ou une webcam