

Analyse et indexation d'images et de vidéos dans des grandes bases Multimédia

Cours n°1

2013-2014

Frederic Precioso

Ce cours s'appuie sur le cours de Nicolas Thome (LIP6 –UPMC)

Introduction à la Recherche d'Informations (RI)

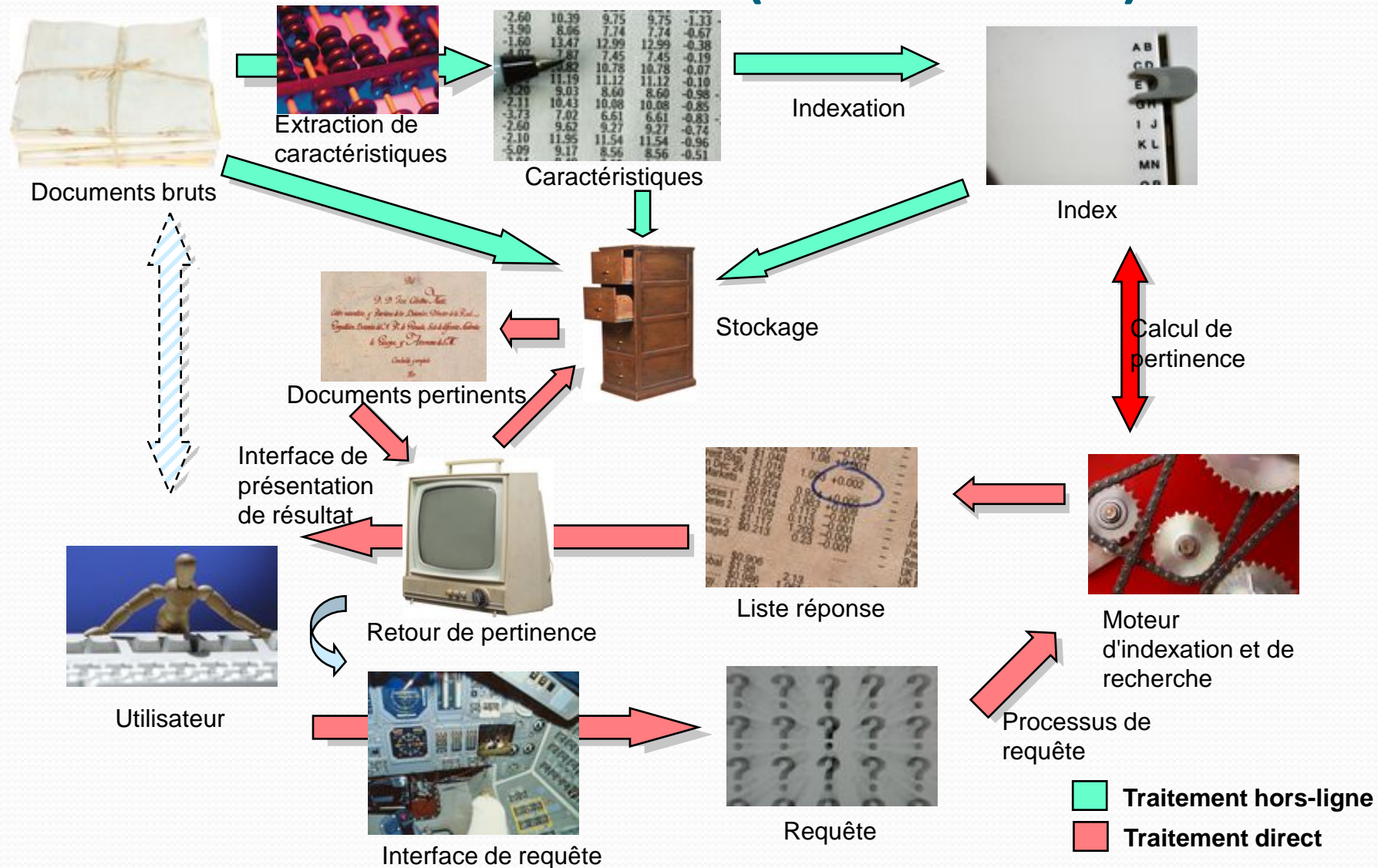
- Quels sont les verrous dans les applications de recherche d'image par le contenu ?
- Problématique : comment représenter efficacement le contenu visuel d'une image ?

Les Systèmes de RI

Un **modèle de RI** a pour but de formaliser:

- La représentation du document
 - Le document en lui-même est alors ignoré
- La représentation des requêtes
 - Expressivité du langage de requêtes
 - Formalisation du **besoin d'information**
- Une **fonction d'ordonnancement** associant à chaque document un **classement (rang)** dans le contexte de la requête
 - Souvent implicite dans la définition de la pertinence (rang = classement par pertinence)

Tâches de la RI (workflow)



Composants d'un système de Recherche d'Information (RI)

- Extraction des caractéristiques
 - Transformer les données en une forme compréhensible par la machine
- Indexation
 - Création d'un résumé par document
 - Structuration des signatures en index, stockage
- Moteur
 - Similarité
 - Calcul de la pertinence en ligne
- Interfaces
 - Requête
 - Résultat
 - Acquisition du retour de pertinence



Composants d'un système RI

- ***Extraction des caractéristiques***
 - *Transformer les données en une forme compréhensible par la machine*
- ***Indexation***
 - *Création d'un résumé par document*
 - *Structuration des signatures en index, stockage*



- Quelles sont les primitives adéquates ?
- Quels sont les descripteurs les plus adaptés ?
- Comment extraire une représentation vectorielle par image ?

Composants d'un système de RI

- Moteur

- Similarité
- Calcul de la pertinence en ligne



- Interfaces

- Requête
- Résultats
- Acquisition du retour de performance



Plan

I. Généralités sur la recherche par le contenu

- De la RI à la recherche par le contenu

II. Etat de l'art en Indexation d'images

- Primitives images
- Descripteurs images
- Formation d'index

III. Modèle Bag of Word (BoW)

- a. Détecteur dense
- b. Dictionnaire visuel : K- Means
- c. BoW : génération histogramme (+info spatiale)
 - Coding + pooling

De la RI à la recherche par le contenu

1. Motivations de la RI
2. Recherche d'image par le contenu
3. 2 contexte d'études classiques
 - a. Recherche de cible
 - b. Recherche de catégories

RI : Motivations

- Avalanche de doc. Multimédia numériques
- Flot croissant d'information multimédia ...
 - Augmentation de la capacité de stockage et de transmission



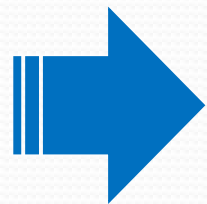
RI avant et après l'ère du numérique

RI : Motivations (Big Data)

- Flot croissant d'information multimédia...
- Méga-banques de données. Quelques chiffres :
 - **YouTube** = en mai 2010, plus de vidéos déposées en 60 jours que ce que les 3 plus grands réseaux de diffusion US ont créé 60 ans [1]; 20 h de vidéo déposées chaque minute en Mai 2009, 24 h chaque minute en mars 2010, pour atteindre **1 h par minute** à l'heure actuelle [1, 4]
 - **Flickr** = 3000 images déposées chaque minute pour atteindre 5 milliards d'images en Septembre 2010 [1]
- L'univers digital était d'environ 281 exaoctets (281 milliards de gigaoctets) en 2007; il a grossi 10 fois depuis [2]
- Images et vidéo, acquises par plusieurs milliards de périphériques (appareils photo, webcams, téléphones) représentent la part la plus importante de ce gigantesque volume de données [3]
- [1] [website-monitoring.com](http://www.website-monitoring.com)
- [2] "digital universe.", <http://eon.businesswire.com/releases/information/digital/prweb509640.htm>
- [3] "major source.", <http://www.emc.com/collateral/analyst-reports/diverse-exploding-digitaluniverse.pdf>
- [4] <http://www.onehourpersecond.com/>

RI : Motivations (Big Data)

- ... mais une « faible » capacité à traiter ces documents
 - Accès, sélection
 - Comparaison, classement
 - Recherche d'information (Moteur de): retrouver une image similaire à ...
 - Fouille de données : structurer une grosse base en groupes homogènes



Nécessité de représentations des données et de méthodes de recherche efficaces

De la RI à la recherche par le contenu

1. Motivations de la RI
2. Recherche d'image par le contenu
3. 2 contexte d'études classiques
 - a. Recherche de cible
 - b. Recherche de catégories

Introduction à la recherche d'images

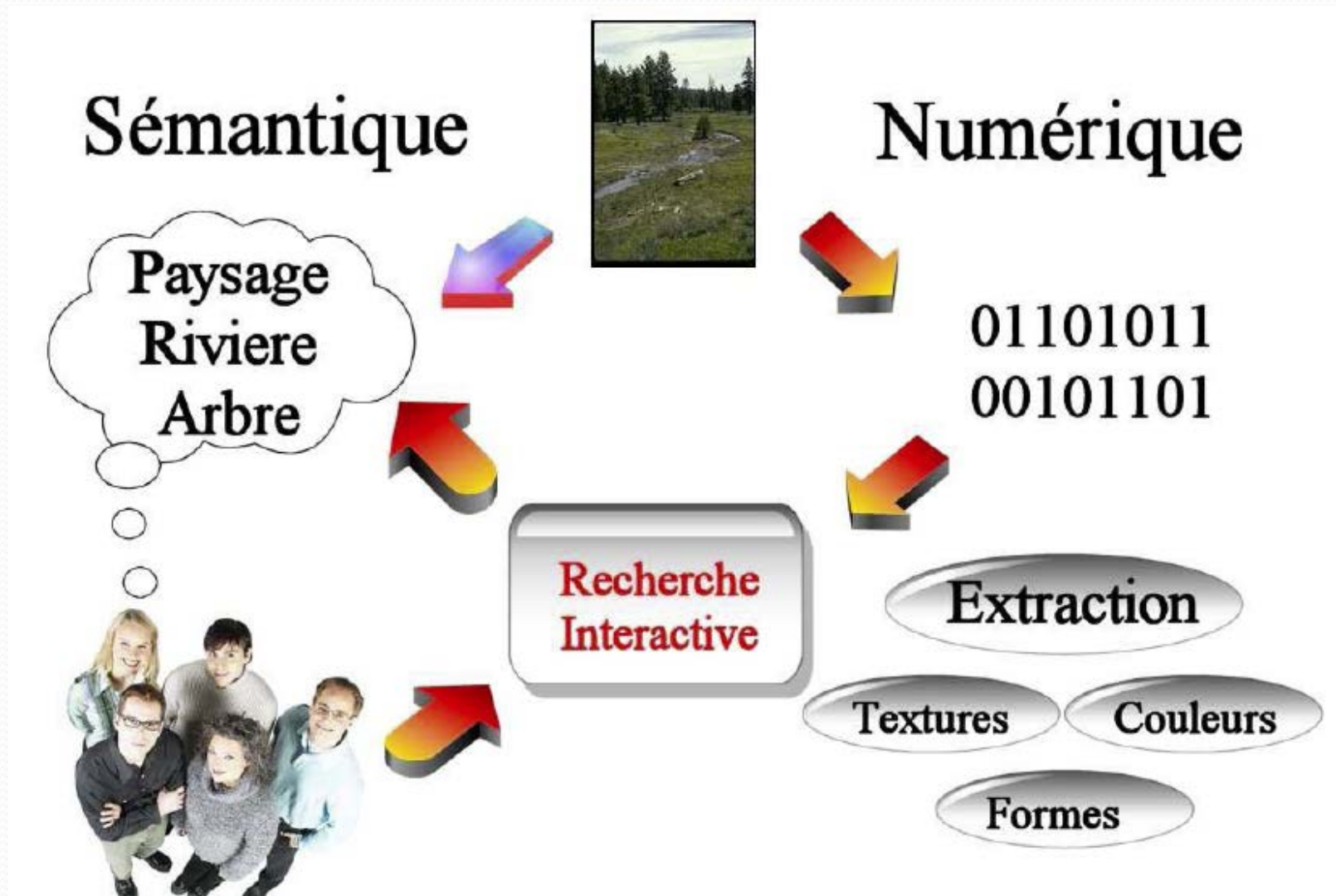
1. Recherche textuelle par mots clés

- Index textuel pour chaque image
- Recherche à partir de mots clés
- Problèmes
 - Disponibilité de l'information: coût de l'indexation
 - Définition du dictionnaire
 - Mise à jour

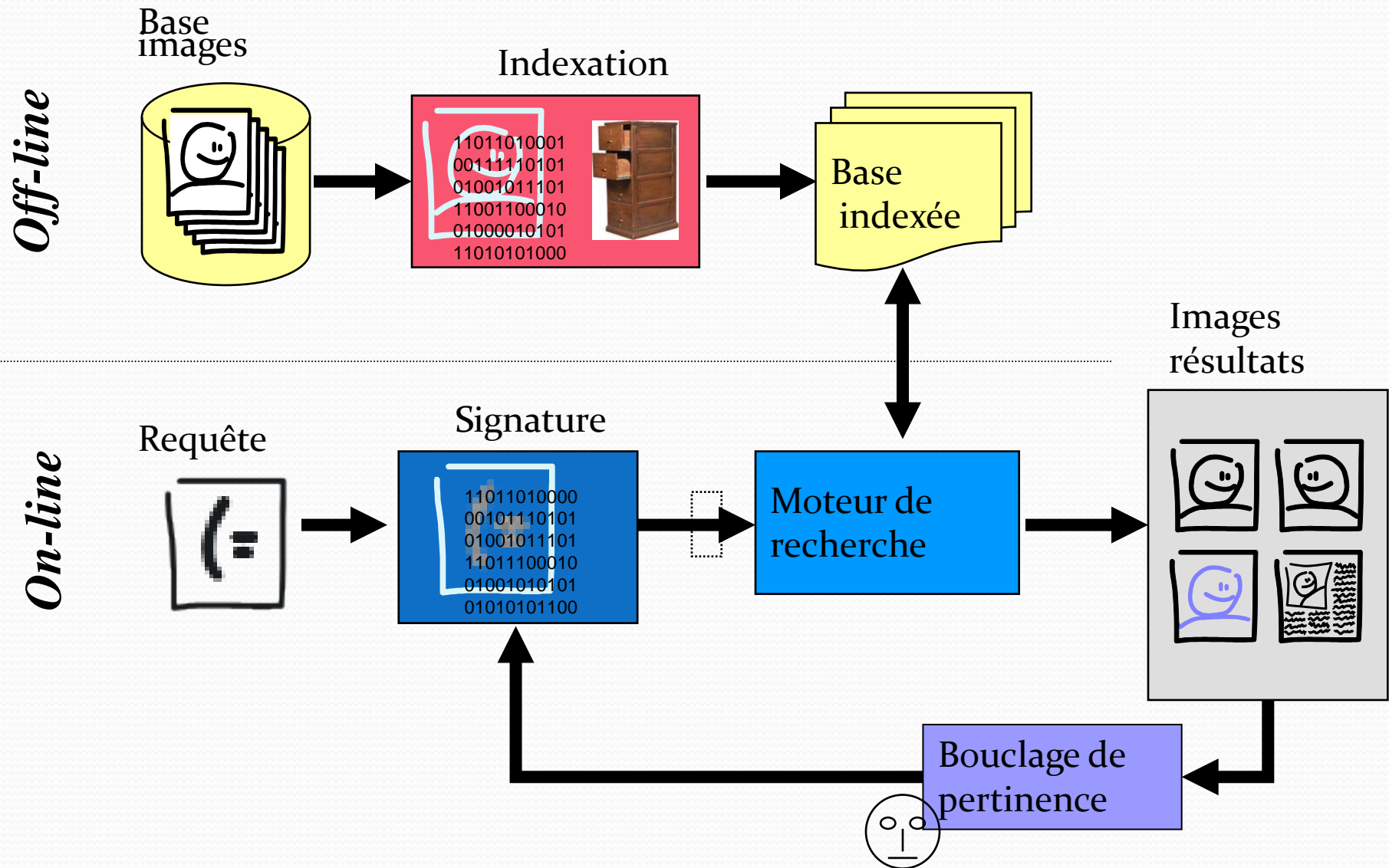
2. Alternative: recherche textuelle dans le contexte

- Indexation des légendes, texte proche de l'image, nom du fichier
- Problèmes: ambiguïtés, synonymes, etc...
 - Exemple: recherche d'images de [Google.com](https://www.google.com)
 - Indexation à partir des tag d'utilisateurs sur les sites collaboratifs (type Flickr)

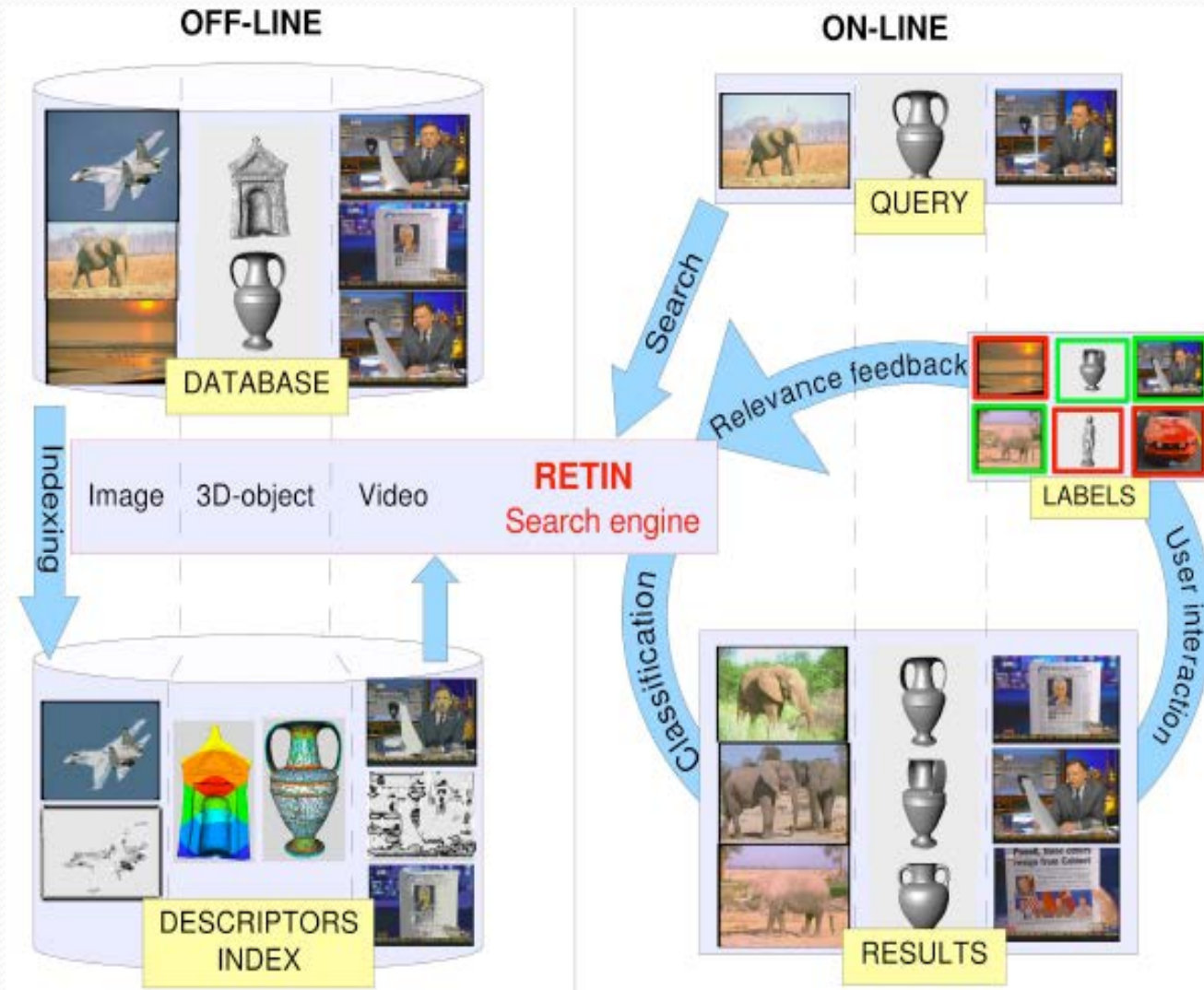
Recherche d'images : le fossé sémantique



Architecture des Systèmes de CBIR



Architecture des Systèmes de CBIR



De la RI à la recherche par le contenu

1. Motivations de la RI
2. Recherche d'image par le contenu
3. 2 contexte d'études classiques
 - a. Recherche de cible
 - b. Recherche de catégories

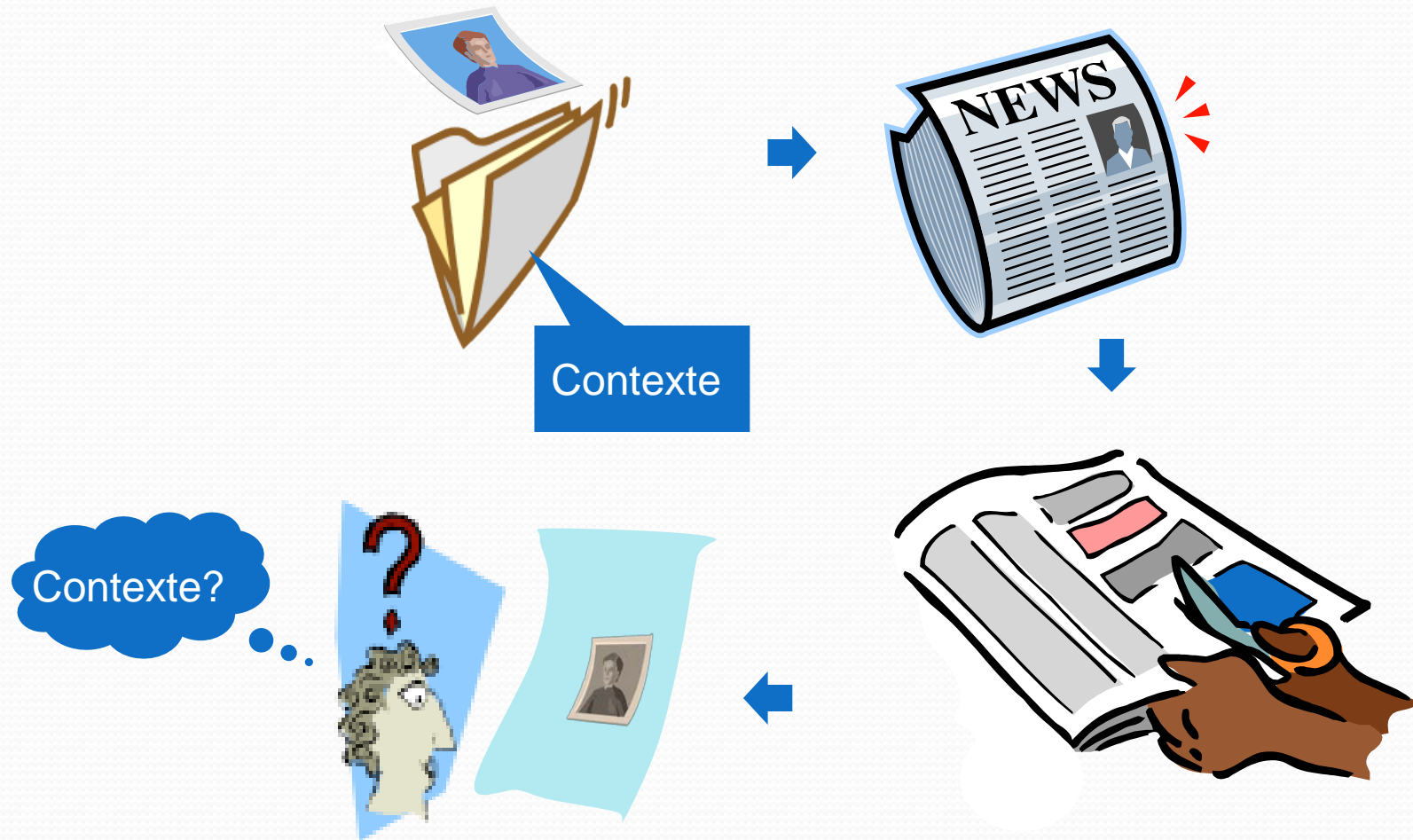
RI : différents Contextes de Travail

Type de recherche, 2 paradigmes :

1. Recherche de cibles ou identification d'images (un tableau, un objet)
 - Trouver le même document, mais (légèrement) déformé
2. Recherche de catégories (une classe d'objets, un concept)
 - Trouver les documents qui contiennent le même concept
 - Problème du fossé sémantique

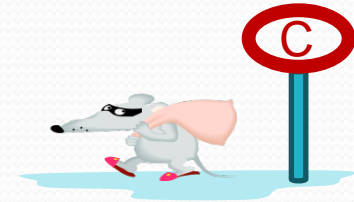
1er paradigme : identification d'image

- Trouver l'original d'un document dont on a « perdu » la référence



1^{er} paradigme : identification d'image

- Détection de copies



requête



Base



1^{er} paradigme : identification d'image

- « Near-duplicate matching »
 - iTOWNS
 - Télévision : INA, dailymotion
 - Moteurs Web



1^{er} paradigme : identification d'image : Projet ANR *iTOWNS*



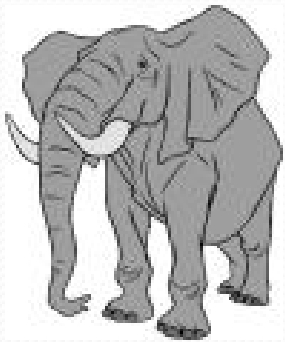
image-based Town On-line Web Navigation and Search engine

1^{er} paradigme : identification d'image

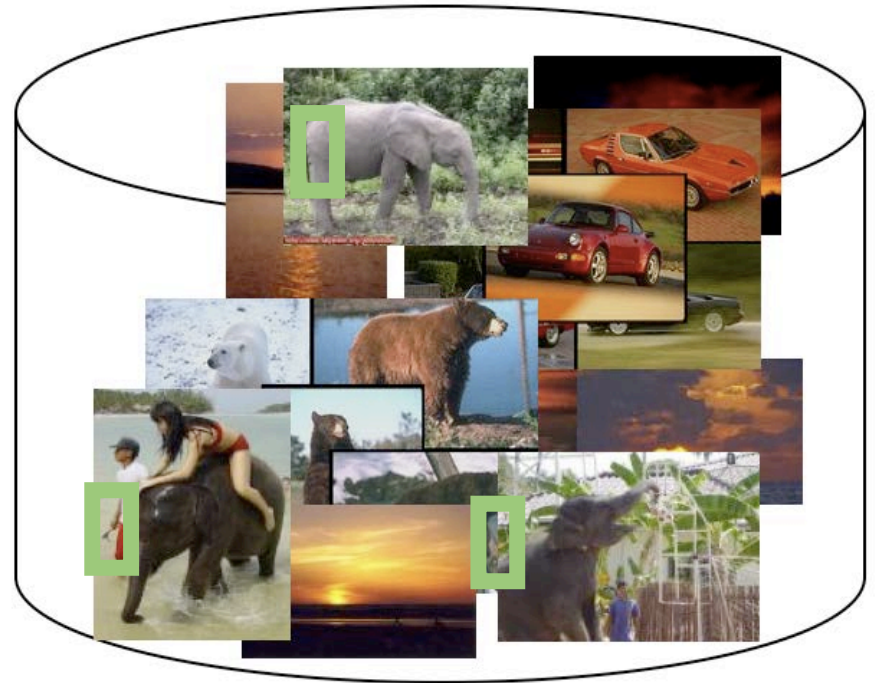
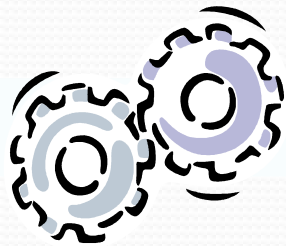
- Dans ce contexte, on veut une représentation
 - Très discriminante : on cherche le même objet
 - Robuste à des transformations
 - Géométriques
 - photométriques,
 - Point de vue
 - ...

2nd paradigme : recherche de catégories

- Recherche de classe d'objets



Requête



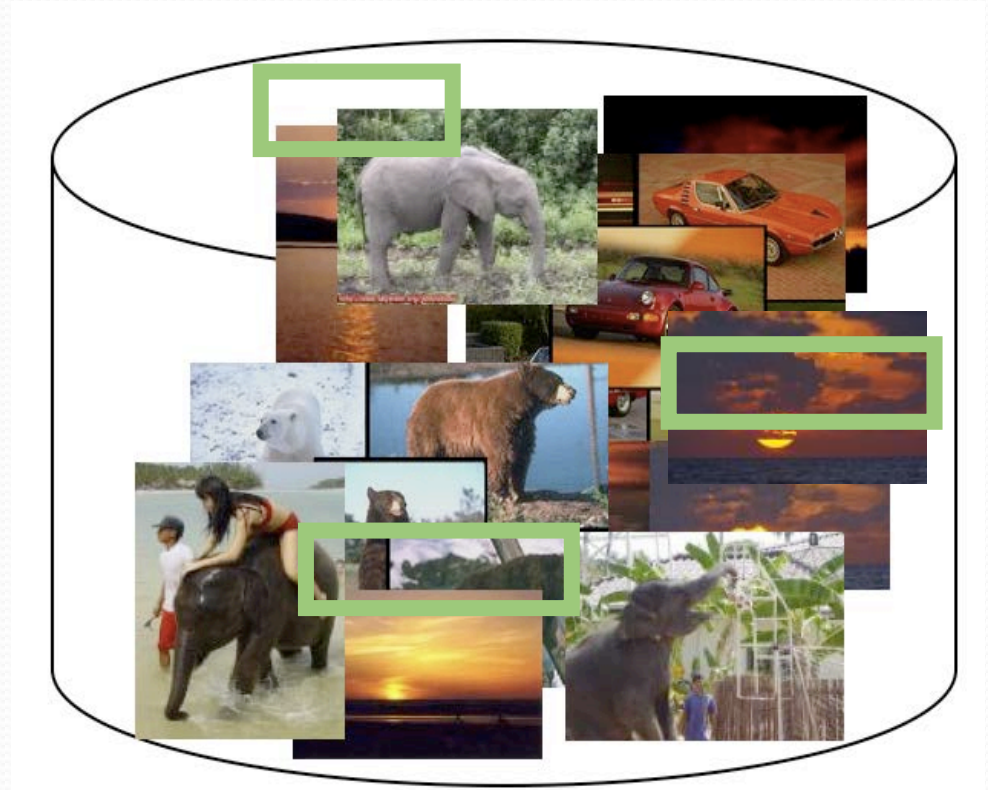
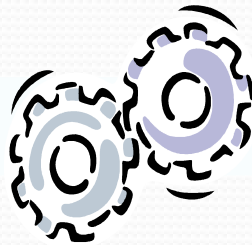
Base d'images

2nd paradigme : recherche de catégories

- Recherche de catégories sémantiques
 - Recherche de *type de scènes*
 - *caractéristiques dites haut niveau*



Requête



Base d'images

2nd paradigme : recherche de catégories

- Dans ce contexte, on veut une représentation
 - Discriminante : on cherche la même catégorie
 - Qui généralise aux autres images de la classe
 - Dans des conditions d'acquisition (géométriques, photométriques, point de vue) qui peuvent être très diverses

2nd paradigme : recherche de catégories

Difficulté de la tâche :

- **Variances intra et inter classes**
 - Var intra doit être faible
 - Var inter doit être grande
- **Conditions d'acquisition**
 - Même Pb que Paradigme 1 mais exacerbé
- **Systèmes**
 - Taille de stockage et temps de calcul
 - Pas nécessairement de catégories connues => recherche en ligne, catégories mixtes, ...

Difficultés : variabilité intra et inter classes

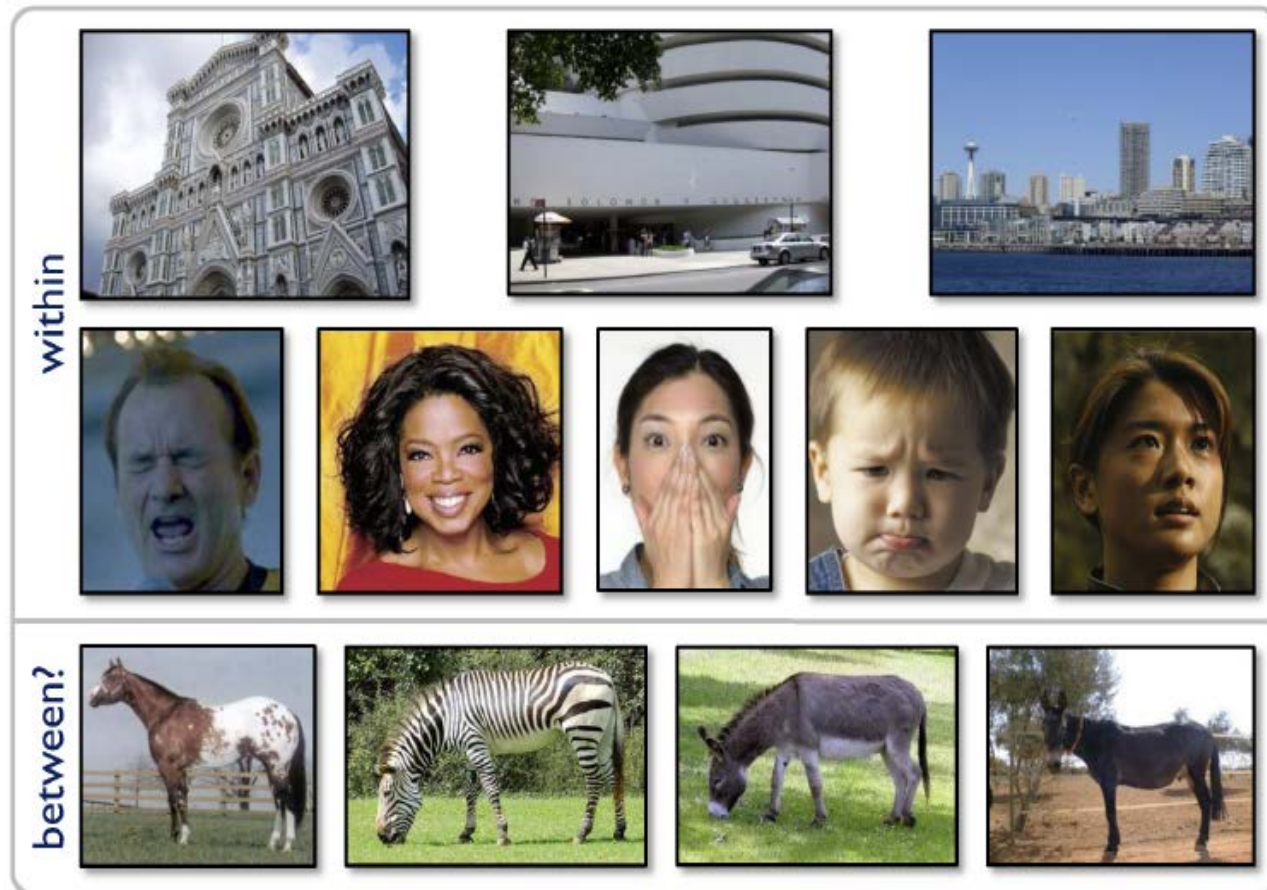


Figure 1.4: Within-class and between-class variability. Above: example images of buildings and faces illustrate the wide within-class variability in appearance of objects from particular categories. Below: depending on the application, horse, zebra, donkey, and mule could be considered (i) different classes, in which case their visual differences are important, or (ii) the same class (*Equidae*), in which case their visual similarities are important.

Difficultés



Figure 1.3: **Sources of visual variability.** Automatic visual recognition systems must deal with visual variability arising from the viewing angle, the pose and articulation of the object, the lighting of the scene, partial occlusions obscuring the object, widely varying scales, and environmental conditions.

Plan

I. Généralités sur la recherche par le contenu

- De la RI à la recherche par le contenu

II. Etat de l'art en Indexation d'images

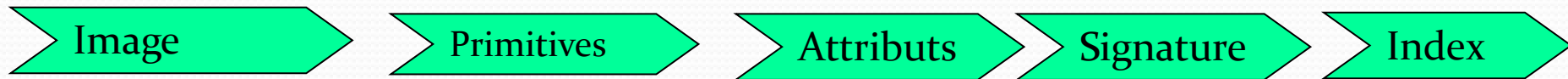
- Primitives images
- Descripteurs images
- Formation d'index

III. Modèle Bag of Word (BoW) : représentation TME

- a. Détecteur dense
- b. Dictionnaire visuel : K- Means
- c. BoW : génération histogramme (+info spatiale)
 - Coding + pooling

Indexation du contenu visuel

- Représentation « structurée » de chaque image de la base
- Terminologie :
 - Signature => relatif au contenu image
 - Index=> relatif à la structure de stockage de l'information (informatique)
- Flowchart du traitement :

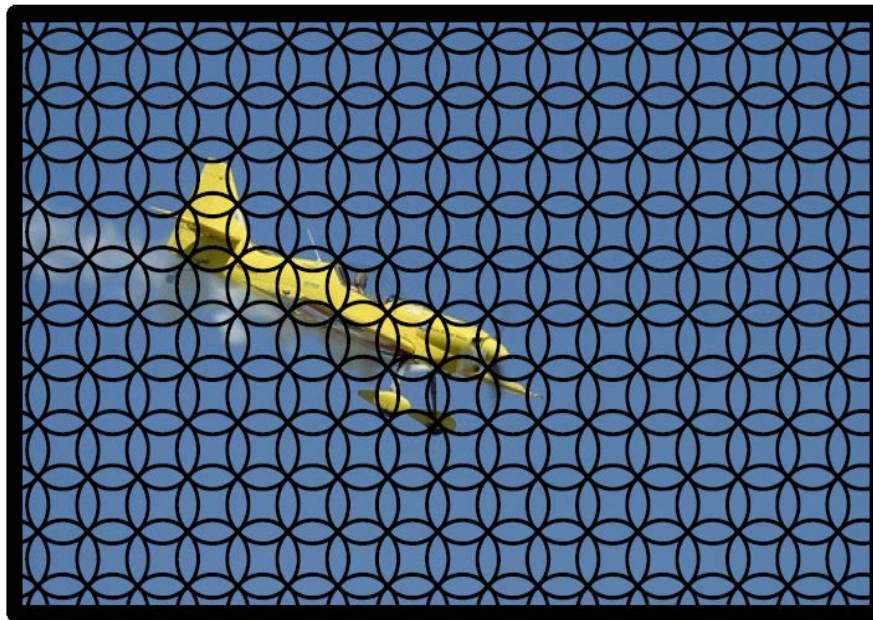


Les primitives Images

1. Les approches « denses »
2. Les approches basées sur le contenu ou « sparse »
 - Extraction de primitives reposant sur le contenu visuel de l'image
 - Détection de points / régions d'intérêt basées sur le contenu de l'image
 - a. *Détection de PoI*
 - b. *Détection multi-échelle*
 - c. *Détection de blobs*
 - d. *Détection de RoI*

Les primitives Images denses

- Pixels, Grille dense
- Régions / pts sélectionnés de manière uniforme
 - Echelle d'analyse fixée
 - Multi-échelle par sélection explicite



Dense sampling detector

Les primitives Images sparse

- Exemple : Régions issues d'une segmentation de l'image



Les primitives Images sparse

- **Défis posés :**

- L'extraction doit être “répérable” (repeatable)
 - décalage (translation), rotation, changement d'illumination, le point est-il encore détecté ?



no chance to match!

- L'extraction doit être précise, bien localisée
- L'extraction doit viser à l'exhaustivité
 - Il faut couvrir suffisamment les zones à décrire
- Les primitives doivent porter de l'information
 - Description exploitable

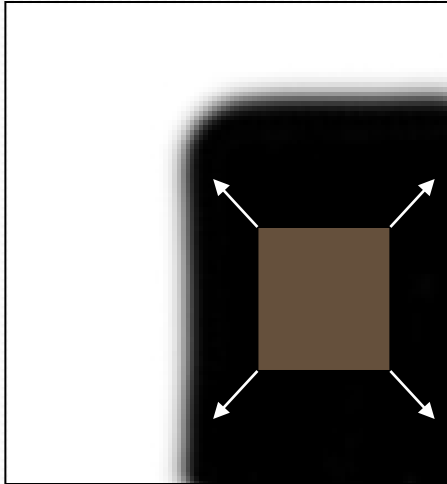
Les primitives Images sparse : points / régions d'intérêt

- Définition :

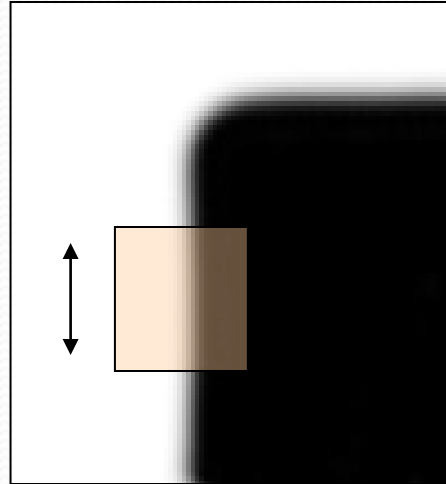
- Contour = discontinuité dans une direction de la fonction d'intensité
- Point d'intérêt (1^{ère} déf.) = discontinuité dans 2 directions de la fonction d'intensité
- Plein de variantes autour de la notion de **saillance** dans les images

Les primitives Images sparse :

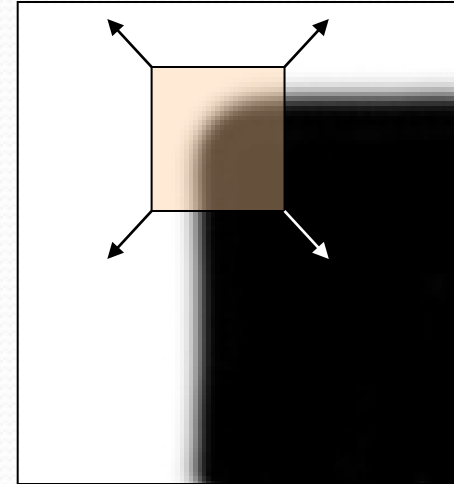
saillance



“flat” region:
pas de
changement
dans
aucune
direction



“edge”:
ne change pas
le long du
contour



“corner”:
changements
significatifs
dans toutes les
directions



Les primitives Images sparse

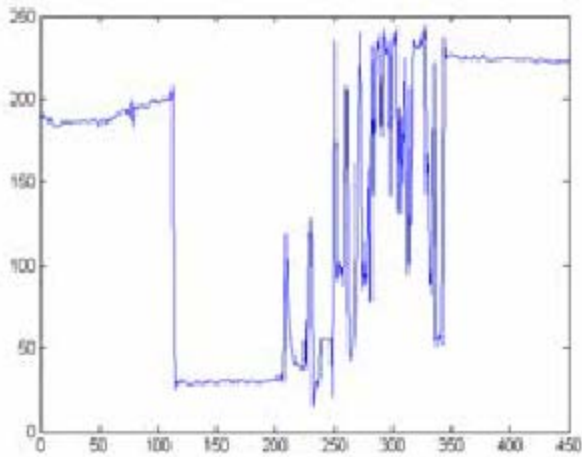
Qu'est-ce qu'un contour ?

- Un contour est une variation brusque d'intensité.



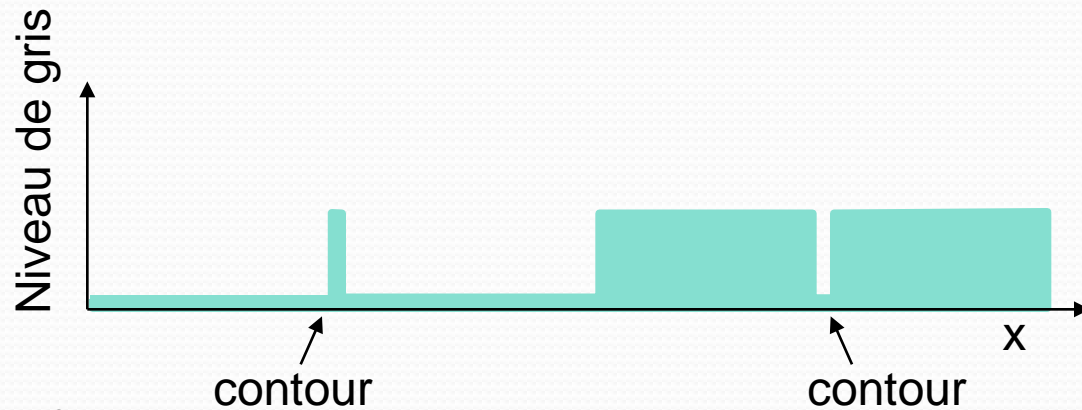
Source : cis.poly.edu/cs664/

Qu'est-ce qu'un contour ?

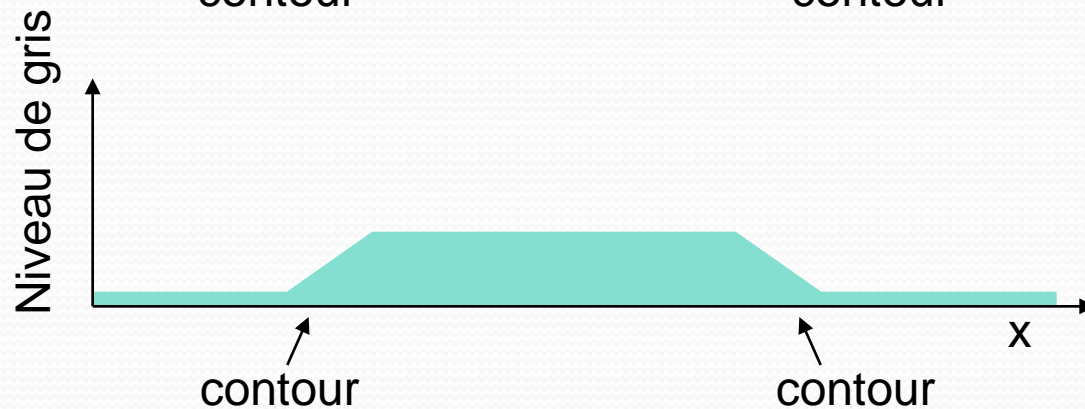


Types de contours dans une image

Marche
d'escalier



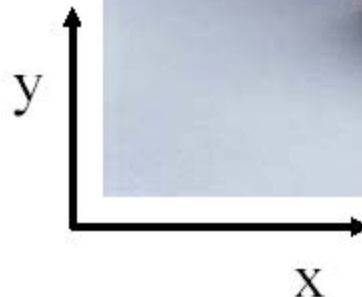
Rampe



Mesure des variations de l'image

- Le gradient (première dérivée) de l'image est l'opérateur de base pour mesurer les contours dans l'image.

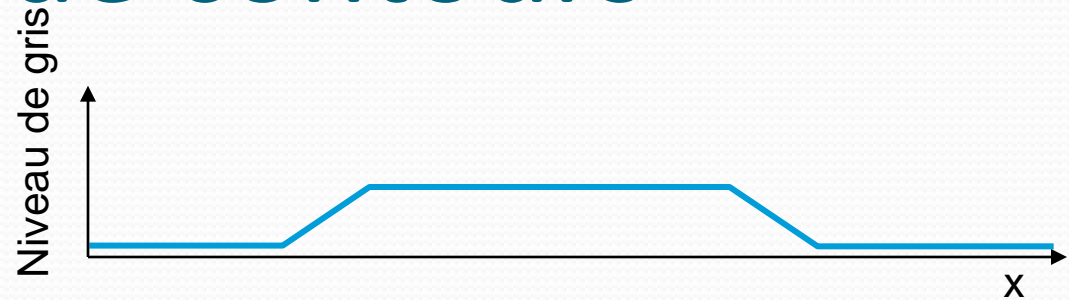
$$|\nabla f| \equiv \left(\frac{\partial f}{\partial x} \right)^2 + \left(\frac{\partial f}{\partial y} \right)^2$$



$f(x, y)$

Détection de contours

Image 1D $f(x)$



1ère dérivée $f'(x)$



$|f'(x)|$



Pixels contours:

$|f'(x)| > \text{Seuil}$



Première dérivée de l'image (Roberts)

- Roberts (1965) fournit une première approximation de la première dérivée d'une image discrète.
- Le calcul se fait avec deux masques de convolution représentant la dérivée dans deux directions orthogonales.

$$\begin{array}{cc} \boxed{0} & 1 \\ -1 & 0 \end{array} \quad \begin{array}{cc} \boxed{1} & 0 \\ 0 & -1 \end{array}$$

I_x I_y

Première dérivée de l'image (Roberts)

- La norme du gradient (plus utilisée) est obtenue par :

$$|\nabla I| = \sqrt{I_x^2 + I_y^2} \approx |I_x| + |I_y|$$

- La direction du gradient (moins utilisée) est obtenue par :

$$\theta = \arctan(I_y / I_x) - 3\pi / 4$$

Opérateurs de détection de contours

- Par la suite, plusieurs autres approximations du gradient discret sont apparus.

- Exemples :

- Prewitt

$$Px = \begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix} \text{ et } Py = \begin{bmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix}$$

- Sobel

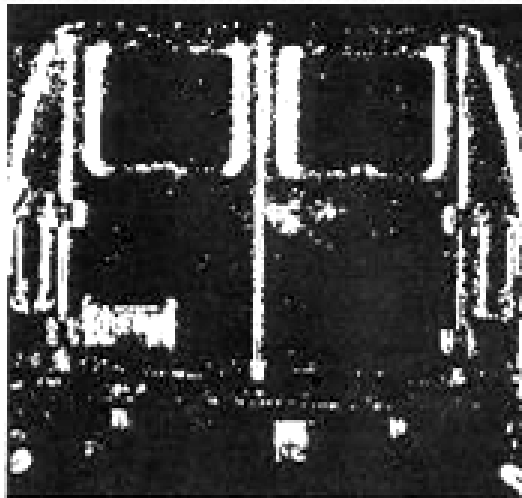
$$Sx = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \text{ et } Sy = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}$$

- Kirsh : 8 masques pour 8 directions
 - Canny, Deriche : optimisation de critères

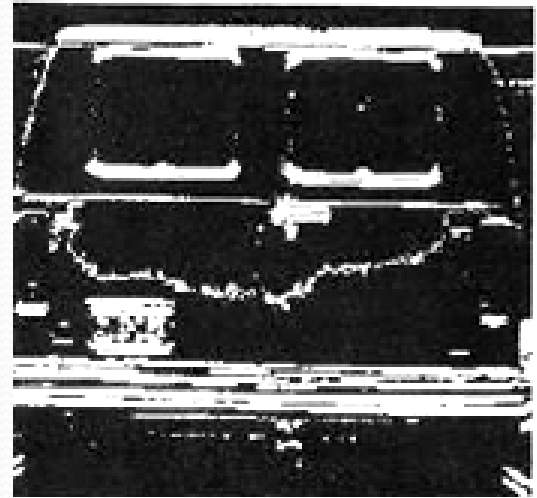
Exemple de détection de contours



Image I

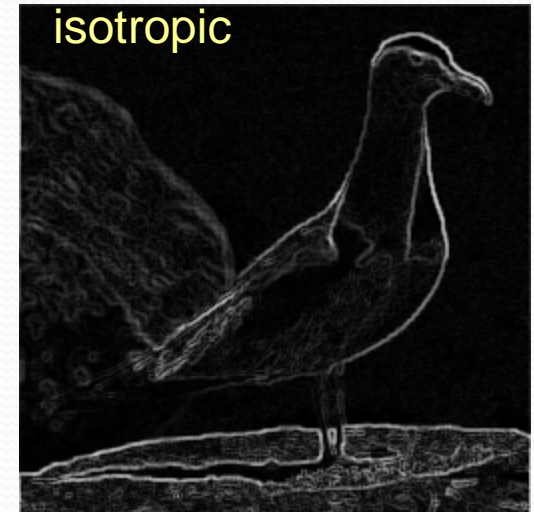
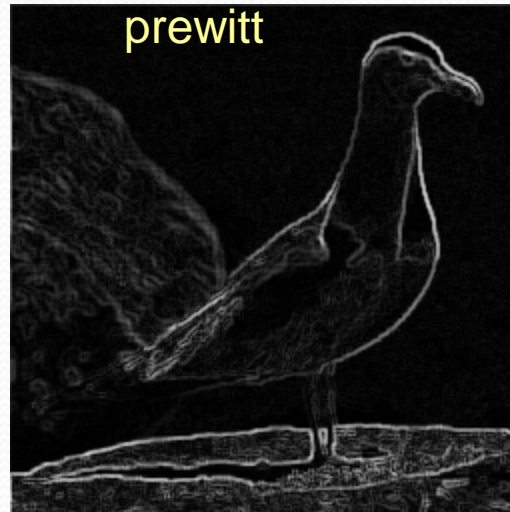
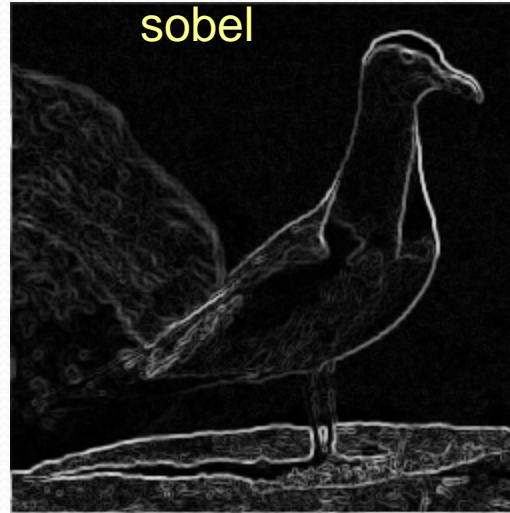


Contours verticaux $\frac{\partial I}{\partial x}$



Contours horizontaux $\frac{\partial I}{\partial y}$

Exemples de détections de contours

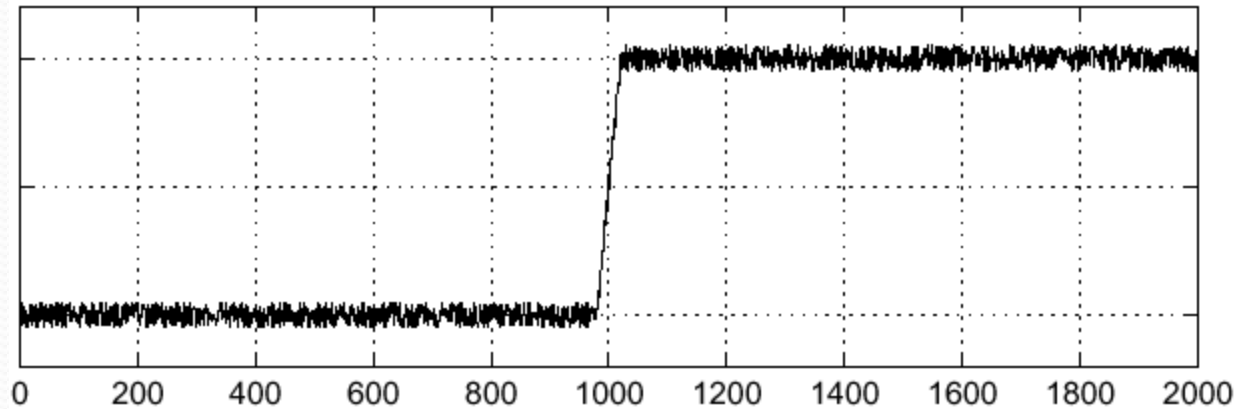


Détecteurs de points de contours

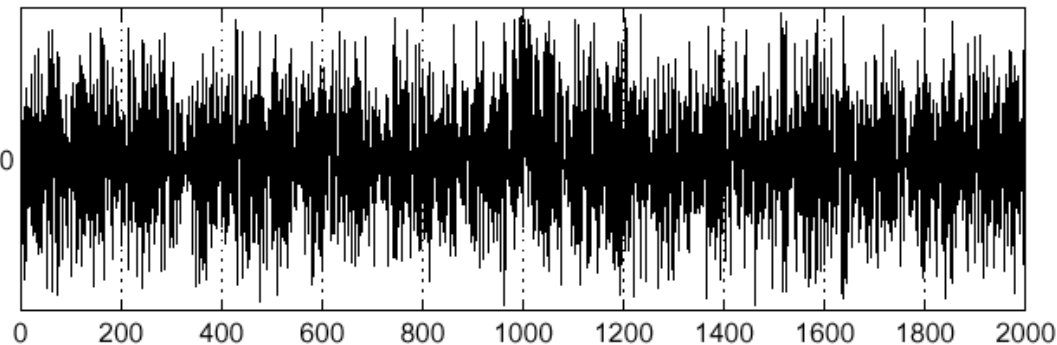
- Il existe plusieurs opérateurs pour détecter les points de contours (dérivée du premier ordre).
- La plupart de ces opérateurs (sauf Roberts) lisse l'image afin d'obtenir un meilleur résultat.
- En pratique, on obtient des contours incomplets
 - il y a des pixels superflus
 - il y a des manques
 - il y a des erreurs dans la position et l'orientation des pixels contours
- Il faudra utiliser d'autres techniques pour compléter les indices de contours obtenus avec ces opérateurs.

Effet du lissage sur la dérivée

$$f(x)$$

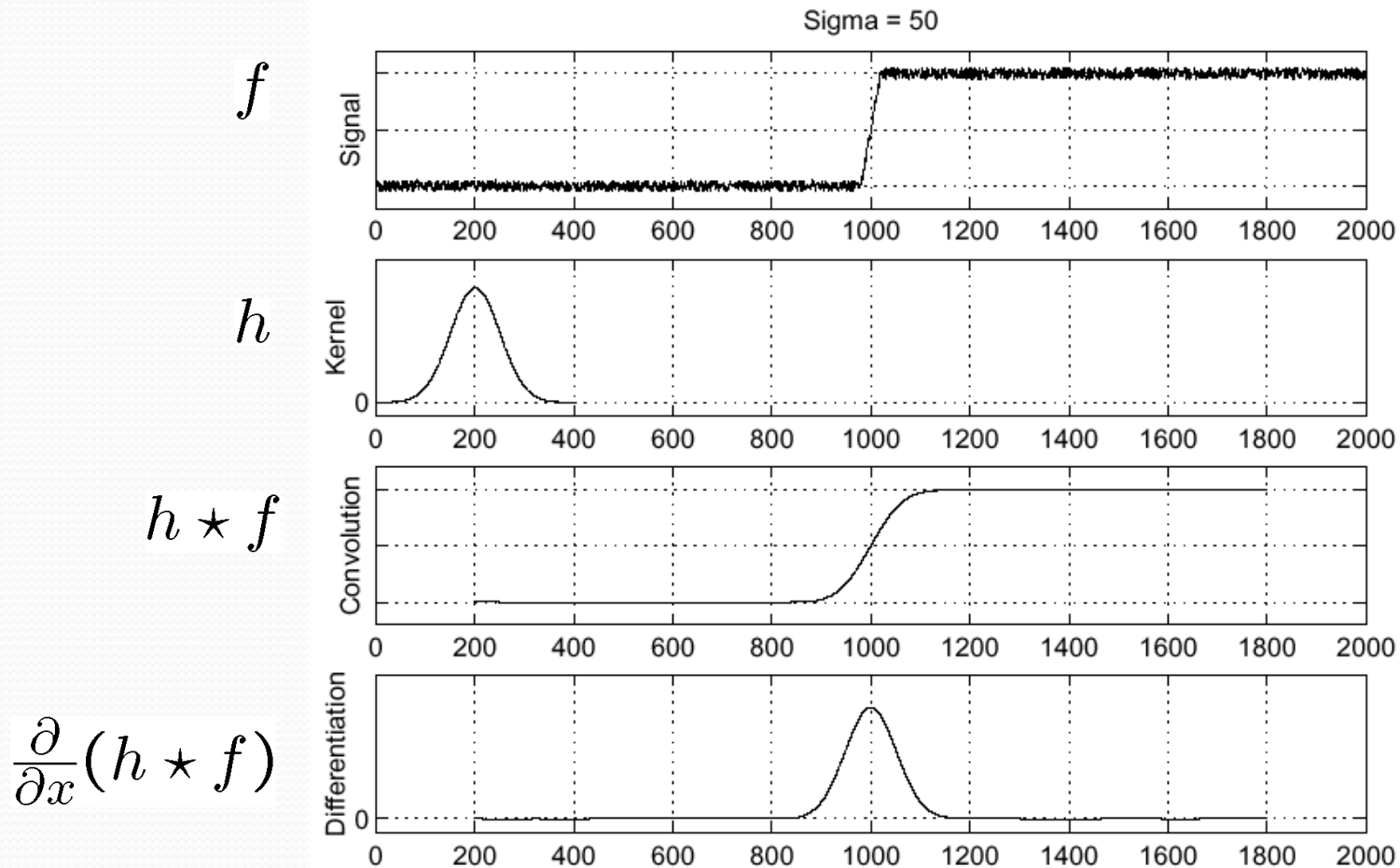


$$\frac{d}{dx}f(x)$$



Where is the edge??

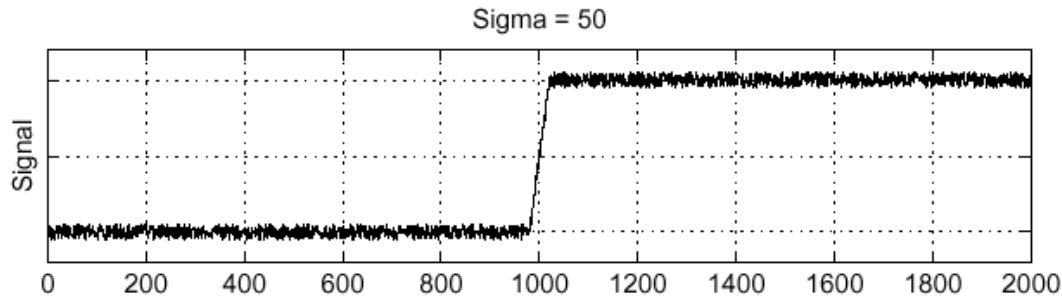
Effet du lissage sur la dérivée



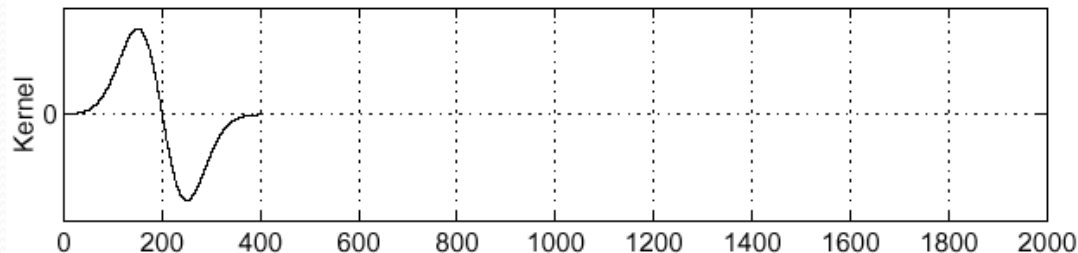
Combiner lissage et dérivation

$$\frac{\partial}{\partial x}(h \star f) = \left(\frac{\partial}{\partial x}h\right) \star f \quad (\text{i.e., économise une opération})$$

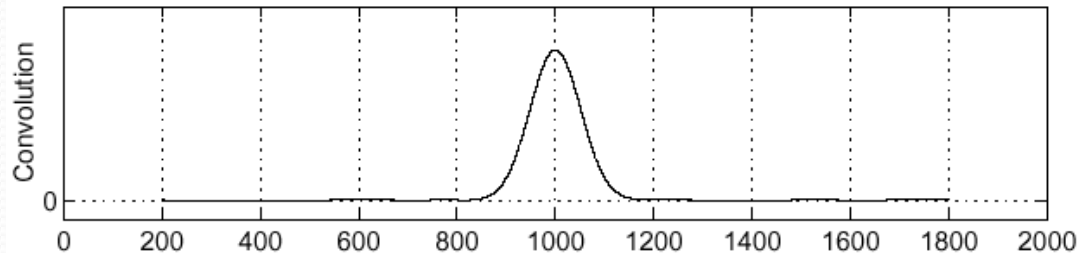
f



$\frac{\partial}{\partial x}h$



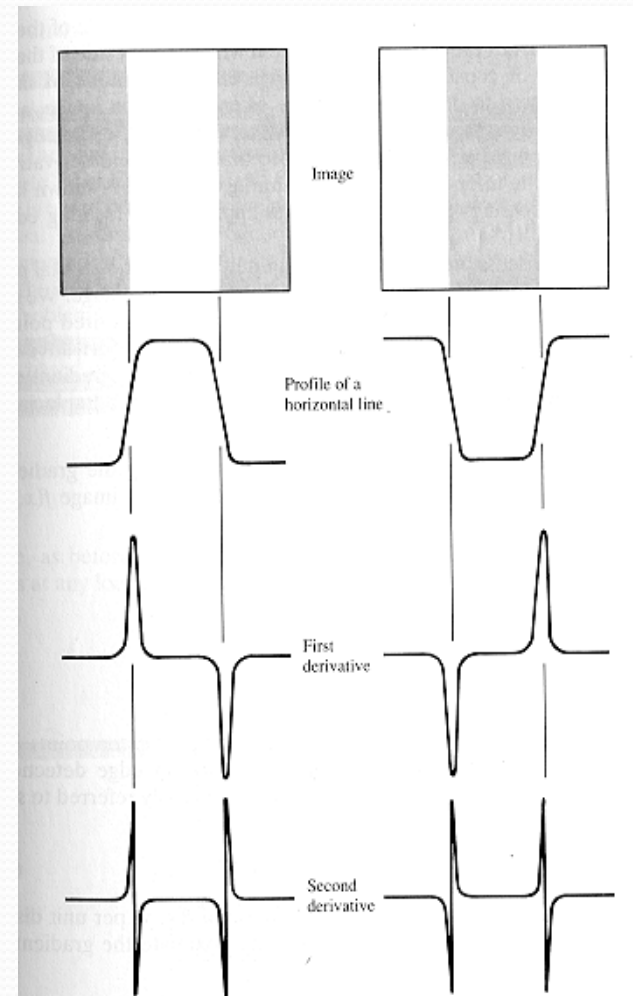
$\left(\frac{\partial}{\partial x}h\right) \star f$



Deuxième dérivée de l'image

- Une autre approche pour trouver les contours de l'image est d'utiliser la seconde dérivée de l'image.
- Pour cela, on utilise le Laplacien comme opérateur
- Les contours correspondent :
 - Aux maxima de la première dérivée
 - Aux passages par zéros de la deuxième dérivée

$$\nabla^2 I = \frac{\partial I}{\partial x^2} + \frac{\partial I}{\partial y^2}$$



Laplacien

- Plusieurs approximations discrètes du Laplacien existent.

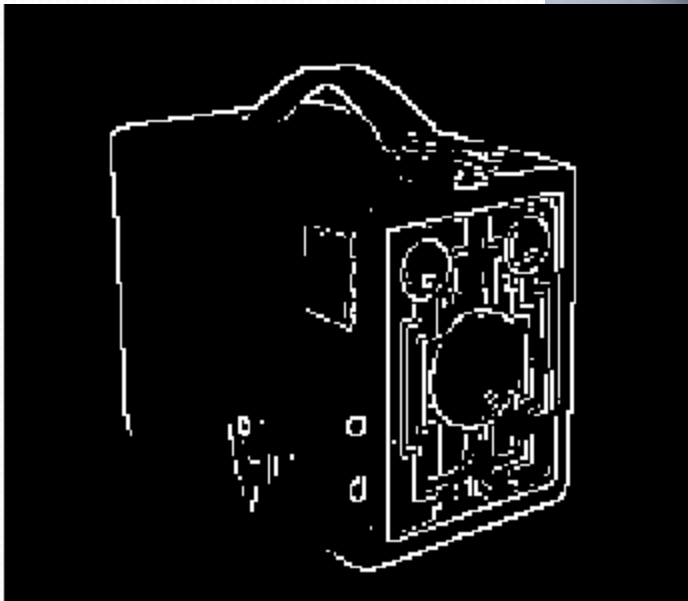
$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix} \quad \text{ou} \quad \begin{bmatrix} 1 & 1 & 1 \\ 1 & -8 & 1 \\ 1 & 1 & 1 \end{bmatrix} \quad \text{ou} \quad \begin{bmatrix} 1 & 4 & 1 \\ 4 & -20 & 4 \\ 1 & 4 & 1 \end{bmatrix} \quad \text{ou ...}$$

- Une seule matrice de convolution
- Symétrique en rotation

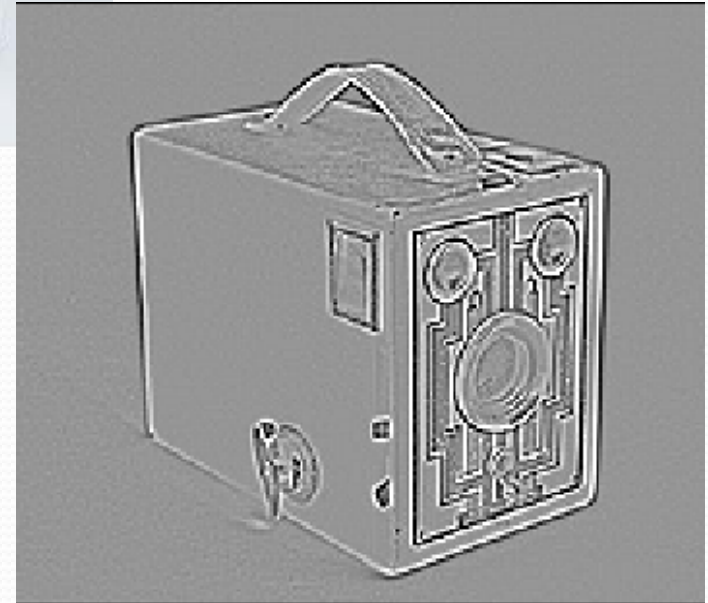
Comparaison "Gradient" / Laplacien



Gradient



Laplacien



Laplacien de Gaussienne

- Le laplacien, comme tous les détecteurs de contours, est très sensible aux bruits.
- On préfère lisser l'image avant de détecter les contours.
- Pour ce faire, on combine un lissage Gaussien avec le Laplacien.
 - La Gaussienne lisse l'image et rend les contours flous, mais conserve leurs positions
 - Le Laplacien donne les passages par zéro aux contours

Laplacien de Gaussienne

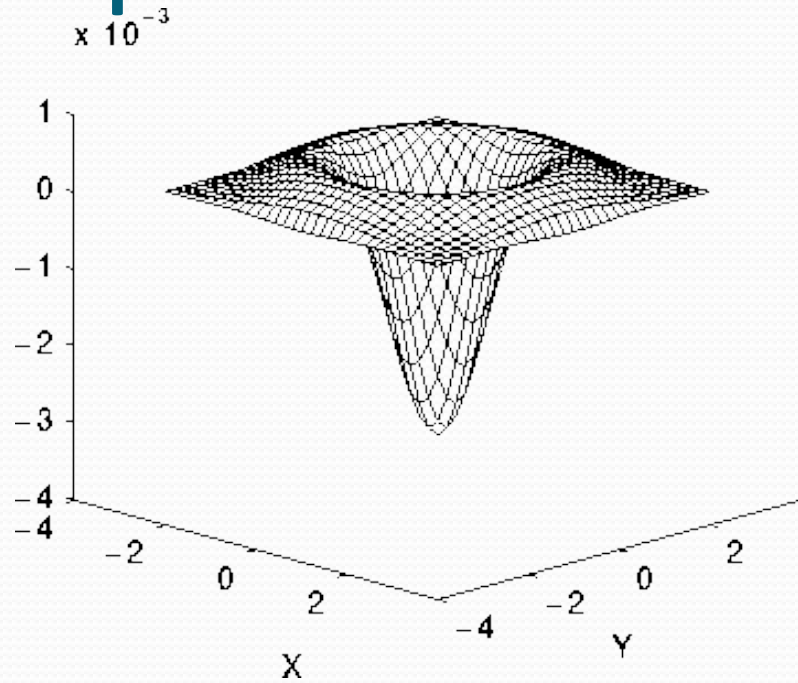
- On réalise cette opération d'un coup en utilisant un filtre Laplacien de Gaussienne (*LoG*) :

$$LoG * I = \nabla^2 * G * I = \nabla^2 G * I$$

$$LoG(x, y) = -\frac{1}{\pi\sigma^4} \left[1 - \frac{x^2 + y^2}{2\sigma^2} \right] e^{-\frac{x^2 + y^2}{2\sigma^2}}$$

- Dans ce filtre, on pré-convolue le Laplacien et la Gaussienne en choisissant le paramètre σ de la Gaussienne.
- Le système visuel humain utiliserait une approche similaire pour reconnaître les contours.

Laplacien de Gaussienne



La fonction LoG à la forme d'un chapeau mexicain inversé.

0	1	1	2	2	2	1	1	0
1	2	4	5	5	5	4	2	1
1	4	5	3	0	3	5	4	1
2	5	3	-12	-24	-12	3	5	2
2	5	0	-24	-40	-24	0	5	2
2	5	3	-12	-24	-12	3	5	2
1	4	5	3	0	3	5	4	1
1	2	4	5	5	5	4	2	1
0	1	1	2	2	2	1	1	0

*Approximation discrète de LoG
avec $\sigma=1.4$*

*On choisit souvent comme
taille de filtre gaussien $5^* \sigma$.*

Source : <http://www.dai.ed.ac.uk/HIPR2/log.htm>

Laplacien de Gaussienne



Sigma = 3



Sigma = 5

En faisant varier σ , on obtient différents niveaux de contours.

Quel filtre choisir pour les contours ?

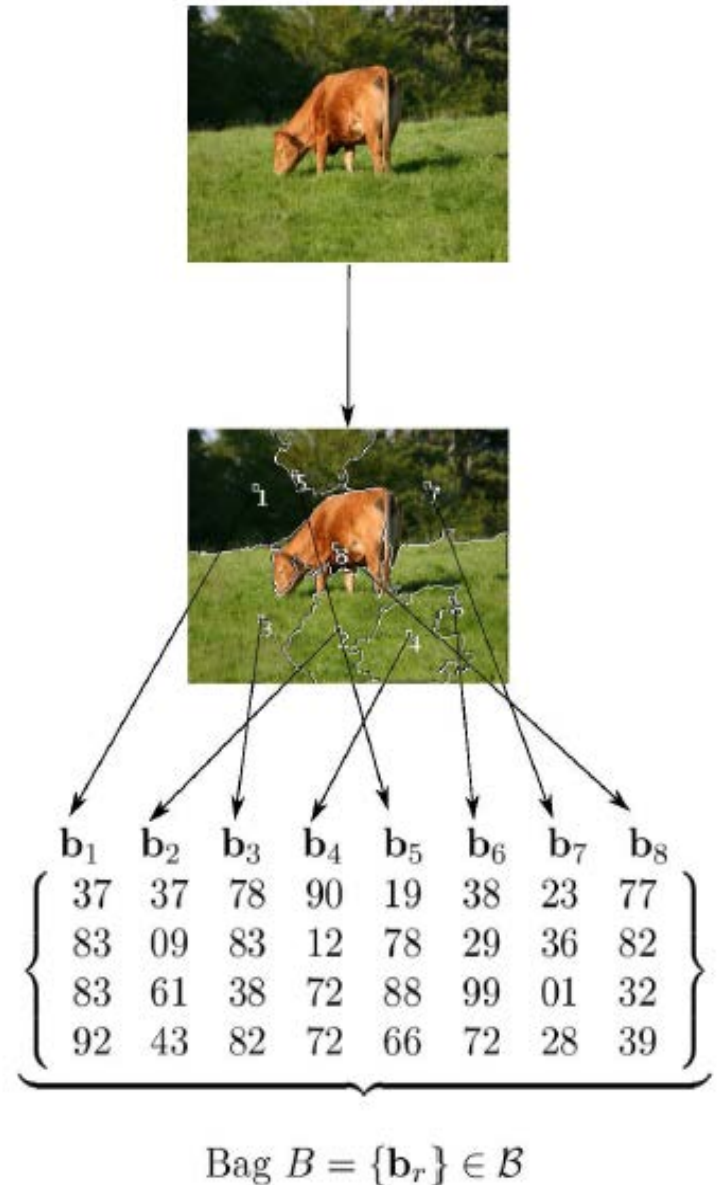
- Aucun opérateur n'est parfait pour détecter les contours.
- En pratique, on obtient des contours incomplets
 - il y a des pixels superflus
 - il y a des manques
 - il y a des erreurs dans la position et l'orientation des pixels contours
- Chacun semble avoir sa préférence pour une méthode ou une autre.
- Un opérateur de détection de contour n'est qu'une première étape dans la chaîne de segmentation.

Plan

- I. Généralités sur la recherche par le contenu
 - De la RI à la recherche par le contenu
- II. **Etat de l'art en Indexation d'images**
 - Primitives images
 - **Descripteurs images**
 - Formation d'index
- III. Modèle Bag of Word (BoW) : représentation TME
 - a. Détecteur dense
 - b. Dictionnaire visuel : K- Means
 - c. BoW : génération histogramme (+info spatiale)
 - Coding + pooling

Descripteur d'image

- Pour chaque primitive détectée dans l'image
 - Pixel, échantillonnage dense
 - Région de segmentation
 - PoI, RoI extrait selon des critères de saillance
- Caractérisation du contenu visuel de la primitive



Descripteur d'image

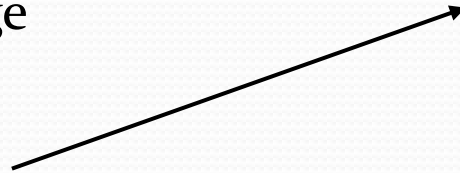
Défis posés

1. **Robustesse** \Rightarrow Invariance aux changements d'illumination, d'échelle, rotation, transformations affines, perspective
2. **Distinctif** \Rightarrow utile pour les recherches dans de grandes bases.
3. **Efficient** \Rightarrow facile et rapide à calculer (temps-réel)
4. **Représentatif** \Rightarrow beaucoup de points/régions détectés même sur des petites zones/objets

Difficultés

- Transformations image : rotation

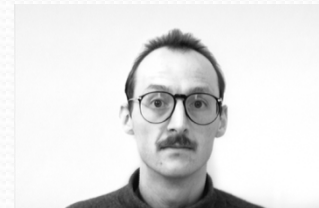
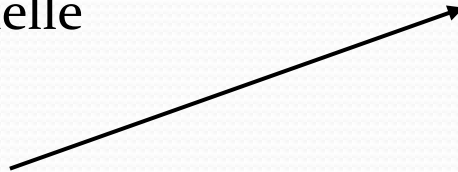
rotation d'image



Difficultés

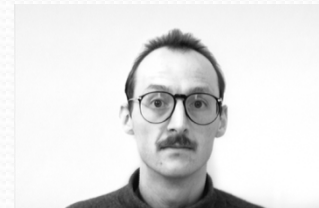
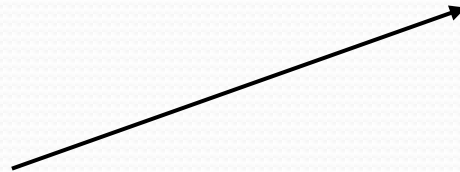
- Transformations image : échelle

changement d'échelle



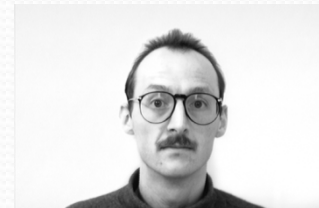
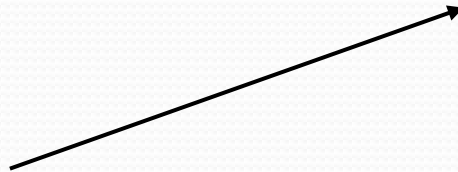
Difficultés

- Transformations image
 - Changement de luminosité



Difficultés

- Transformations image
 - Visibilité partielle / occultation



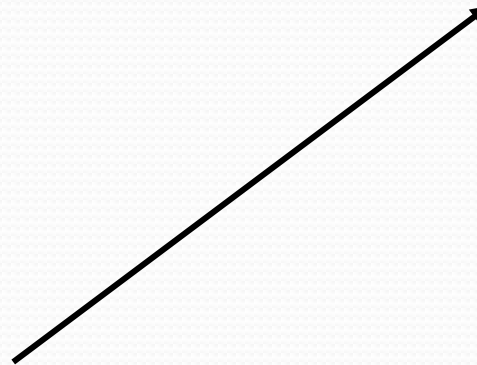
Difficultés

- Transformations image
 - Objets supplémentaires, fond



Difficultés

- Transformations image
 - Objets supplémentaires, fond



Descripteurs d'image

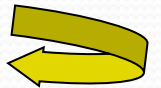
- Classification des caractéristiques extraites
 1. Couleur
 2. Texture
 3. Forme
 4. Gradient / contour

Descripteurs couleur : Signature colorimétrique

- Palette couleur



- Signature image : histogramme sur la palette :

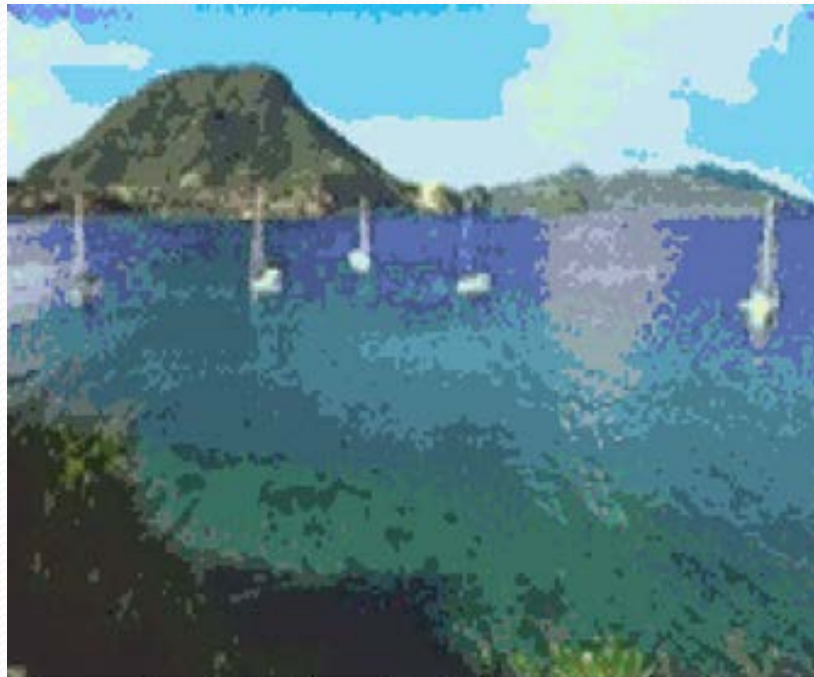


[8,27,35,48,1,10,32,9,1, 4, 26, 6,19,7, 3, 8, 3, 7, 2, 0, 0 , 0, 3, 0, 2]



Descripteurs couleur : Signature colorimétrique

- Différents espace colorimétrique : RGB, HSV, L^*a^*b
- Palette globale ou optimisée sur une base
 - Partie Bag of Words



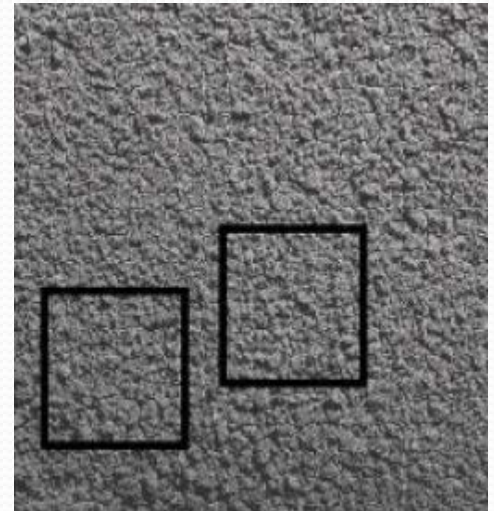
Descripteurs d'image

- Classification des caractéristiques extraites
 1. Couleur
 2. **Texture**
 3. Forme
 4. Gradient / contour

Descripteurs texture

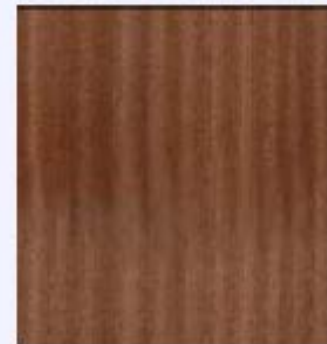
Qu'est-ce que la texture ?

- Une texture représente, à une échelle donnée, le même aspect quelle que soit la zone observée.
 - Répétition spatiale d'un même motif dans différentes directions de l'espace



Qu'est ce que la texture

- Plus formellement :
 - On considère l'image image comme la réalisation d'un processus stochastique local et stationnaire.
 - Caractérisation du voisinage de chaque pixel identique (même modèle statistique) pour tous les pixels de l'image.



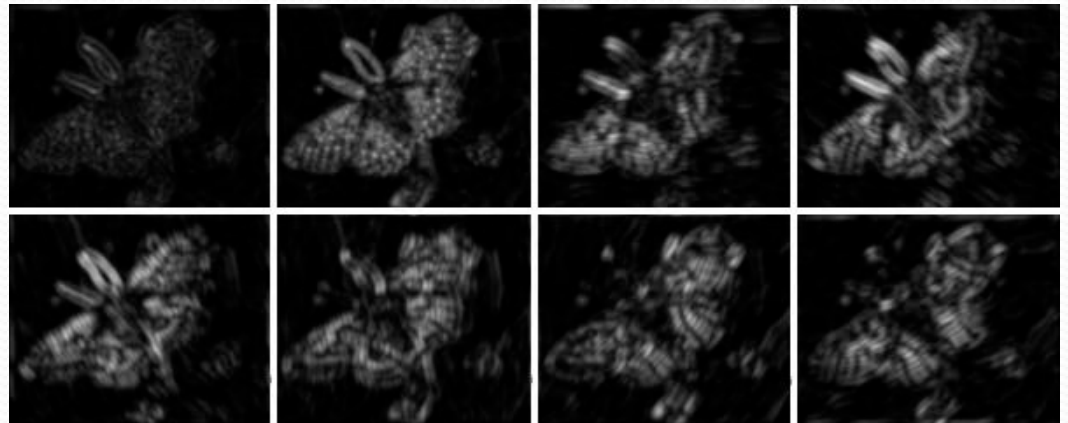
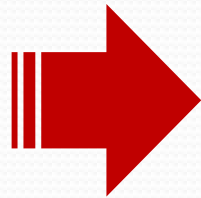
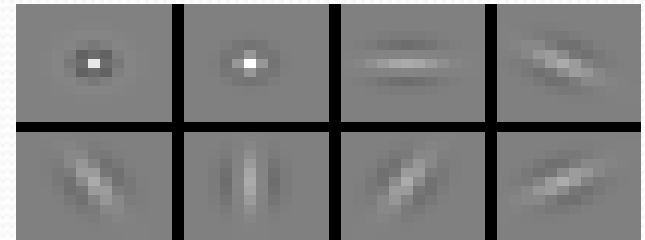
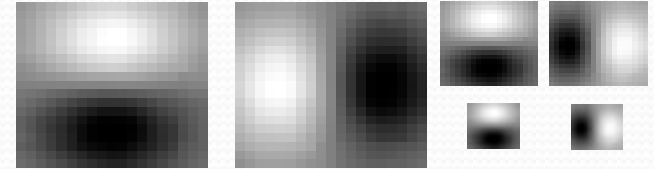
Descripteurs texture

- La plupart des descripteurs texture sont initiés par une étape de pré-traitement
 - convolution de l'image par des filtres fréquentiels
 - Avec certain paramètres fréquentiels caractéristiques de motifs texturels
 - Fréquence spatiale
 - Orientation de la texture
 - *etc*



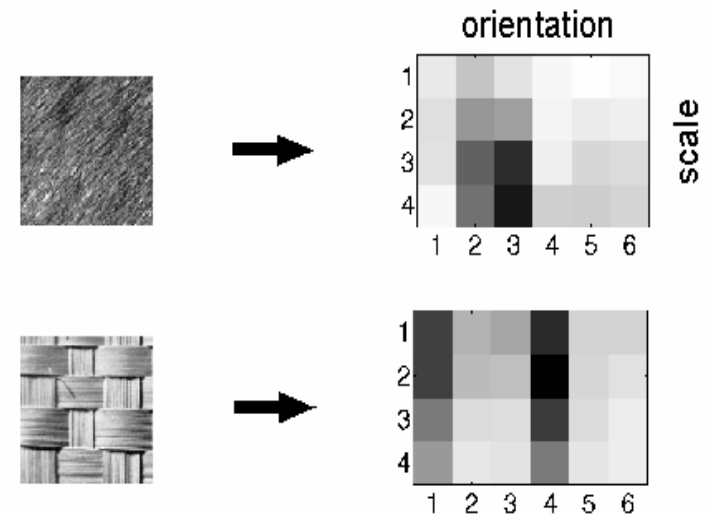
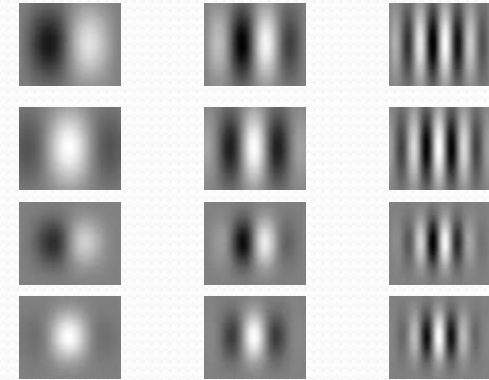
Descripteurs texture

- Difference of Gaussian Filters
- Spots and Oriented Bars (Malik and Perona)



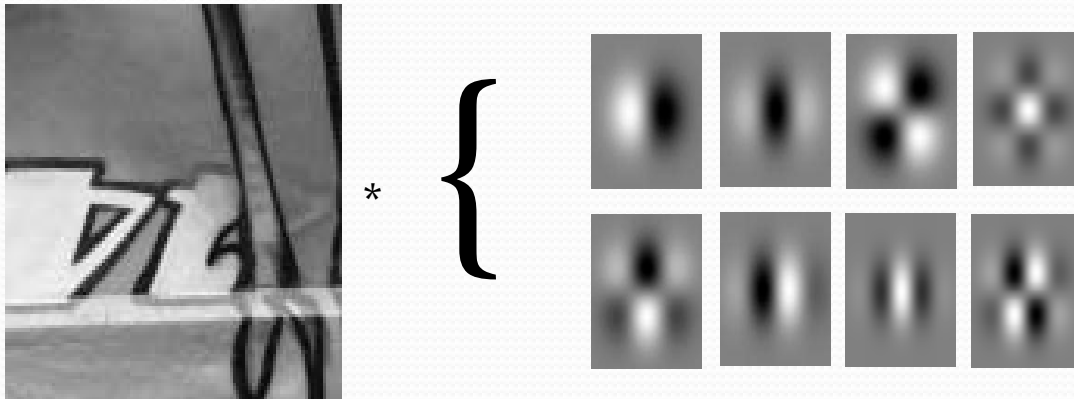
Descripteurs texture (suite)

- Bancs de filtre de Gabor à différentes échelles et fréquences spatiales
 - Filtres antisymétriques ou symétriques
- Attributs basés sur des bancs de filtres de Gabor
 - Vecteur de dimension 6x4

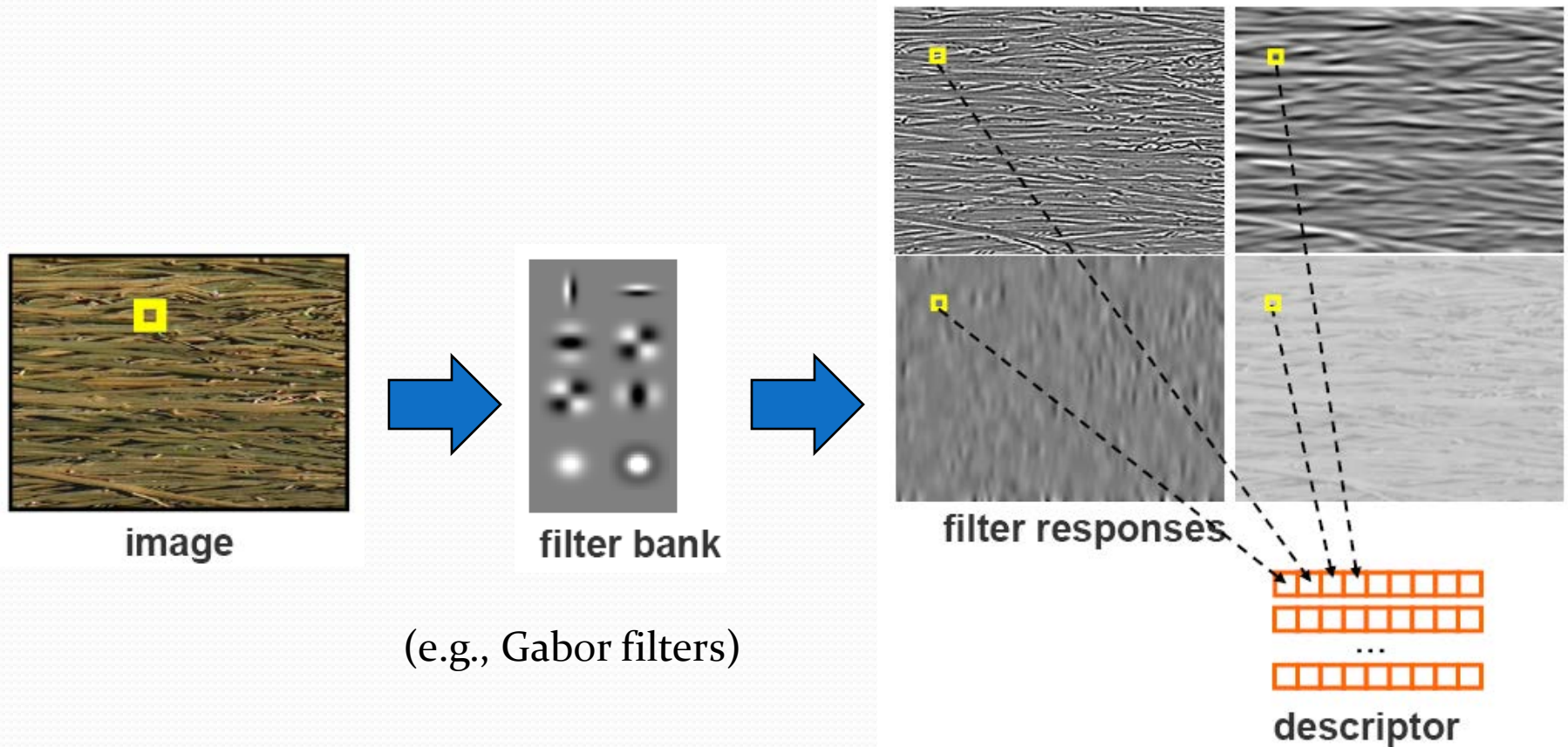


Descripteurs texture (suite)

- Steerable filters (*Freeman and Adelson'91*) (dim. 13)



Descripteurs de texture



Descripteurs d'image

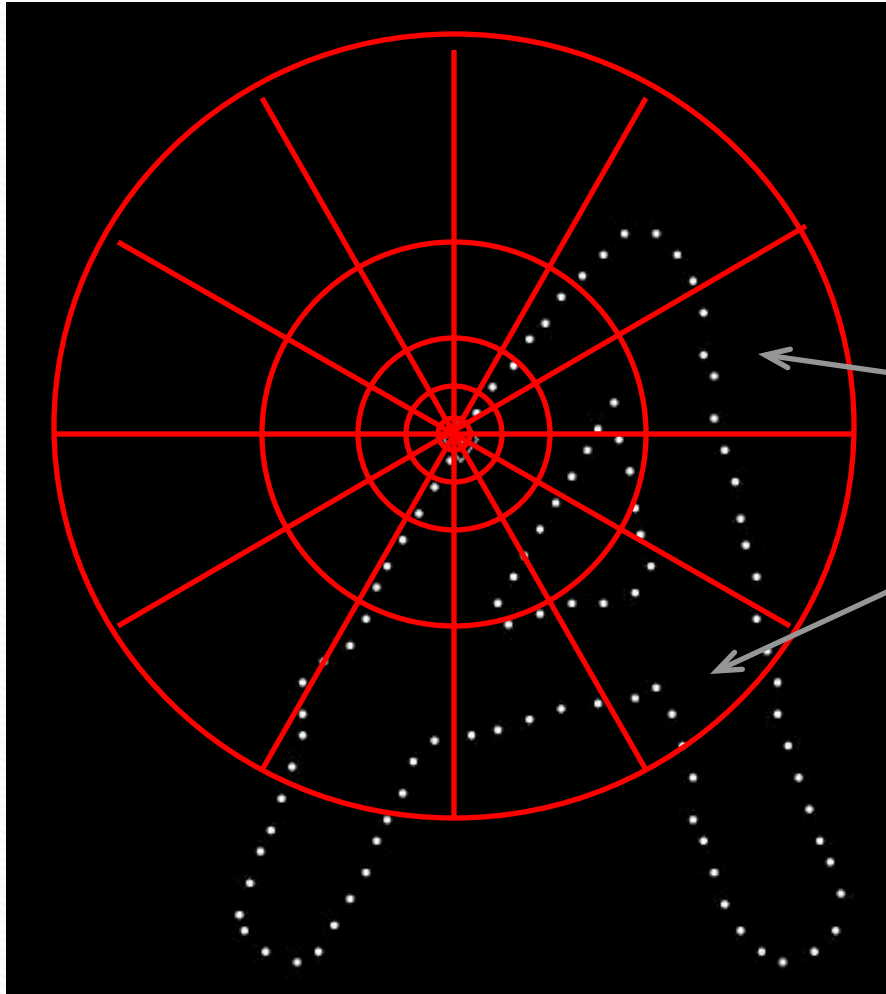
- Classification des caractéristiques extraites
 1. Couleur
 2. Texture
 3. **Forme**
 4. Gradient / contour

Descripteurs Forme

- Analyse de la forme
 - Région binaire : *e.g.* Moments de Zernike
 - Contour : *e.g.* Descripteurs de Fourier, Shape Context
- Performances dépendantes du succès de l'étape de segmentation / binarisation



Autre descripteur : Shape Context(forme)



Comptage du nb de points de chaque case

nb= 4

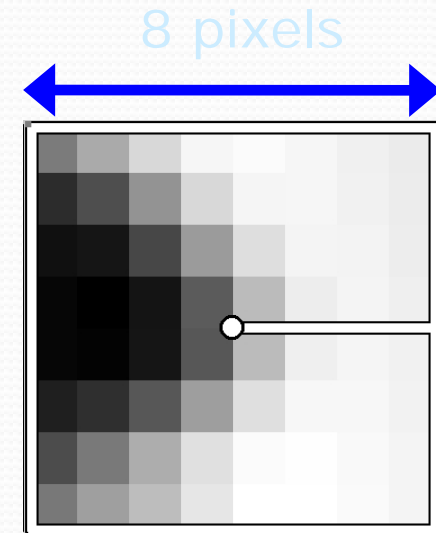
⋮

nb= 10

Histo Log-polaire : + de précision pour les points proches, plus de tolérance vis-à-vis des points éloignés.

MOPS :descripteur type patch normalisé

- Patch 8x8 avec une normalisation d'orientation
 - Echantillonné à 5 échelles
- Renormalisation /r biais/gain: $I' = (I - \mu)/\sigma$



Descripteurs d'image

- Classification des caractéristiques extraites
 1. Couleur
 2. Texture
 3. Forme
 4. Gradient / contour

Plan

I. Généralités sur la recherche par le contenu

- De la RI à la recherche par le contenu

II. Etat de l'art en Indexation d'images

- Primitives images
- Descripteurs images
- Formation d'index

III. Modèle Bag of Word (BoW)

Bag of Word : introduction

- Retour sur notre problématique de recherche de cible
 - Nécessite de trouver le même objets avec des déformations
 - Combinaison gagnante : détecteur sparse PoI + descripteurs discriminants et invariants type SIFT
 - Local => donc intrinsèquement robustes aux occultations et au bruit (du fond)
 - Challenge : recherche rapide de plus proches voisins dans les (très) grandes bases
 - Vérification géométrique pour contraindre les correspondances

Le modèle Bag of Words (BoW) issu de la recherche textuelle

Of all the sensory impressions proceeding to the brain, the visual experiences are the dominant ones. Our perception of the world around us is based on the messages that reach the brain. For a long time it was believed that the image was transmitted as a whole into the brain. However, Hubel and Wiesel have been able to show that the message about the image falling on the retina undergoes a step-wise analysis in a system of nerve cells stored in columns. In this system each cell has its specific function and is responsible for a specific detail in the perception of the retinal image.

**sensory, brain,
visual, perception,
retinal, cerebral cortex,
eye, cell, optical
nerve, image
Hubel, Wiesel**

China is forecasting a trade surplus of \$90bn (£51bn) to \$100bn this year, a threefold increase on 2004's \$32bn. The Commerce Ministry said the surplus would be created by a predicted 30% jump in exports, offset by a 18% rise in imports. The ministry argued that China's trade surplus was a deliberate policy to support the yuan, bank, domestic, foreign, increase, trade, value.

**China, trade,
surplus, commerce,
exports, imports, US,
yuan, bank, domestic,
foreign, increase,
trade, value**

On compare deux documents en comparant leurs histogrammes d'occurrence de mots

Documents (Google)

Créer des moteurs de recherche comprenant le langage naturel

-> évolution de *Google* au-delà des mots clés

-> le « text-mining » ou le « web-mining »

Collection : ensemble des documents

Vector Space Model (VSM) : espace à N dimensions où N est le nombre de termes utiles dans le langage (« le », « la »... sont inutiles et sont appelés **stop-words**). L'espace est donc découpé selon N clusters...

Documents (suite)

- Etape 1 : Pour chacun des N termes du langage, création d'un **Inverse Index** qui stocke pour chaque terme du langage les documents l'utilisant
 - Calcul du **Document Frequency** = nb doc utilisant ce terme / nb total de document = **df** (Plus df grand, moins le terme a d'importance d'un point de vue informatif)
 - soit **idf** = $\log(1/df)$
- Etape 2 : Pour chaque document,
 - Pour chaque terme, calcul du **Term Frequency** (Plus **tf** est grand dans un document plus ce terme doit être important par rapport au sujet du document)
 - Calcul d'un vecteur caractéristique VSM :
 - Pour i allant de 0 à N , $VSM[i] = \mathbf{tf}(\text{terme } i \text{ dans ce document}) * \mathbf{idf}(\text{terme } i)$
 - On normalise ce vecteur pour que $\|VSM(\text{document})\|=1$

Chaque requête de l'utilisateur est associée de la même façon à un $VSM(\text{requête})$ et la Similarité est calculée par produit scalaire (voir clustering pour indexation et optimisation, voir ontologies pour web sémantique)

Documents (suite)

- Si on veut être plus statisticien = Mesures de proximité entre 2 vecteurs-distributions le cadre fréquentiste
- Pour les documents (un feature = un mot)...

Table 2: Tableau de contingence

		Mots					
		m_1	...	m_j	...	m_L	somme
	D_1	$f_{1,1}$...	f_{1j}	...	f_{1L}	...

Textes	D_i	f_{i1}	...	f_{ij}	...	f_{iL}	$f_{i.} = \sum_{j=1}^L f_{ij}$

	D_D	f_{D1}	...	f_{Dj}	...	f_{DL}	...
	somme	$f_{.j} = \sum_{i=1}^D f_{ij}$	f

Documents (suite)

- Dans le cadre du texte mining, la **distance du χ^2** (*chiz* ou *khiz*) est la distance euclidienne entre deux vecteurs-documents normalisés par leur longueur (nombre de mots),

$$d_{eNid}^2(VSM_1, VSM_2) = \sum_{j=1}^L \left(\frac{f_{1j}}{f_{1.}} - \frac{f_{2j}}{f_{2.}} \right)^2$$

pondérée par la masse de chacun des mots par rapport à l'ensemble des textes (nombre total d'un mot dans l'ensemble des textes)

$$\chi^2(VSM_1, VSM_2) = \sum_{j=1}^L \frac{f}{f_{.j}} \left(\frac{f_{1j}}{f_{1.}} - \frac{f_{2j}}{f_{2.}} \right)^2$$

- f désigne le nombre total de mots dans l'ensemble des documents
 - f_{ij} désigne la fréquence du mot j dans le document i
 - $f_{i.}$ désigne le nombre total de mots du document i
 - $f_{.j}$ désigne la fréquence du mot j dans l'ensemble des documents
- Très utile dès qu'on veut comparer des histogrammes, des distributions de probabilités....

Documents (suite)

- Mesure de proximité entre 2 vecteurs-distributions = le cadre probabiliste
- La divergence de **Kullback-Leibler** entre les documents D_1 et D_2 :

$$d_{KL}^2(VSM_1, VSM_2) = \frac{1}{f} \sum_{j=1}^L f_{.j} \log \left(\frac{p(j|1)}{p(j|2)} \right)$$

$$d_{KL}^2(VSM_1, VSM_2) = \frac{1}{f} \sum_{j=1}^L f_{.j} \left(\log \frac{f_{1j}}{f_{1.}} - \log \frac{f_{2j}}{f_{2.}} \right)$$

- f désigne le nombre total de mots dans l'ensemble des documents
 - f_{ij} désigne la fréquence du mot j dans le document i
 - $f_{i.}$ désigne le nombre total de mots du document i
 - $f_{.j}$ désigne la fréquence du mot j dans l'ensemble des documents
- Pas une vraie distance, pas symétrique \Rightarrow appelée divergence

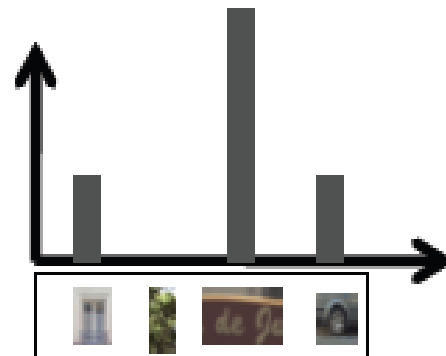
Du Bag of Features (BoF) => (BoW)

- BoF : ensemble de descripteurs locaux
 - Comment comparer deux images ? => signature vectorielle
 - Compromis discriminabilité / généralisation

Sac de descripteurs
(features)



BoW : histogramme
sur un dictionnaire

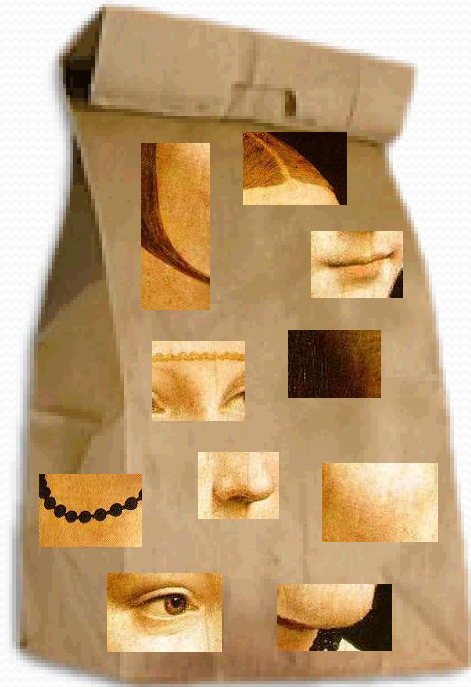


Le modèle Bag of Visual Words (BoW)

Image

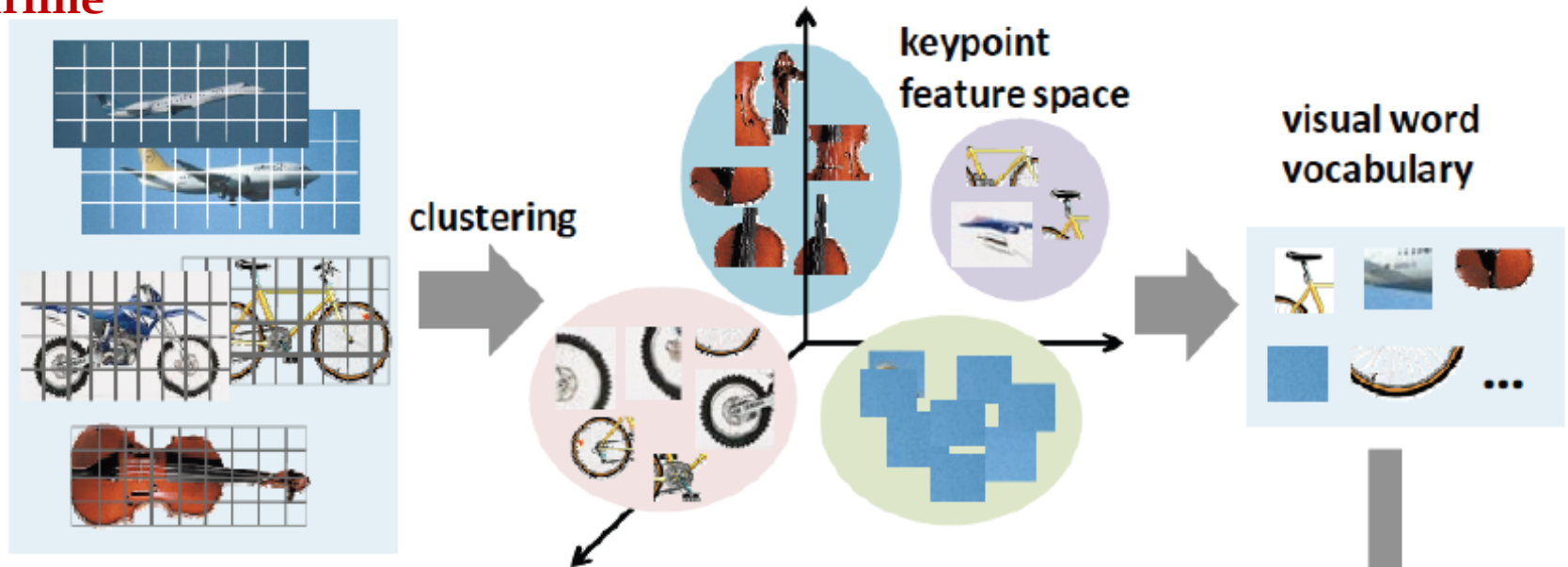


Bag of Visual
'Words'



Bag of Visual Words (BoW)

offline



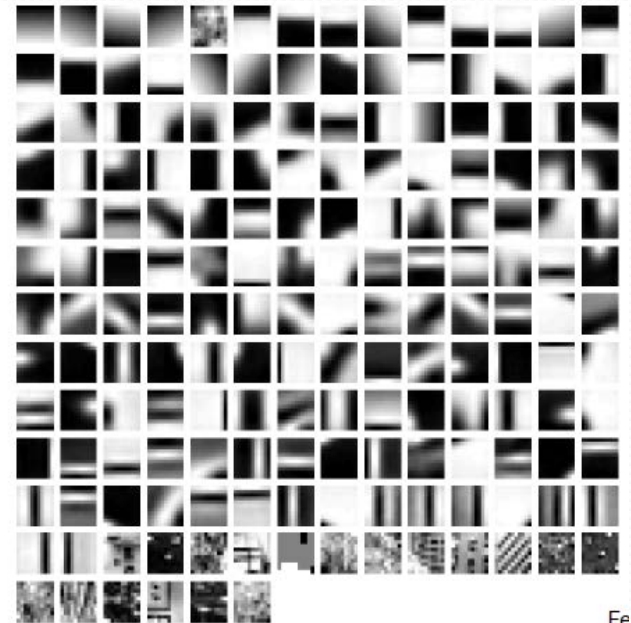
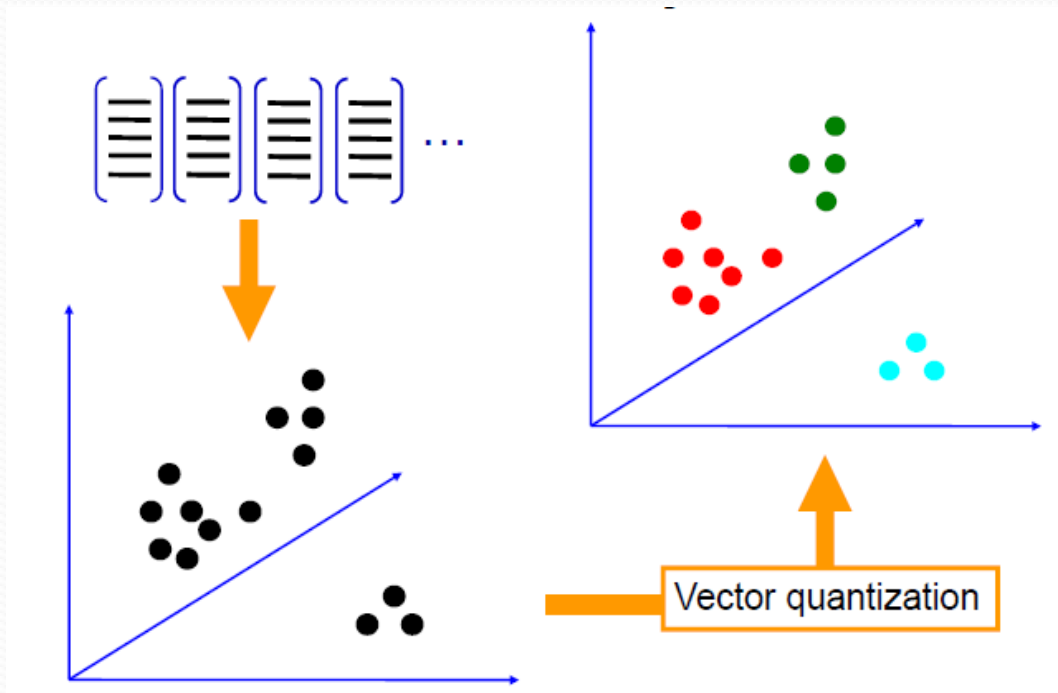
online



Bag of Visual Words : partie offline

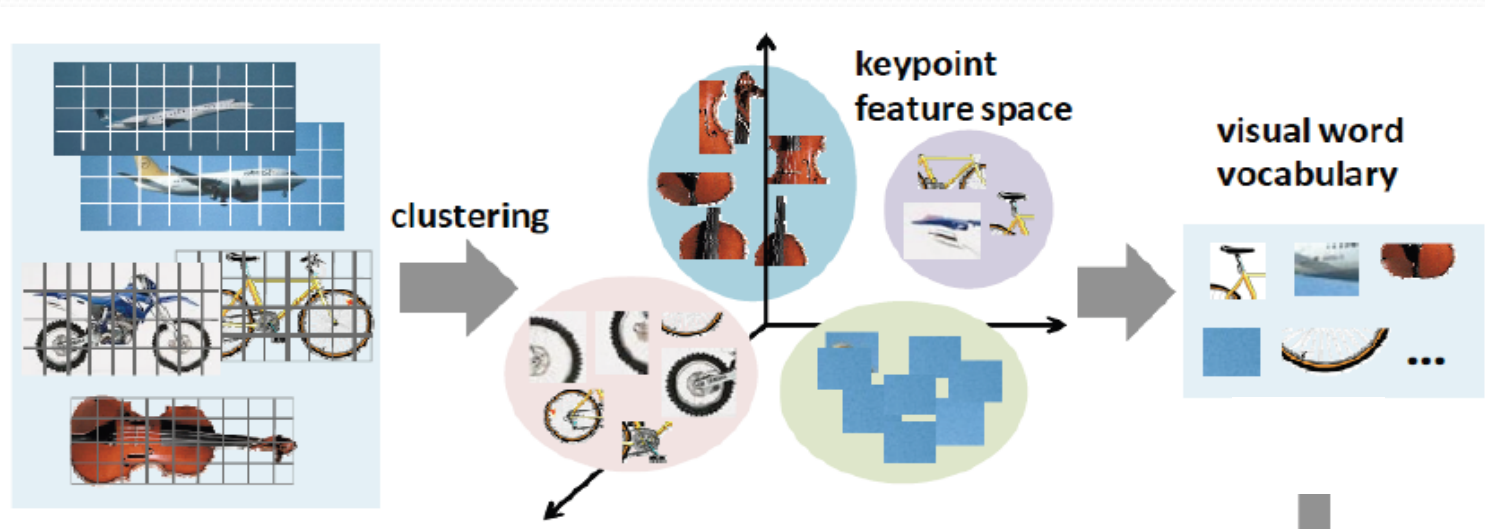
➤ Extraction descripteurs, approches basées dictionnaire « visuel »

- Construction d'un dictionnaire de motifs (mots) visuel adapté à la base traitée, *e.g.* K-Means



Bag of Visual Words : partie offline


- Extraction descripteurs, approches basées dictionnaire « visuel »
 - K-Means
 - Limitations : concentre les clusters sur les zones de fortes densités de points, délaisse les autres
 - Alternatives : méthodes à bases radiales, mais inconvénients inverses
 - En pratique le K-Means reste couramment utilisé
 - Apprentissage de dictionnaire (non supervisé ou supervisé)





THE END

Questions ??



Merci aux différents collègues du LIP6, de l'UPMC (Thome en premier, Cord, Iovan, Valle), et autres (Leibe, Grauman, Lowe, Bernard, Baillie, Verbeek, Schmid, Wu, Guigue, Gosselin..., à qui j'ai emprunté de nombreux documents et illustrations pour faire ce cours