

Indexation textuelle d'images par le contenu visuel pour un moteur de recherche d'images sur le Web

Sabrina Tollari, Hervé Glotin, Jacques Le Maitre

Université du Sud Toulon-Var

Toulon, le 25 novembre 2004

Plan

- Problématique
- Etat de l'art
- Système d'auto-annotation : DIMATEX
 - Protocole
 - Expérimentations
- Système de filtrage visuel d'indexation textuelle
 - Protocole
 - Premières expérimentations
- Perspectives

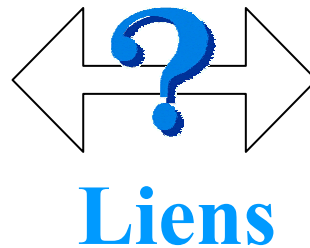
Problématique



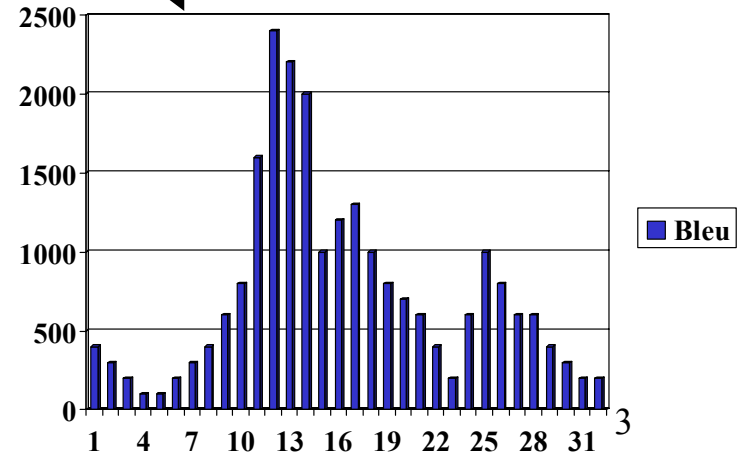
Indexation textuelle



mer ciel sable
fleurs arbres



Indexation visuelle



Rechercher une image



Images

Résultats 1 - 20 sur un total d'environ 510 pour **house water man**. (0,07 secondes)

Afficher: **Toutes les tailles** - [Grandes](#) - [Moyennes](#) - [Petites](#)

Essayez avec cette orthographe : house **waterman**



indxrvsn.jpg
448 x 260 pixels - 23 ko
[www.usbr.gov/mp/watershare/
resources/catalog.cfm](http://www.usbr.gov/mp/watershare/resources/catalog.cfm)



flag.jpg
329 x 265 pixels - 17 ko
[www.sendai-shi.com/
iwo_jima/images/flag.jpg](http://www.sendai-shi.com/iwo_jima/images/flag.jpg)



snow_l.jpg
600 x 400 pixels - 49 ko
[www.redfish.com/Silk-Road/
images/snow_l.jpg](http://www.redfish.com/Silk-Road/images/snow_l.jpg)



oasis_l.jpg
600 x 399 pixels - 46 ko
[www.redfish.com/Silk-Road/
images/oasis_l.jpg](http://www.redfish.com/Silk-Road/images/oasis_l.jpg)



11683.jpg
300 x 462 pixels - 49 ko
[www.kamat.com/.../
architecture/common/11683.htm](http://www.kamat.com/.../architecture/common/11683.htm)



aj0895cut2.jpg
200 x 264 pixels - 8 ko
[www.arne-jacobsen.com/
neobuilder.200202072003...](http://www.arne-jacobsen.com/neobuilder.200202072003...)



Bevan5.jpg
434 x 314 pixels - 19 ko
[www.lutsk.ukrpack.net/
~scm/Pic/Bevan5.jpg](http://www.lutsk.ukrpack.net/~scm/Pic/Bevan5.jpg)



15N-England-House.jpg
350 x 263 pixels - 43 ko
[boss.ethz.ch/.../pages/
15N-England-House.htm](http://boss.ethz.ch/.../pages/15N-England-House.htm)

Problématique

Indexation d'images pour la recherche d'images sur le Web

- Indexation textuelle
 - Manuelle : coûteuse, subjective
 - Automatique à partir du nom, de la légende ou du texte entourant l'image : ne décrit pas le contenu de l'image, beaucoup d'erreurs d'indexation car ce texte ne correspond que rarement avec l'image
- Indexation visuelle
 - Couleurs, formes, textures
 - Localisation, régions d'intérêts, segmentation
- Extraction de la sémantique difficile !

Indexer automatiquement une image

- House for sale in Saint-Zenon Quebec
- Mail delivery by mail man.
- Water has been tested is very good.
- Many mature trees on property.

www.zenwaiter.com/house.htm



Recherche d'images : house water man

Recherche d'images : house snow

Systemes de recherche d'images

Indexation textuelle seule	Indexation visuelle seule	Indexation textuelle et visuelle
Google MSN Search Moteurs de recherche d'images sur le web	Virage(1996) NeTra(1997) SurfImage(INRIA,1998) IKONA(INRIA,2001)	Chabot(Berkeley,1995) QBIC(IBM,1995) ImageSeeker5.4(INRIA,LTU)

- Peu de systèmes utilisent à la fois le texte et le contenu de l'image pour indexer, filtrer, combiner, rechercher les informations textuelles et visuelles efficacement. 7

Méthodes d'annotation automatique textuelle à partir du visuelle

- [REF1] **Kobus Barnard**, P. Duygulu, N. de Freitas, D. Forsyth, D. Blei, and M. I. Jordan, « *Matching Words and Pictures* », Journal of Machine Learning Research, Vol 3, pp 1107-1135, 2003.
- [REF2] **F. Monay and D. Gatica-Perez**, « *On Image Auto-Annotation with Latent Space Models* », in ACM MM, 2003
- [REF3] **Jia Li, James Z. Wang**, « *Automatic linguistic indexing of pictures by a statistical modeling approach* », *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2003.

- Tous utilisent la même base d'images COREL.

Systeme 1 : DIMATEX (DIchotomic IMAge TEXt system)

Un système d'annotation automatique
d'images à partir du contenu visuel

Protocole pour associer automatiquement des mots à une image

Corpus d'images

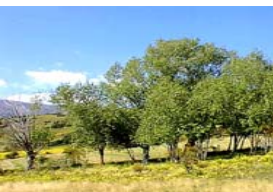
(indexées textuellement et visuellement)

Base de test

Base de référence



ciel mer fleur arbre



ciel arbre



arbre fleur route

Estimation des probabilités

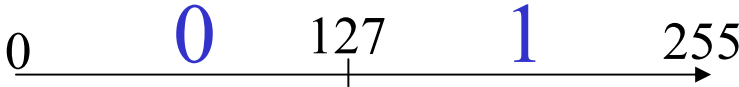
	Arbre	Ciel	Fleur	Mer
Rouge	2	1	2	1
Vert	3	2	2	1
Bleu	2	2	1	1

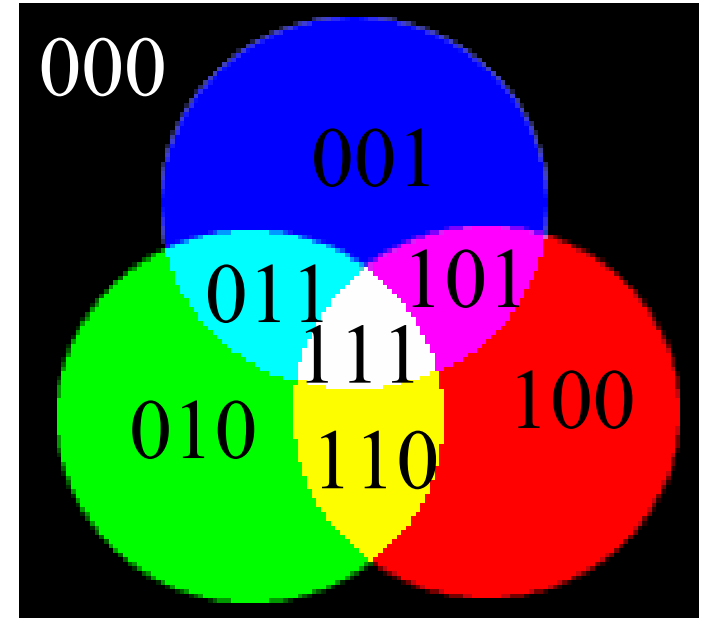
arbre (5fois), fleur (4 fois)

fusion

Classes visuelles

Construction de classes visuelles par méthode dichotomique

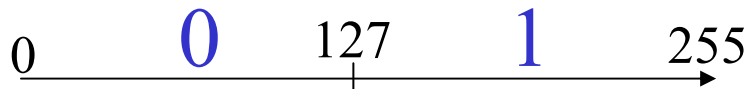
- Principe : séparer uniformément l'espace visuel
 - Exemple : découpage de l'espace RGB
 - chaque dimension est séparé en deux
- 0 0 127 1 255
- 
- il y a $2^3 = 8$ classes



Construction de classes visuelles par méthode dichotomique

- Principe : séparer uniformément l'espace visuel
- Exemple : découpage de l'espace RGB

- chaque dimension est séparé en deux



- il y a $2^3 = 8$ classes

$V = \{230, 190, 30\} \longrightarrow 110$

Classe	Mot 1	Mot 2	...	Mot n
000				
001				
010				
011				
100				
101				
110	+1		+1	
111				

Classement par dichotomie des mots associés à l'image

Calcul de la probabilité jointe d'un mot w et d'une classe visuelle C_k

- L'association entre une région (un « blob ») de l'image et une classe visuelle est unique.
- L'ensemble d'apprentissage T est composé des $|T|$ images J_j .
- Si w légende J_j alors il légende $b_i \in J_j$.

$$\begin{aligned} P(w, c_k | T) &= P(w, b_i | T) \\ &= \sum_j P(w, b_i | J_j, T) P(J_j | T) \\ &= \frac{1}{|T|} \cdot \sum_j P(w, b_i | J_j, T) \\ &= \frac{1}{|T|} \cdot \sum_j P(w | J_j, b_i, T) P(b_i | J_j, T). \end{aligned}$$

$$P(b_i | J_j, T) = \frac{\text{area}(b_i)}{\sum_{l=1}^m \text{area}(b_l)}$$

Auto-annotation d'images par K mots

- Soit I l'image de TEST composée des blobs b_1, b_2, \dots, b_m qui forment une partition.

$$P(w|I, T) = \sum_{i=1}^m P(w|b_i, I, T)P(b_i|I, T)$$

On légende l'image de TEST avec les K mots qui ont la plus grande probabilité.

Expérimentations

Corpus

Corpus

- Base d'images de COREL
- 10 000 images
- 200 mot-clés différents en anglais
- Chaque image possède :
 - De 1 à 5 mot-clés choisis manuellement
 - De 2 à 10 « blobs », des segments de l'image
 - Chaque blob de l'image possède un vecteur visuel de 40 composantes extrait par Kobus Barnard (aire, RGB, RGS, LAB, 12 coefficients de texture (filtres gaussiens),...)

Corpus

Corpus

- Exemples de segmentation par « normalized cuts »
- Adaptation du corpus :



- Normalisation par estimation MLE de distributions Gamma des vecteurs visuels pour la génération de distributions de probabilités
- 3 dimensions de forme (F), 6 pour LAB, 4 pour la texture (T)

Quelques images qui ont un blob dans la classe



pumpkin
vegetable



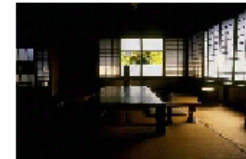
flower leaf
plants rose



flower leaf plants



pillar stone tree
wall



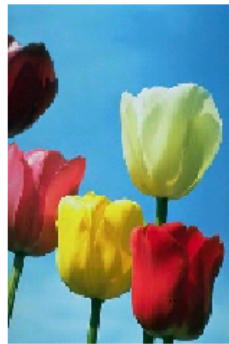
door floor tables
wall



flower leaf plants



closeup model
people woman



flower sky stem
tulip



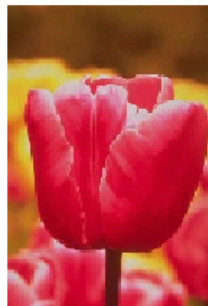
costume people



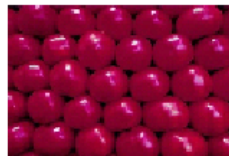
closeup people
woman



closeup fish
water



flower petals
stem tulip



closeup pattern
vegetable



closeup flower
garden



cactus saguaro
sunset tree

DIMATEX

Quelques images qui ont un blob dans la classe 0111010110000



field grass
horse tree



field foal horse
mare



field foal horse
mare



closeup flower
garden



forest road tree



grass house
tree



crop field plants



bird nest



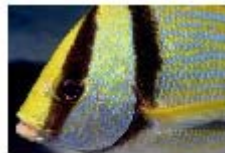
leaf moss



forest hills tree



field horse



closeup fish
water



column palace
tree



field foal horse
mare



flower leaf plants

Calcul du score de la classification « Normalized Score »

$$\text{Score NS} = \underbrace{\text{right}/n}_{\text{sensibilité}} - \underbrace{\text{wrong}/(N-n)}_{1\text{-spécificité}}$$

Où :

- right est le nombre de mots de références acceptés
- wrong est le nombre de mots acceptés qui ne sont pas des mots de références
- n est le nombre de mots de références de l'image
- N est le nombre de mots du vocabulaire

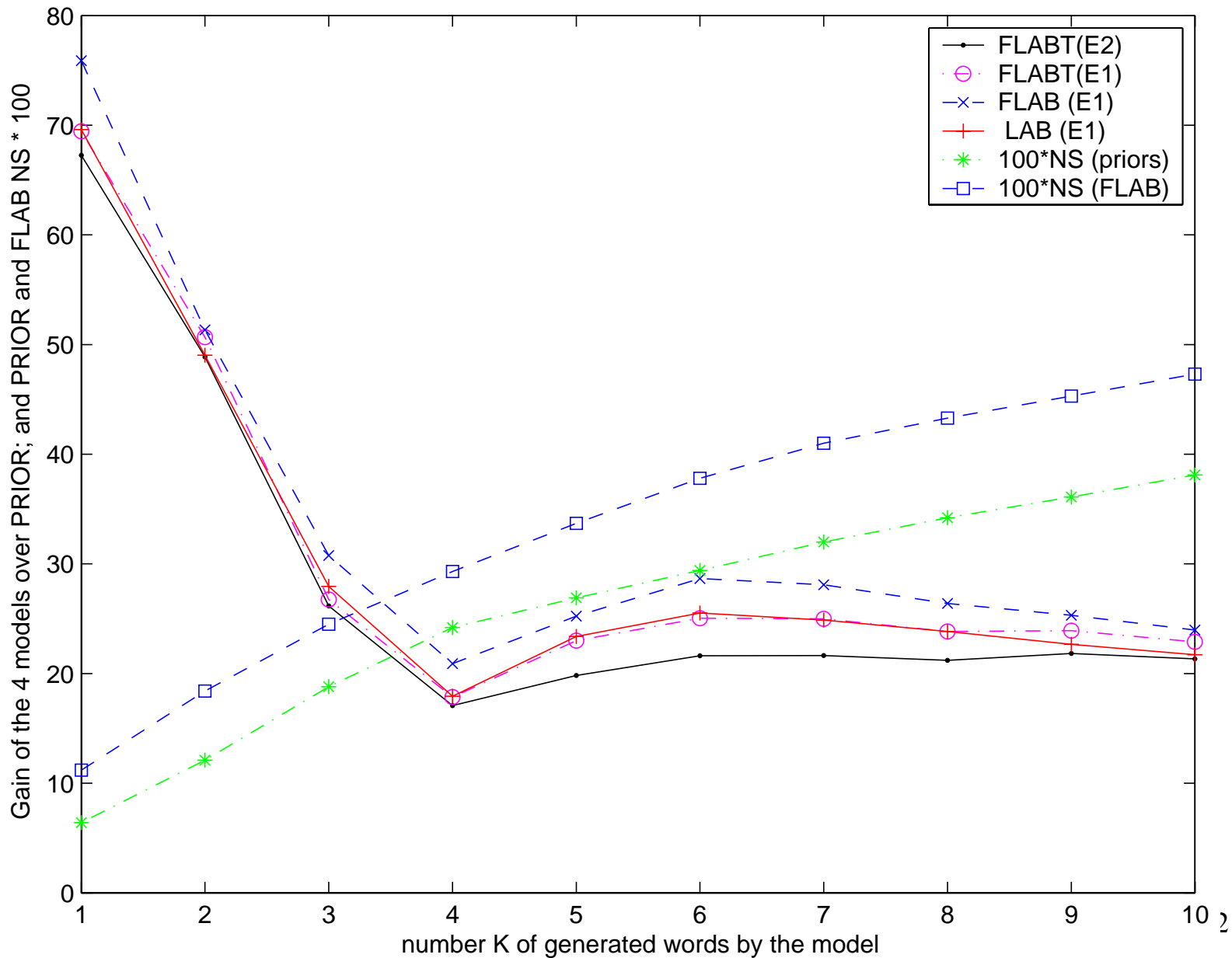
Calcul du score de la classification

Image de la base de test

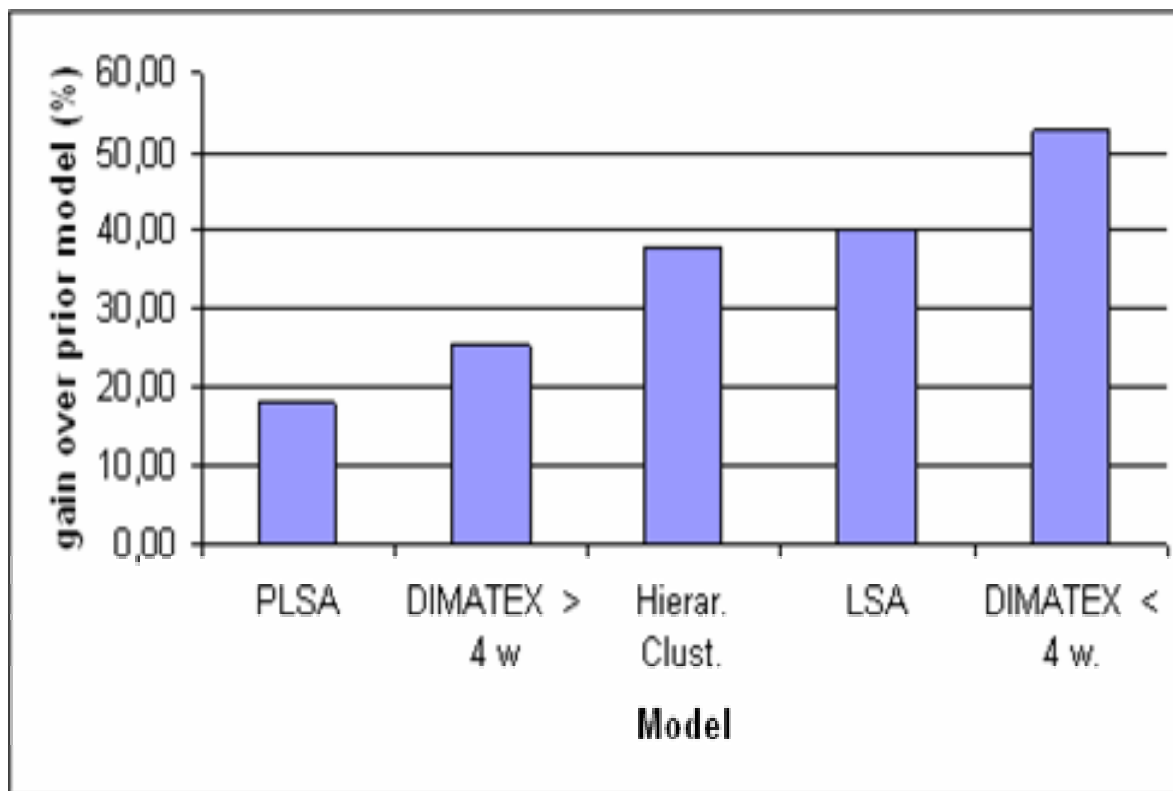


Mots de références de l'image	Mots trouvés par le système
fleur arbre lampadaire	arbre fleur ciel

Score NS=2/3-1/200



Etude comparative des modèles de l'état de l'art



Modèle Hierar. Clust. [REF1]

Modèle LSA, PLSA [REF2]

Exemples d'indexation par DIMATEX



horizon mountain
palace tree
sky tree rock building
grass



people tables tree
people flower tree
water plants



display
people water tree
flower sky



baby closeup people
people tree water
closeup flower



park sky tree
sky water tree people
bird



roof snow tree
people sky tree water
snow



field people tree
water tree grass
people mountain



field sky
tree sky forest
cougar fish



reflection temple tree
water
water tree sky people
bird



field sky tree
sky water tree people
grass



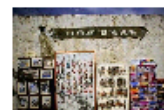
dune river water
tree sky water grass
people



street
people tree water
building sky



cloud hotel sky tree
tree people water sky
flower



shop street wall
people tree sky water
building



people road
people building sky
tree water

Avantages / Inconvénients du système

- Avantages
 - Annotation très rapide
 - Scores équivalents aux autres systèmes
- Inconvénients
 - Utilisable seulement avec des vecteurs de petites dimensions (pas plus de 20)
 - Soumis aux problèmes de cooccurrences
- Amélioration du système
 - Fiabiliser les légendes de l'ensemble d'apprentissage

Systeme 2

Systeme de filtrage visuel
d'indexation textuelle

Comment trouver les caractéristiques visuelles d'un mot ?

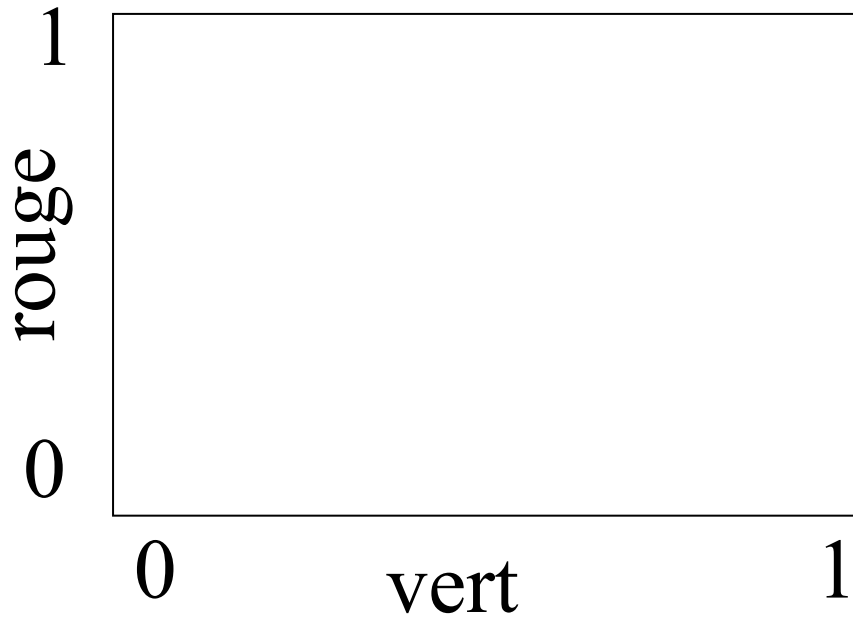
- Un mot peut avoir plusieurs représentations visuelles
 - Exemple : flower (différentes couleurs)
- Un mot peut ne pas avoir de traits caractéristiques visuels
 - Exemple : market, street, guard, restaurant, race
- Il est difficile de trouver des différences visuelles significatives entre de nombreux mots, particulièrement avec une faible résolution d'image
 - Exemple : flower, tulip

Construction de classes visuelles

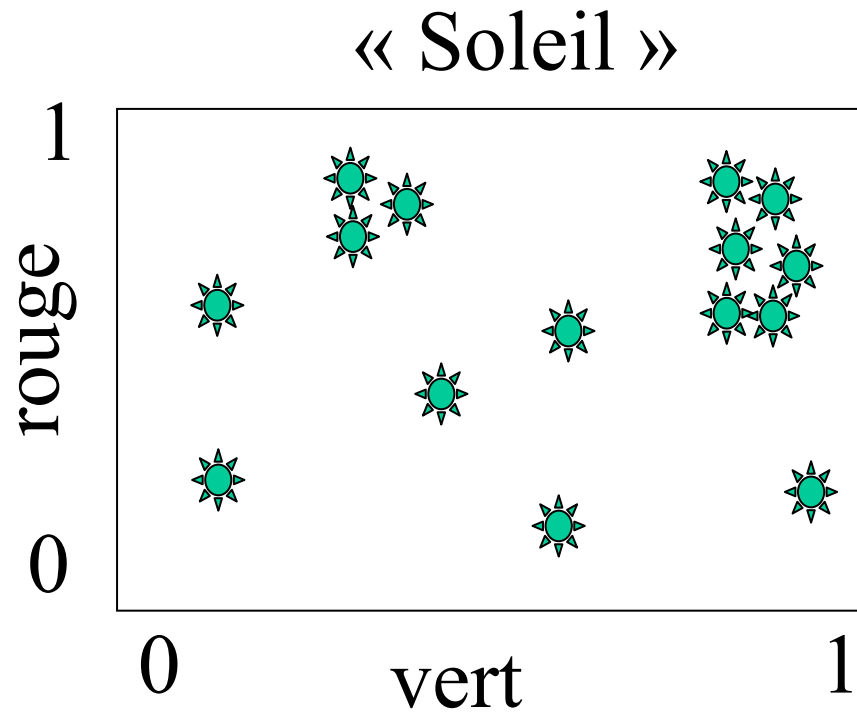
- Pour chaque mot,
 - On réalise une CAH (Classification Ascendante Hiérarchique) sur les traits visuels des blobs des images contenant ce mot à la recherche de noyaux de blobs
 - Critère d'arrêt de la CAH : le score NS

Système 2

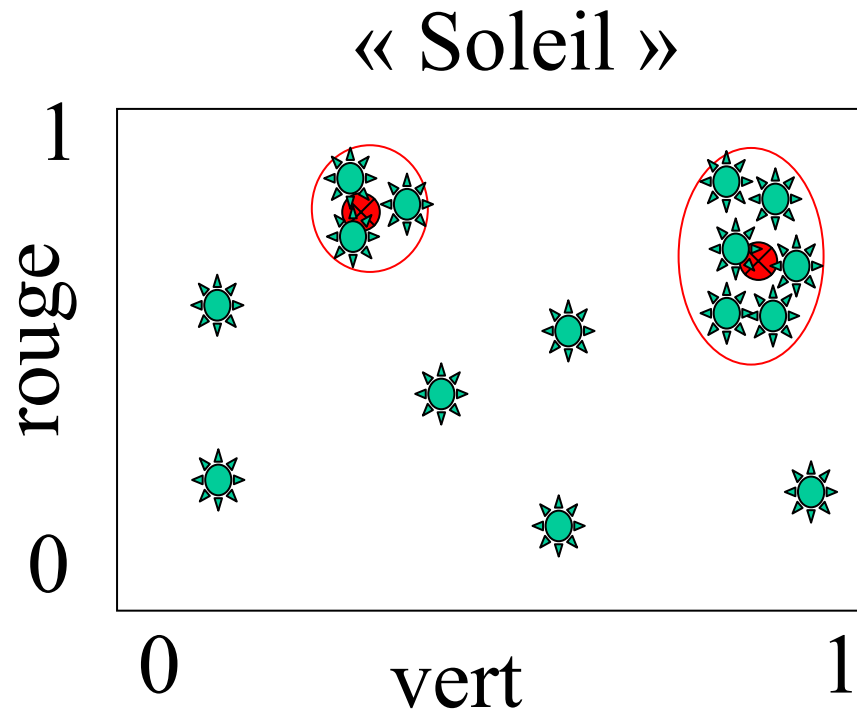
« Soleil »



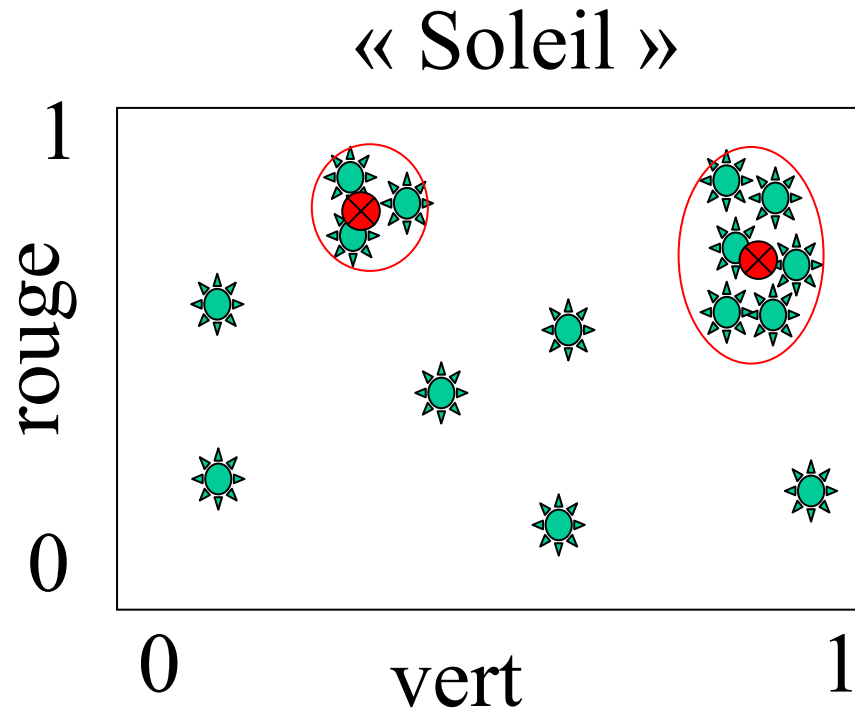
Système 2



Système 2



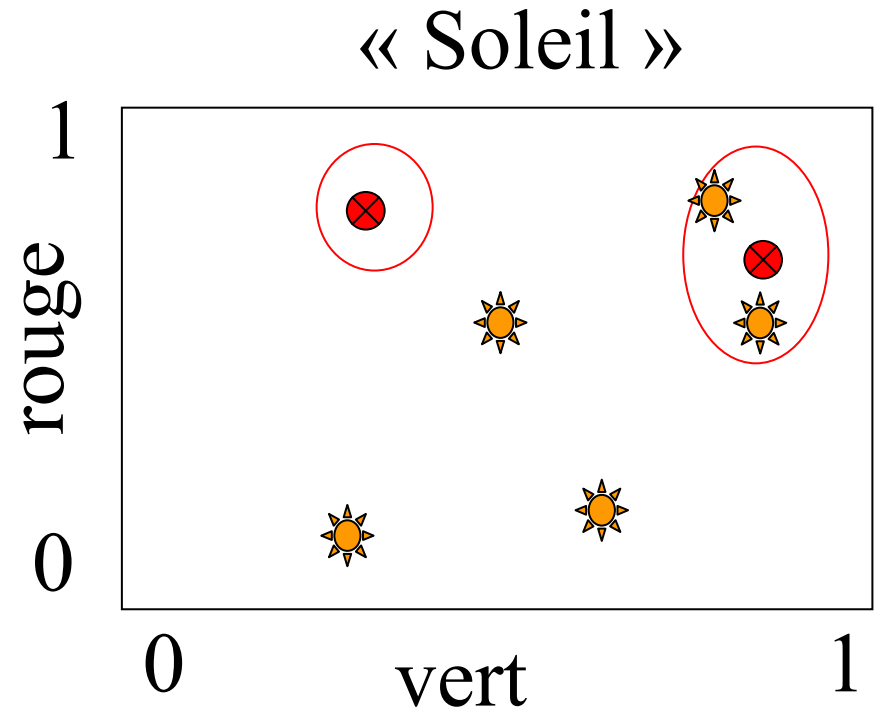
Système 2



- Pour chaque classe, on garde seulement :
 - le vecteur centroïde
 - les valeurs des écart-types de la classe pour chaque dimension du vecteur

Phase de test

- Pour chaque image de la base de test, on regarde combien de blobs sont dans l'un des clusters du mot.
- Si un ou plusieurs blobs sont dans l'une des classes d'un mot alors on indexe l'image avec le mot.
- Finalement, on calcule le score NS.

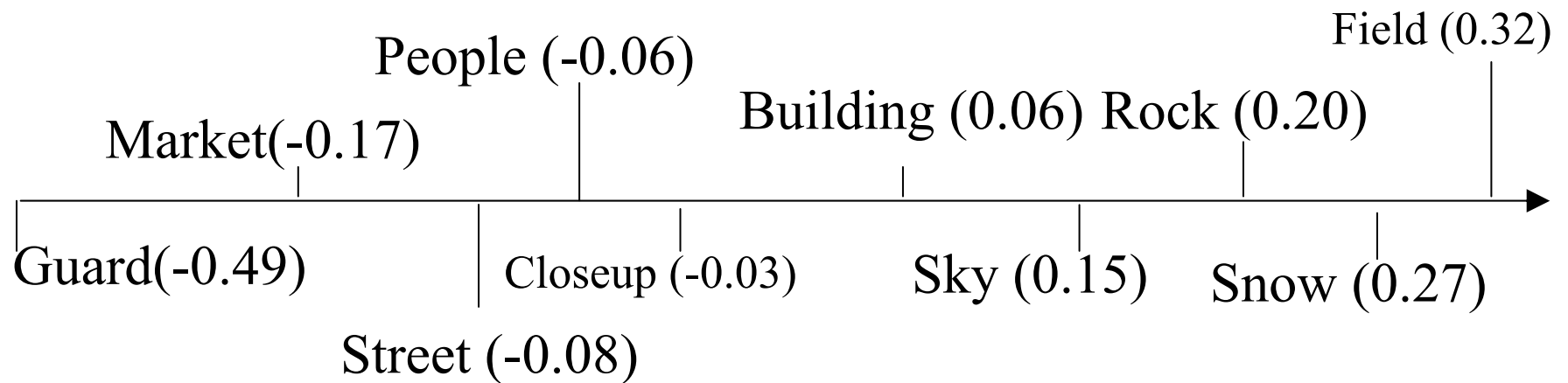


Premiers résultats pour quelques mots

	r	n	w	N	sensibilité	spécificité	NS
Field	103	122	1437	2922	0.84	0.48	0.32
Iceberg	19	20	867	2922	0.95	0.70	0.65
Leaf	76	119	585	2922	0.64	0.79	0.43
Lion	31	39	576	2922	0.79	0.80	0.60
Lizard	26	30	1453	2922	0.87	0.49	0.36
Sailboat	25	27	1377	2922	0.93	0.52	0.45
Rose	19	22	1324	2922	0.86	0.54	0.40

Systeme 2

Ordonnancement de la consistante visuelle
des mots à l'aide de la valeur du score NS



Conclusion sur le système 2

- Premières expérimentations encourageantes
- Ce système une fois bien paramétré permet de tester rapidement la validité visuelle des mots entourant une image.

- House for sale in Saint-Zenon Quebec
- Mail delivery by mail man.
- Water has been tested is very good.
- Many mature trees on property.



Perspectives

- Méthode indépendante des traits visuels utilisés, possibilité de comparer l'efficacité de nouveaux traits visuels par rapport aux autres traits.
- Réduire l'espace de recherche en fonction du mot et des traits visuels efficaces pour ce mot
- Appliquer ces méthodes sur le Web