

Filtrage de l'indexation textuelle d'une image au moyen du contenu visuel pour un moteur de recherche d'images sur le web

Sabrina Tollari
Université du Sud Toulon-Var

Directeurs de thèse :
Hervé Glotin et Jacques Le Maitre

Grenoble, le 10 mars 2005
CORIA'05

Problématique : Rechercher une image



[Web](#) [Images](#) [Groupes](#) [Annuaire](#) [Actualités](#)

house water man

Rechercher

[Images – Recherche avancée](#)
[Préférences](#)

Images

Résultats **1 - 20** sur un total d'environ **520** pour **house water man filetype:jpg -lukas.** (0,23 secondes)

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



Water Gap House.jpg

795 x 535 pixels - 45 ko
www.dutotmuseum.com/.../Water%20Gap%20House.jpg



Clear WindowPlants HouseB...

640 x 480 pixels - 69 ko
www.briannelsonconsulting.com/images-clearwin...



indxrvsn.jpg

448 x 260 pixels - 23 ko
www.usbr.gov/mp/watershare/resources/catalog.cfm



tsunami01.jpg

742 x 557 pixels - 63 ko
www.hanneketravels.net/alaska/tsunami01.jpg



snow_l.jpg

600 x 400 pixels - 49 ko
www.software-e-commerce.com/.../don+t++its+over



Bevan5.jpg

434 x 314 pixels - 19 ko
www.lutsk.ukrpack.net/~scm/Pic/Bevan5.jpg



LoyalWedellBulldozerTarps...

640 x 480 pixels - 87 ko
www.surpluscitysales.com/partytencitycustome...



11683.jpg

300 x 462 pixels - 49 ko
www.kamat.com/.../architecture/common/11683.htm

Problématique : Indexation textuelle d'images sur le web

- Mots du titre, mots les plus fréquents de la page
 - Toutes les images de la page sont indexées avec les mêmes mots
- Nom de l'image, URL
 - Beaucoup d'images portent des noms génériques comme img001.jpg
- Texte de l'attribut ALT de la balise HTML IMG
 - Revient à de l'indexation manuelle : coûteuse et subjective
- Texte proche de l'image
 - Pas toujours en rapport avec l'image

Problématique : Exemple d'indexation d'une image du web

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



trees water house man

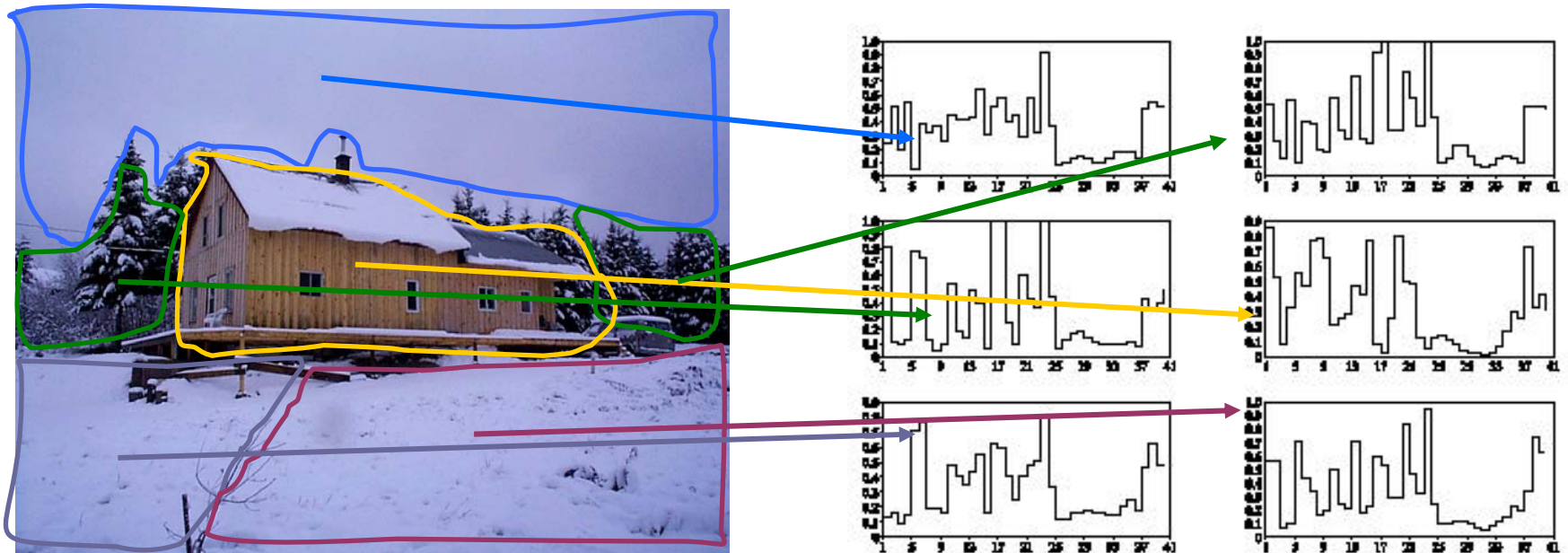
www.zenwaiter.com/house.htm

Recherche d'images : house water man

Problématique : Indexation visuelle d'images généralistes

Couleurs, texture, formes

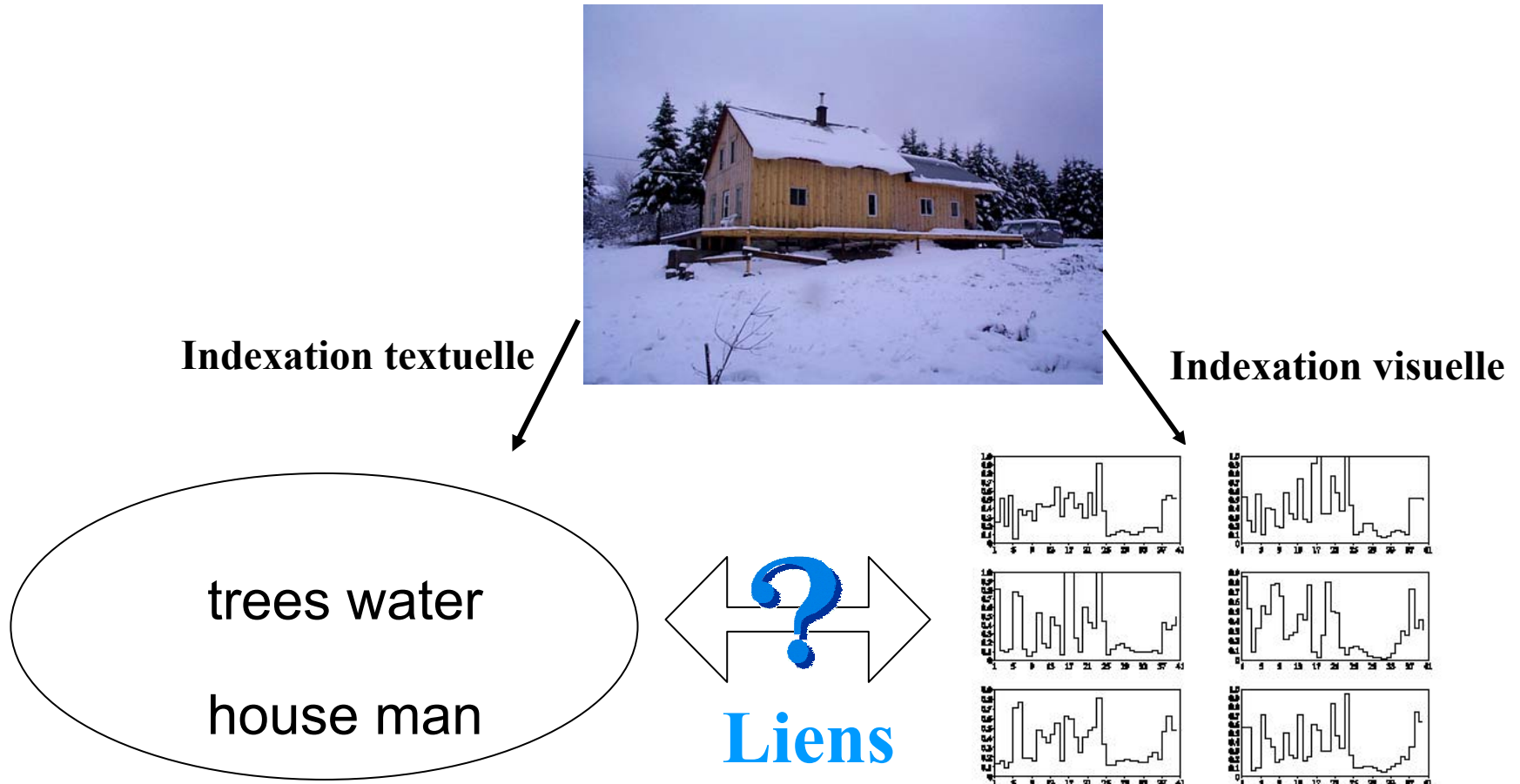
Localisation, régions d'intérêts, segmentation



Chaque région de l'image est appelée « segment visuel » ou « blob ».

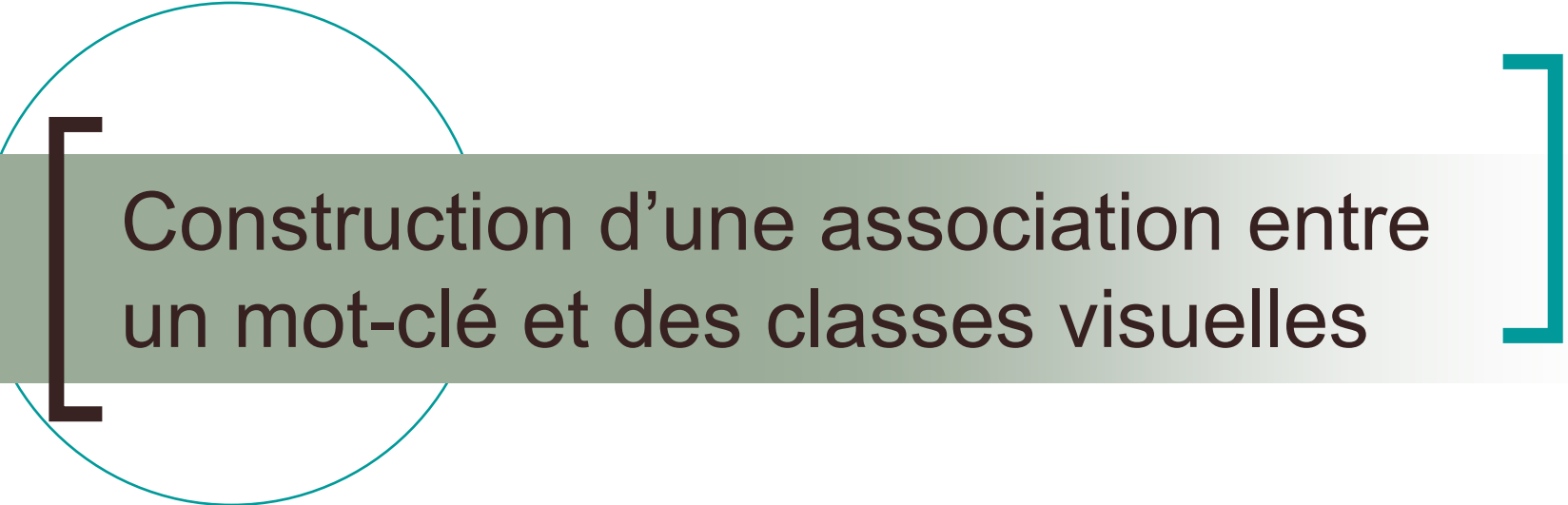
Chaque « blob » peut être décrit par un vecteur dans l'espace multidimensionnel visuel (couleurs, texture, forme,...).

Problématique : Associer des classes visuelles à un mot-clé



Plan

- Problématique
- Construction d'une association entre un mot-clé et des classes visuelles
- Évaluation de cette association
- Résultats expérimentaux
- Application au filtrage d'une indexation textuelle par le contenu visuel
- Conclusion et perspectives



Construction d'une association entre
un mot-clé et des classes visuelles

Schéma général

Corpus d'images indexées
textuellement et visuellement

Séparation aléatoire

Base d'apprentissage

Base de test

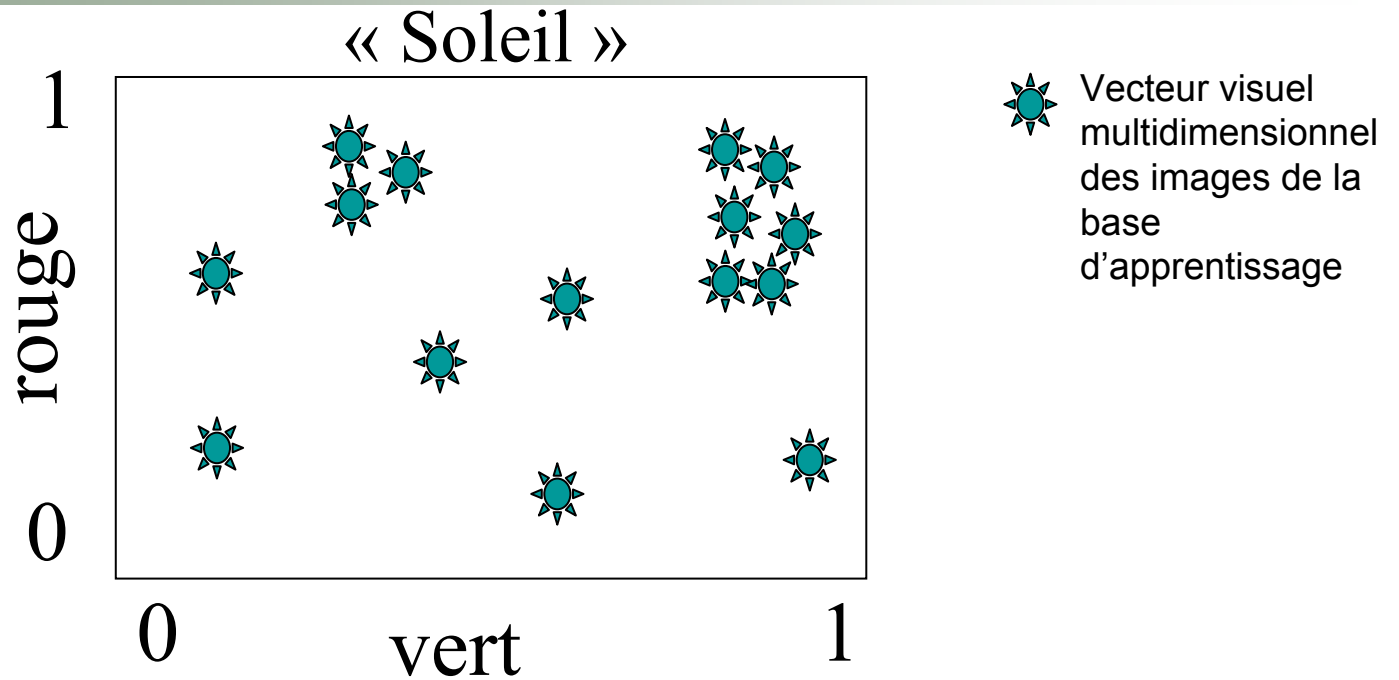
Sélection des images
indexées par le mot étudié

Base d'apprentissage
pour le mot-clé étudié

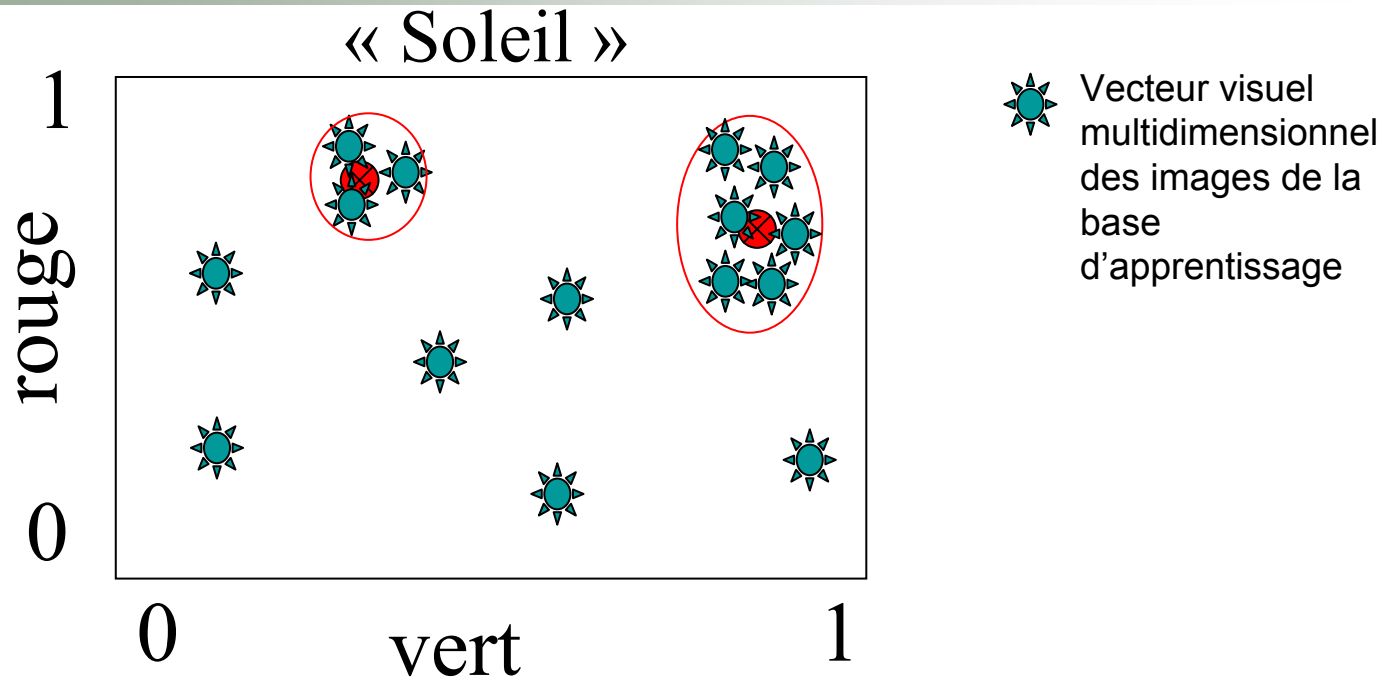
Construction d'une association entre un mot-clé et des classes visuelles

- Pour chaque mot-clé du lexique
 - Construire le sous-ensemble des images de la base d'apprentissage possédant ce mot-clé
 - Rechercher les regroupements de vecteurs visuels des images dans l'espace multidimensionnel au moyen de *clustering* fait par **Classification Ascendante Hiérarchique** (CAH) en prenant :
 - Comme critère d'agrégation, le plus proche voisin
 - Comme critère d'arrêt, la classification qui donne le meilleur score

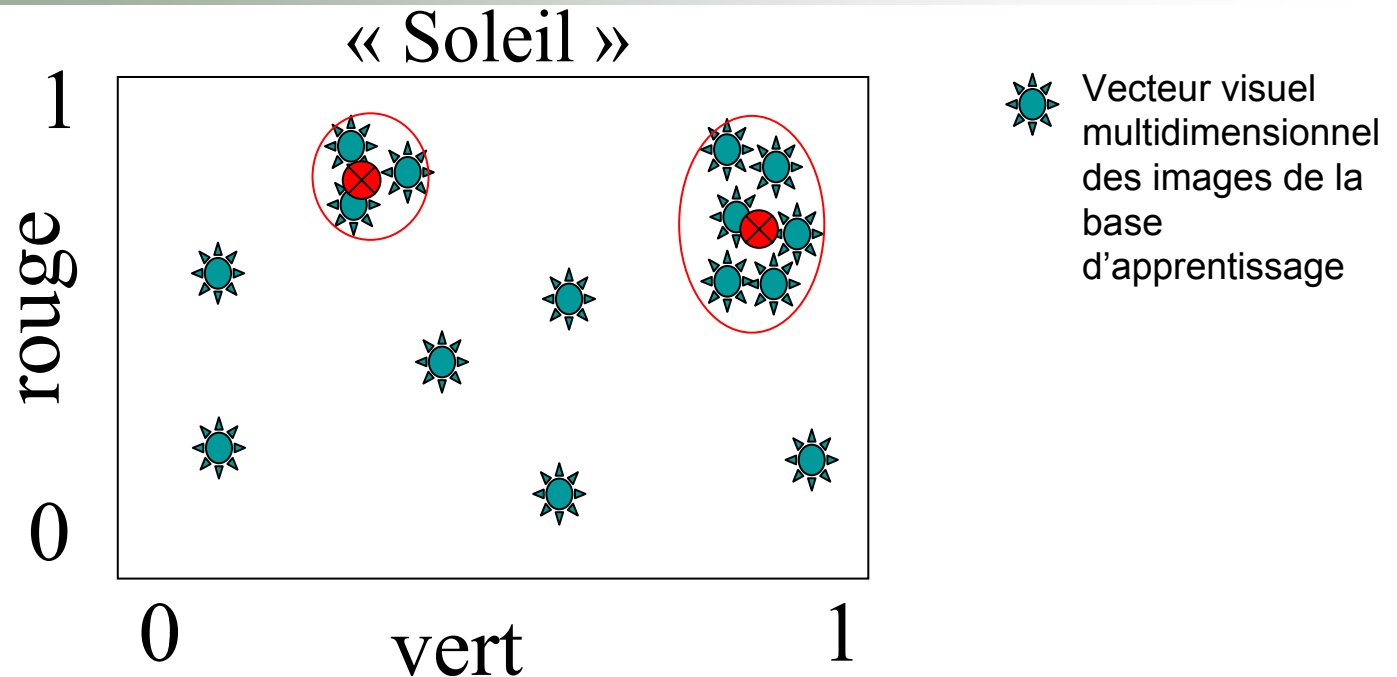
Construction d'une association entre un mot-clé et des classes visuelles



Construction d'une association entre un mot-clé et des classes visuelles

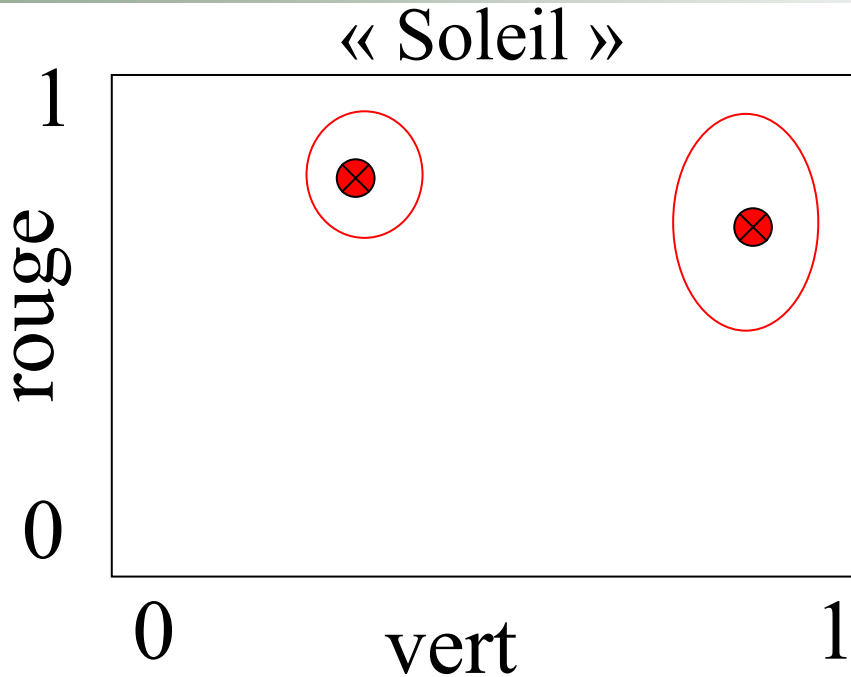


Construction d'une association entre un mot-clé et des classes visuelles



- Pour chaque classe C_k de chaque mot w_i , on garde seulement :
 - le vecteur centroïde : $\bar{C}_k(w_i)$
 - les valeurs des écarts types de la classe pour chaque dimension du vecteur : $\bar{\sigma}(C_k(w_i))$

Construction d'une association entre un mot-clé et des classes visuelles



- Pour chaque classe C_k de chaque mot w_i , on garde seulement :
 - le vecteur centroïde : $\bar{C}_k(w_i)$
 - les valeurs des écarts types de la classe pour chaque dimension du vecteur : $\bar{\sigma}(C_k(w_i))$



Évaluation de l'association

Schéma général

Corpus d'images indexées
textuellement et visuellement

Séparation aléatoire

Base d'apprentissage

Base de test

Sélection des images
indexées par le mot étudié

Base d'apprentissage
pour le mot-clé étudié

Construction de
classes visuelles

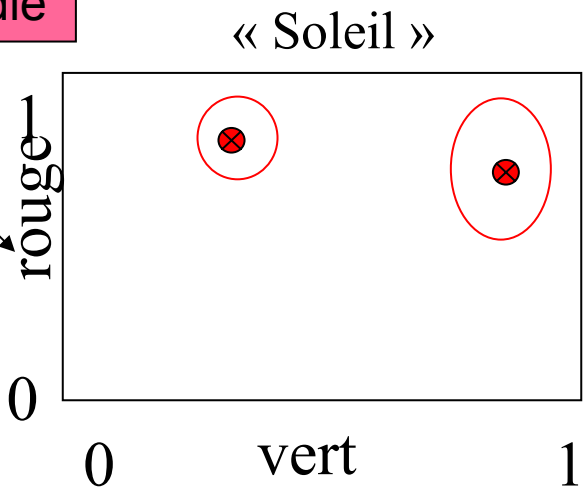
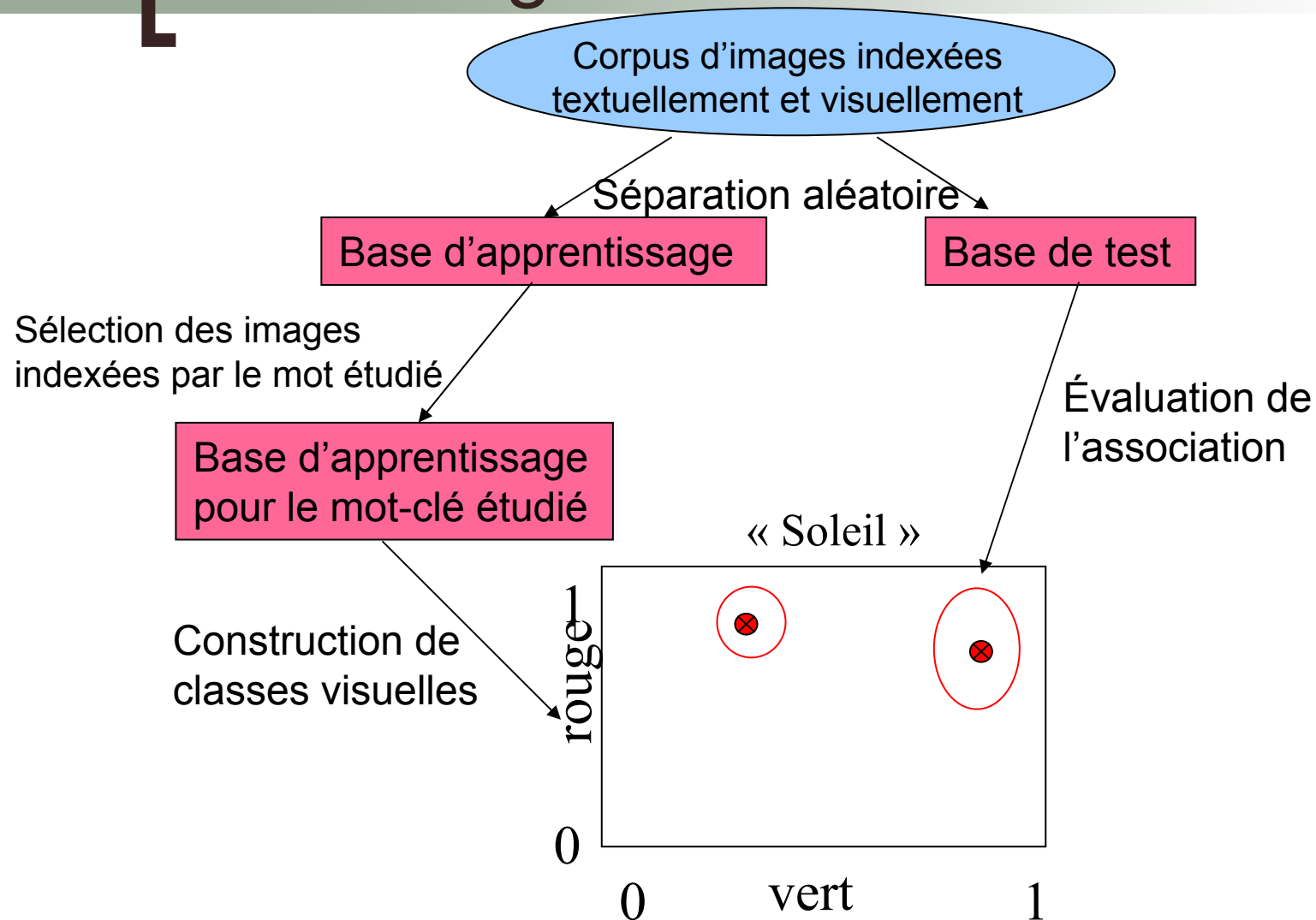
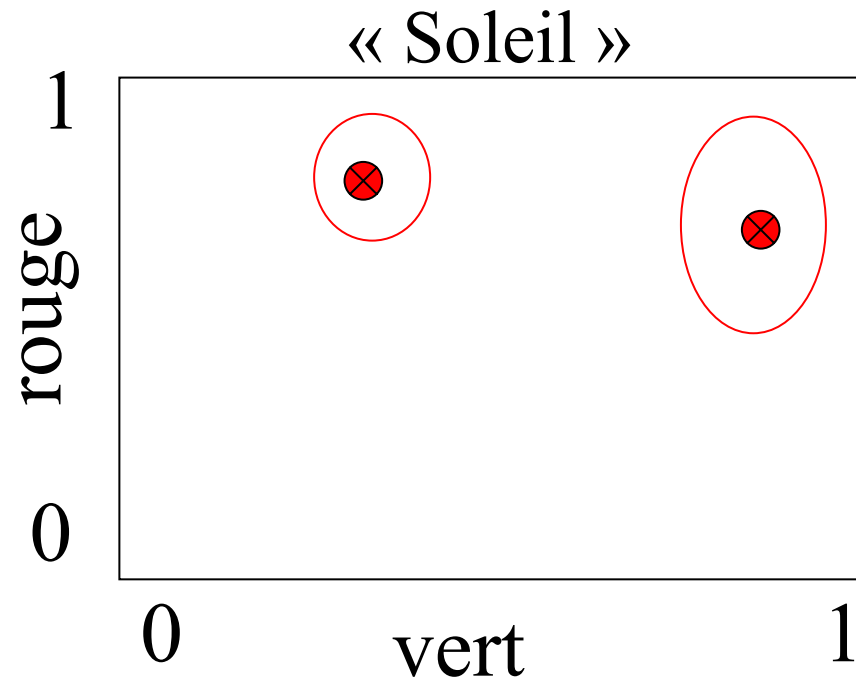


Schéma général

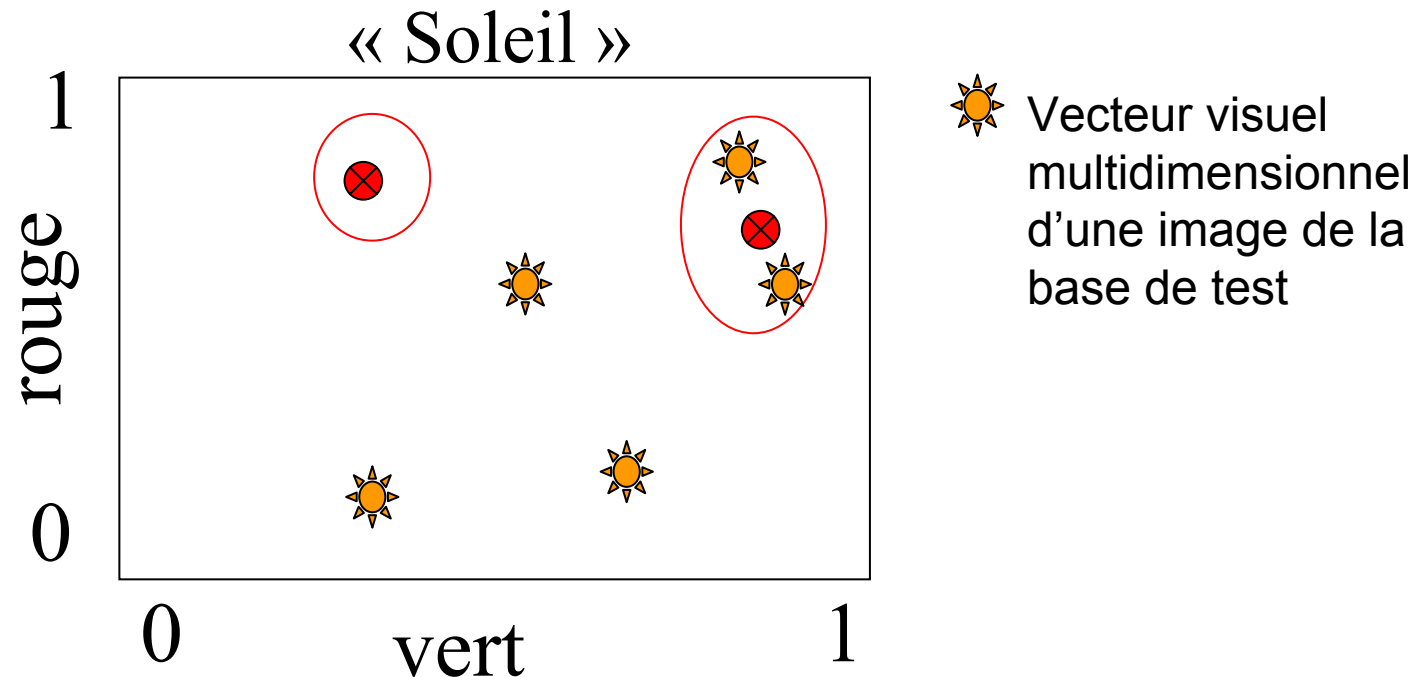


Évaluation de l'association : Classer les images de test



- Classer les images de la base de test
- Pour chaque image, compter le nombre de blobs contenus dans les classes visuelles du mot.

Évaluation de l'association : Classer les images de test



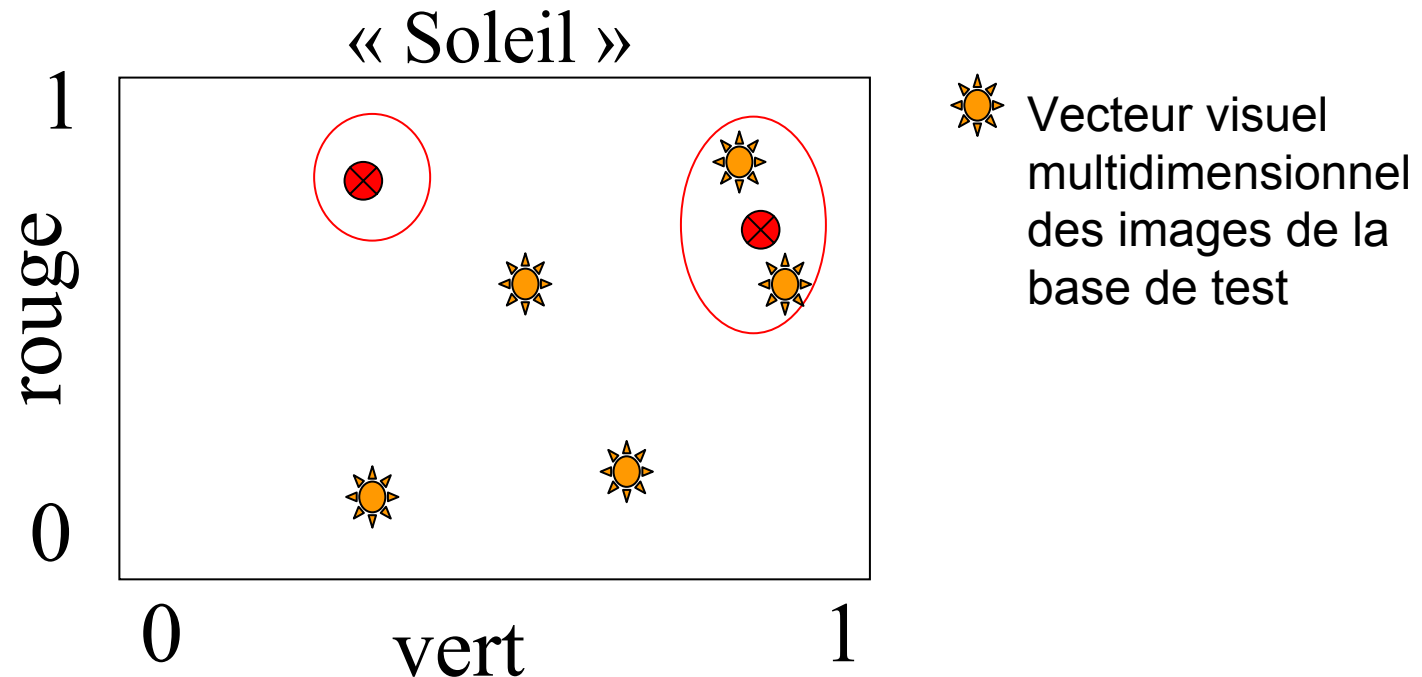
- Classer les images de la base de test
- Pour chaque image, compter le nombre de blobs contenus dans les classes visuelles du mot.
- Ici, 2 blobs sur 5 sont dans les classes visuelles du mot « soleil ».

Évaluation de l'association : Formalisation de l'appartenance d'un vecteur de test à une classe

- Un blob b_j de test appartient à la classe C_k du mot w_i ssi pour toute dimension visuelle p , le vecteur visuel du blob est compris dans l'hyper-rectangle définie par le vecteur centroïde et le vecteur d'écart types.

$$b_j \in C_k(w_i) \text{ ssi } \forall p \left| C_{k,p}^-(w_i) - \vec{V}_p(b_j) \right| \leq D \times \sigma_p(C_k(w_i))$$

Évaluation de l'association : Associer un mot-clé à une image



- Un mot est associé à un blob si ce blob est dans l'une des classes visuelles de ce mot.
- Un mot est associé à une image si B blobs de cette image appartiennent aux classes visuelles de ce mot.

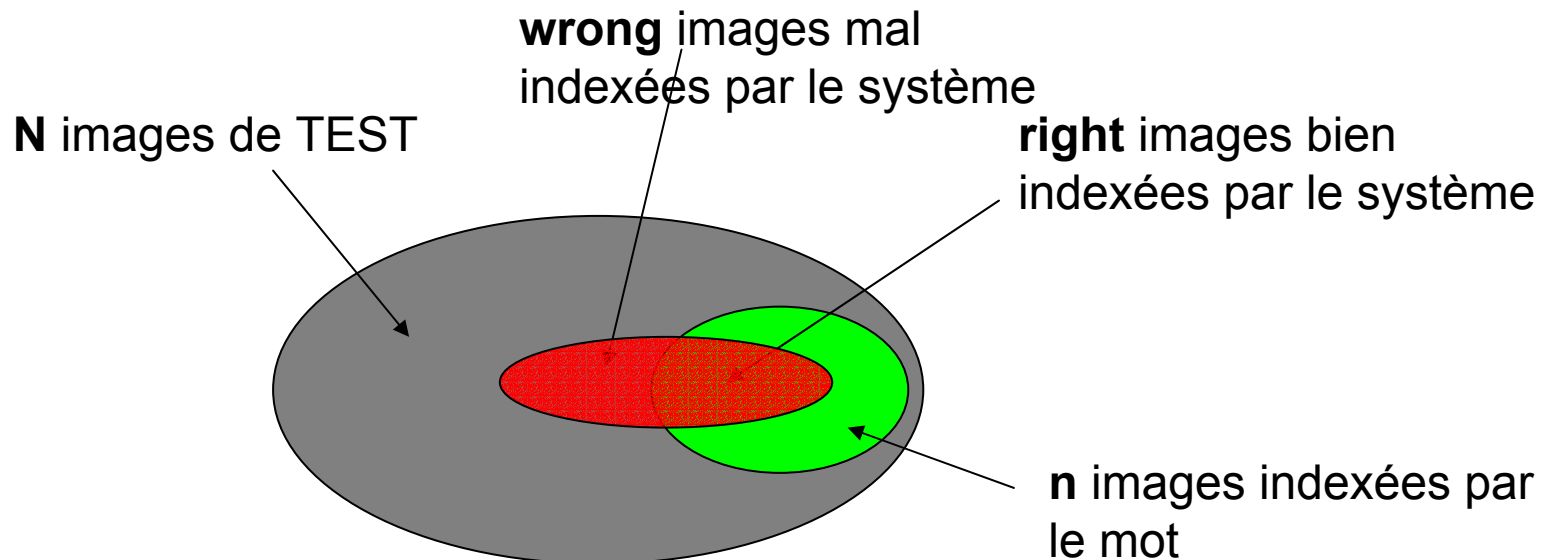
Évaluation de l'association : Calcul du score de la classification



Pour chaque mot, on peut calculer le score « **Normalized Score** » :

$$\text{Score NS} = \frac{\text{right}}{n} - \frac{\text{wrong}}{N-n}$$

sensibilité 1-spécificité





Résultats expérimentaux

Corpus :

Bases d'images COREL

- 10 000 images
- 161 mot-clés différents en anglais
- Chaque image possède :
 - De 1 à 5 mot-clés choisis manuellement
 - De 2 à 10 « blobs », des blobs 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),...)

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.

http://vision.cs.arizona.edu/kobus/research/data/jmlr_2003/index.html

<http://wang.ist.psu.edu/docs/home.shtml>

Corpus : Bases d'images COREL

- 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

Exemples d'images de la base d'apprentissage possédant le mot « horse »



field horse tree



farm horse



field horse



horse street town



horse



head horse



farm horse



field horse



horse



field horse



farm field hills horse



horse



horse people town



head horse sky



farm fence horse



fence horse



field horse



field horse



fence horse



field forest horse



horse people



horse people



grass horse tree



horse people tree



horse people tracks



field grass horse tree



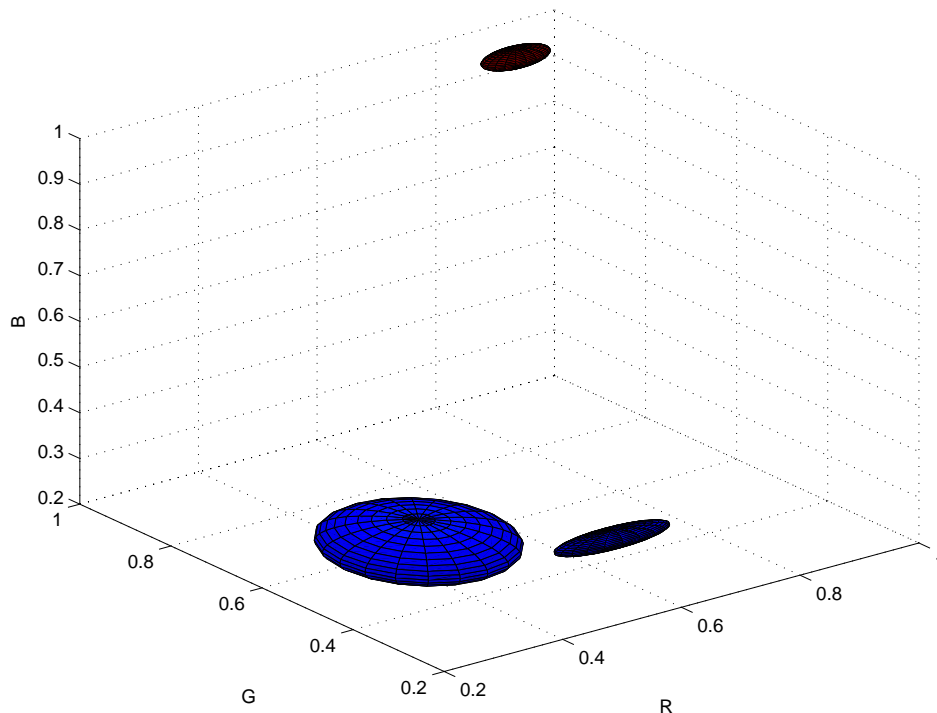
horse water



horse people sky

Exemple : classes visuelles construites à partir des images de la base d'apprentissage possédant le mot « horse »

« horse »

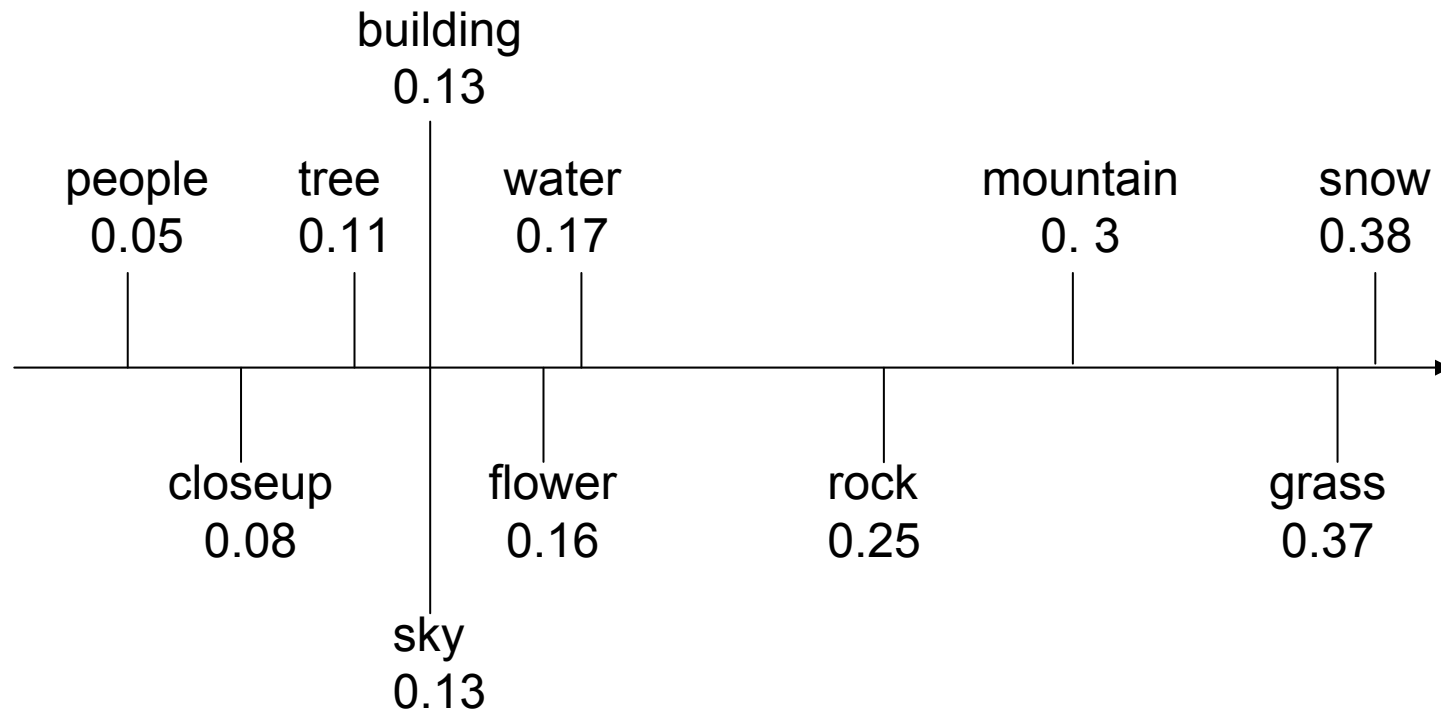


- Nombres d'images de la base d'apprentissage :
 - 6998 parmi lesquelles 173 sont indexées par « horse »
- Nombres d'images de TEST :
 - 2922 parmi lesquelles 96 sont indexées par « horse »
- Nombre de classes obtenues :
 - 3 classes visuelles
- Right : 55 sur 96
- Wrong : 879 sur 2826
- Sensibilité : 0.57
- Spécificité : 0.69
- Score NS : 0.26

Résultats sur 2922 images de TEST

Mot	Nombre images TEST	D	Nombre de classes	Sensibilité	Spécificité	Score NS
Snow	174	3	1	0.69	0.69	0.38
Grass	233	2.5	1	0.65	0.72	0.37
Mountain	191	3	1	0.61	0.69	0.3
Rock	181	2.5	1	0.65	0.6	0.25
Water	583	3	1	0.73	0.43	0.17
Flower	222	2	1	0.66	0.5	0.16
Sky	532	3.5	1	0.45	0.68	0.13
Building	242	3	4	0.86	0.27	0.13
Tree	534	3	3	0.89	0.23	0.11
Closeup	202	2.5	4	0.77	0.31	0.08
People	491	2.5	1	0.91	0.14	0.05

Notions de « consistance visuelle »



Certains mots sont plus facilement détectables dans une image que d'autres.



Application : filtrage de
l'indexation textuelle d'images

Application :

Filtrage de l'indexation textuelle d'une image du web

- **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



Application : Méthodologie

■ Problème :

- Nous ne disposons pas d'une base d'images du web bien indexée pour évaluer la performance

■ Idée :

- A chaque image de test, on associe les 161 mots du lexique, puis on les filtre avec les classes visuelles de chaque mot, on évalue alors le nombre de bons mots associés à l'image à l'aide du score NS

Calcul du score

Pour chaque image de test, on peut calculer le score « **Normalized Score** » :

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

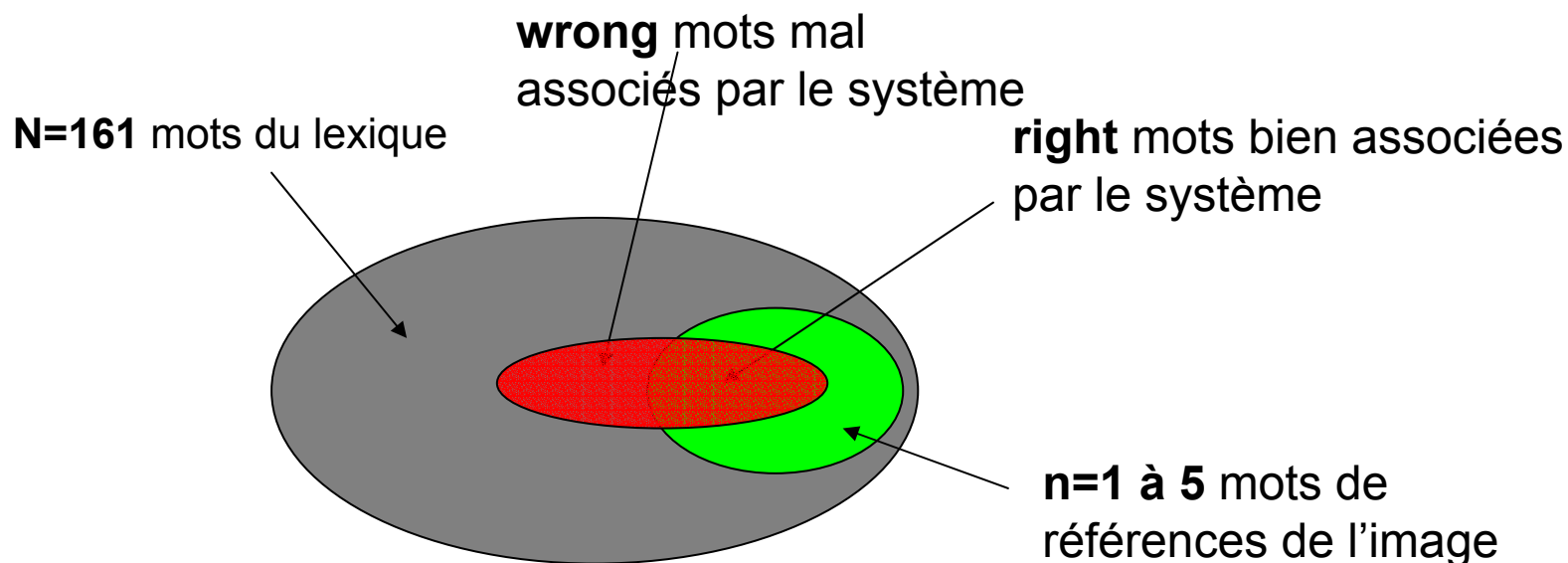




Image 184059 (9 blobs)

Mots de Références présents (3 sur 4)

snow(6) ice(6) iceberg(3)

people

sensi=0.75 specif=0.64

preci=0.05 NS=0.39

Index (59 mots acceptés sur 111 présents parmi 161)

car(7) ruins(6) kauai(6) texture(6) sign(6) desert(6)
parade(6) cliff(6) reflection(6) snow(6) sand(6) pattern(6)
horizon(6) statue(6) stone(6) woodland(6) military(6) ice(6)
sheep(6) hat(6) temple(6) ski(5) design(5) fox(5) mountain(5)
race(4) saguaro(4) boat(4) sky(4) sailboat(4) hippo(4)
branch(4) building(4) river(4) valley(4) feline(4) street(4)
beach(4) guard(4) water(3) deer(3) church(3) light(3) hills(3)
tower(3) bobcat(3) lion(3) town(3) lizard(3) frost(3) coast(3)
sea(3) road(3) art(3) rock(3) petals(3) iceberg(3) wall(3)
crystal(3)

Non-Index (52+50=102 parmi 161)

museum(2) bird(2) landscape(2) head(2) wave(2) city(2)
park(2) wings(2) tree(2) eagle(2) sculpture(2) wood(2)
grass(2) entrance(2) cougar(2) perch(2) vegetable(2)
canyon(2) fence(2) coyote(2) field(2) cat(1) bighorn(1)
hawk(1) fish(1) ram(1) palace(1) food(1) tiger(1) sail(1)
reef(1) wildlife(1) forest(1) door(1) closeup(1) rhino(1)
wolf(1) penguin(1) log(1) horse(1) cactus(1) lynx(1) house(1)
ground(1) courtyard(1) shadow(1) village(1) sunset(1)
rose(1) bengal(1) pumpkin(1) shore(1) baby background
bear bridge bush cheetah cloud column coral costume
display dune elephant face flower foal fungus garden glass
harbor kitten leaf leopard lichen mane mare market model
mushroom nest night ocean orchid owl people pillar plane
plants pyramid reptile restaurant rodent shop snake textile
trunk tulip window wine woman

Application :
Résultats sur les 2922 images de test

Méthode	Sensibilité	Spécificité	NS
BESTALLNS	0.70	0.59	0.29
BESTSENSINS	0.77	0.50	0.27

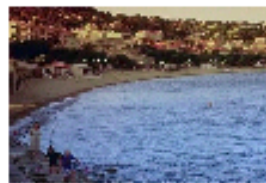
- Score NS moyens sur l'ensemble des 2922 images de test
- BESTALLNS : classes visuelles qui maximise le score NS
- BESTSENSINS : classes visuelles qui maximise la sensibilité



costume(1) people (0)
 sensi=0.00
 specif=0.51 NS=-0.49
 78 mots sur 161



wall(9)
 sensi=1.00 specif=0.17
 NS=0.18
 133 mots sur 161



building(3) water(3)
 landscape(3)
 sunset(2)
 sensi=0.75
 specif=0.50 NS=0.25
 81 mots sur 161



sky(3) pattern(1)
 cloud (0)
 sensi=0.33
 specif=0.64
 NS=-0.03
 58 mots sur 161



leaf(5) flower(3)
 rose(3) plants(2)
 sensi=0.75
 specif=0.91 NS=0.66
 17 mots sur 161



tree(3) water(2)
 sensi=0.50
 specif=0.67
 NS=0.17
 54 mots sur 161



field(4) building(4)
 tree(4)
 sensi=1.00
 specif=0.57 NS=0.57
 71 mots sur 161



tree(5) forest(4) cat(4)
 sensi=1.00 specif=0.24
 NS=0.24
 123 mots sur 161



people(1) woman (0)
 sensi=0.00
 specif=0.28
 NS=-0.72
 114 mots sur 161



closeup(1) crystal (0)
 pattern (0) texture (0)
 sensi=0.00
 specif=1.00
 NS=0.00
 0 mots sur 160



field(4) grass(4)
 sensi=1.00 specif=0.26
 NS=0.26
 119 mots sur 161



mountain(3) sky(3)
 tree(2) snow(1)
 sensi=0.50
 specif=0.77
 NS=0.27
 38 mots sur 161



market(4) people (0)
 street (0)
 sensi=0.33 specif=0.94
 NS=0.27
 11 mots sur 160



market(8) people (0)
 sensi=0.50
 specif=0.35 NS=-0.15
 104 mots sur 161



leaf(7) flower(3)
 rose(3) plants(2)
 sensi=0.75
 specif=0.78 NS=0.53
 37 mots sur 161



people(5) street (0)
 sensi=0.50
 specif=0.99
 NS=0.49
 2 mots sur 160



ocean(6) closeup(3)
 sea (0)
 sensi=0.67 specif=0.74
 NS=0.41
 43 mots sur 161



cat(5) lynx(5)
 feline(5)
 sensi=1.00
 specif=0.20 NS=0.20
 129 mots sur 161

Conclusion

- Méthode permettant d'associer un mot-clé à des classes visuelles
- Notion de « consistance visuelle »
- Filtrage d'indexation textuelle par le contenu visuel

- Avantages :
 - Extraction des descripteurs visuels dans la phase « indexation »
 - Les méthodes traditionnelles en RI peuvent être utilisées
 - Stockage uniquement des vecteurs centroïdes et d'écart types des classes visuelles

Perspectives

- Proposer une association entre mot-clé et classes visuelles plus fine et non-booléenne qui permette d'ordonner les résultats
- Réduire le nombre de dimensions visuelles pour construire de meilleures classes visuelles en fonction du mot