



Sélection adaptative des descripteurs visuels et dérivation de métadescripteurs contextuels dépendant du mot-clé pour l'indexation automatique d'images

Sabrina Tollari, Hervé Glotin
Laboratoire LSIS - Equipe INCOD
UMR CNRS 6168
Université du Sud Toulon-Var

Atelier MetSI2005
Grenoble, le 24 mai 2005

Problématique : Recherche d'images web par mots clés



[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 : 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.



www.zenwaiter.com/house.htm

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.



www.zenwaiter.com/house.htm

Recherche d'images : house water man

Problématique

K. Barnard, P. Duygulu, N. de Freitas, D. Forsyth, D. Blei, and M. I. Jordan : « It remains an interesting open question to construct feature sets that (...) offer very good performance for a particular vision task », *Matching words and Pictures*, Journal of Machine Learning Research, 2003

- Quels traits visuels doit-on utiliser pour construire des systèmes de recherche d'images qui prennent en compte aussi la sémantique étant donné :
 - le problème du trop grand nombre de dimensions [Berrani, Amsaleg, IRISA, 2004] et
 - l'existence de traits visuels anti-mots ?

Plan

- Problématiques
- Description du corpus
- Description d'une méthode d'évaluation pour le choix des traits visuels
- Proposition d'une méthode de choix des traits visuels
- Résultats expérimentaux
- Conclusion et perspectives

Corpus : Bases d'images COREL

- 10 000 images
- 250 mot-clés environs 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.

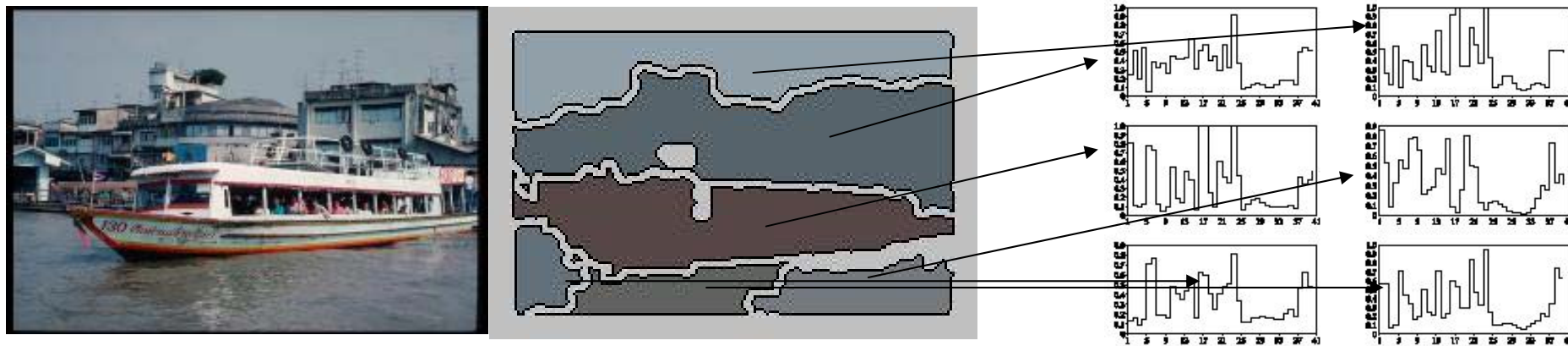
http://vision.cs.arizona.edu/kobus/research/data/jmlr_2003/index.html

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

Corpus : Segmentation des images

- Exemples de segmentation par « normalized cuts »

J. Shi, J. Malik, « Normalized Cuts and Image Segmentation », IEEE on Patterns Analysis and Machine Intelligence, vol.22, n°8, 2000



water, boat harbor building

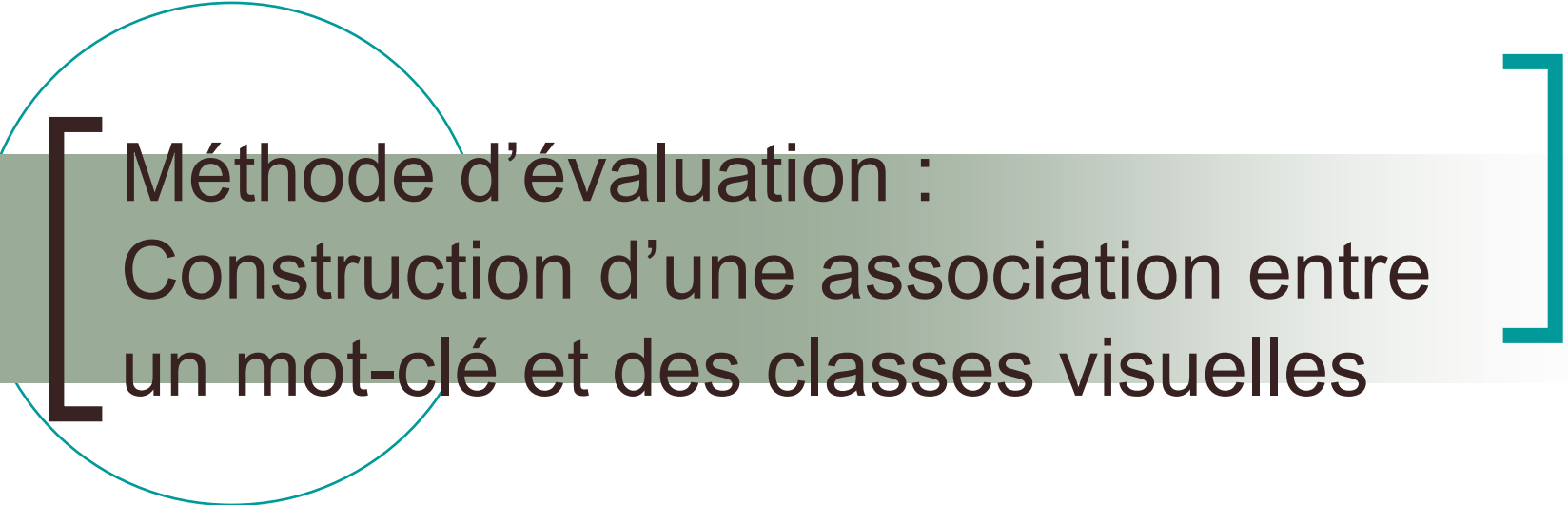
- Normalisation du corpus :

par estimation MLE de distributions Gamma des vecteurs visuels pour la génération de distributions de probabilités. Les valeurs sont comprises entre 0 et 1.

Corpus :

Exemples d'images de la base COREL



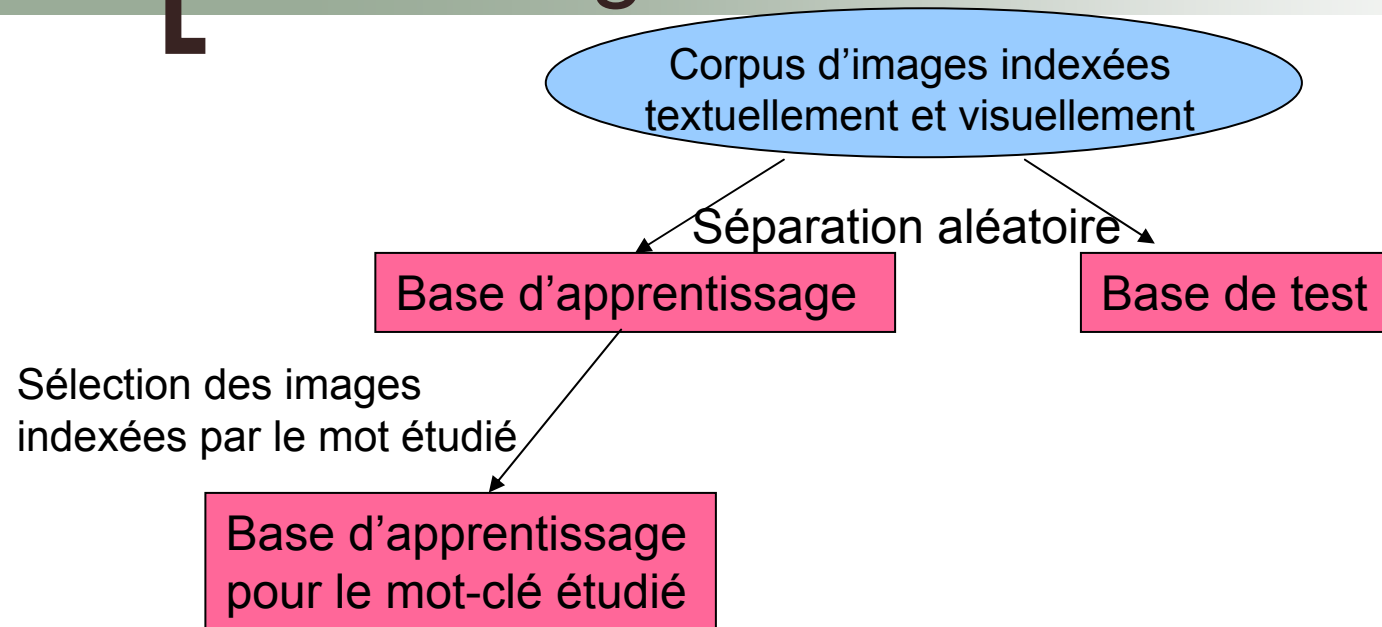


Méthode d'évaluation :
Construction d'une association entre
un mot-clé et des classes visuelles

Schéma général

Corpus d'images indexées
textuellement et visuellement

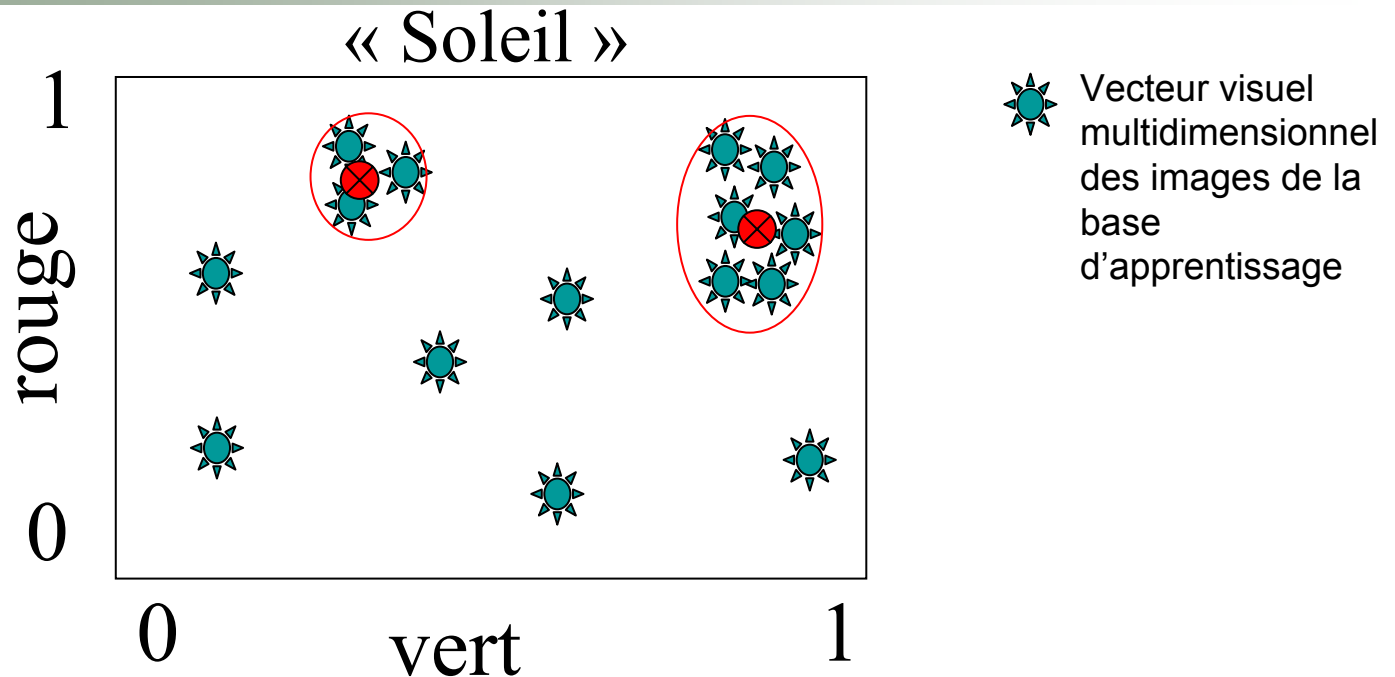
Schéma général



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

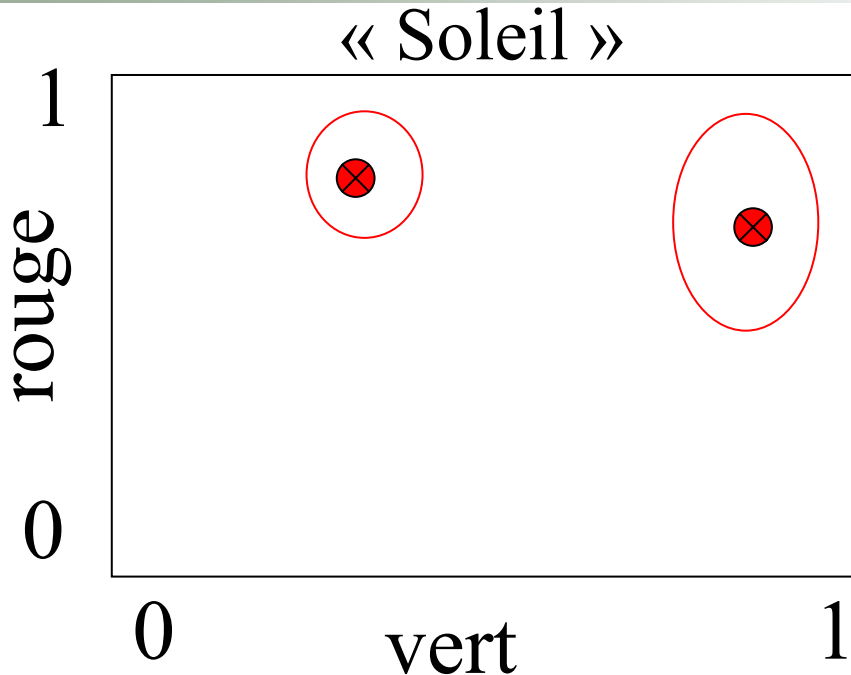
- Pour chaque mot-clé du lexique
 - Construire le sous-ensemble des images de la base d'apprentissage possédant ce mot-clé
 - Rechercher les regroupements de vecteurs visuels des images dans l'espace multidimensionnel au moyen de *clustering* fait par **Classification Ascendante Hiérarchique** (CAH) 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



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

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



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



Évaluation de l'association

Schéma général

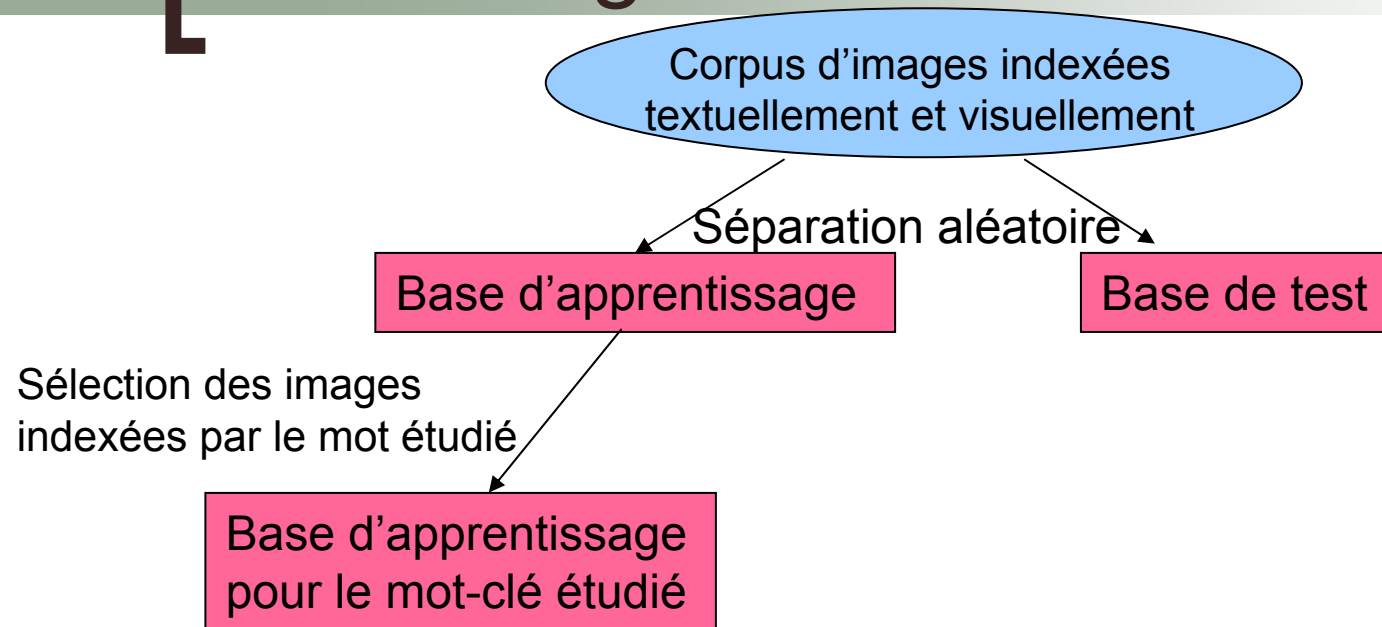
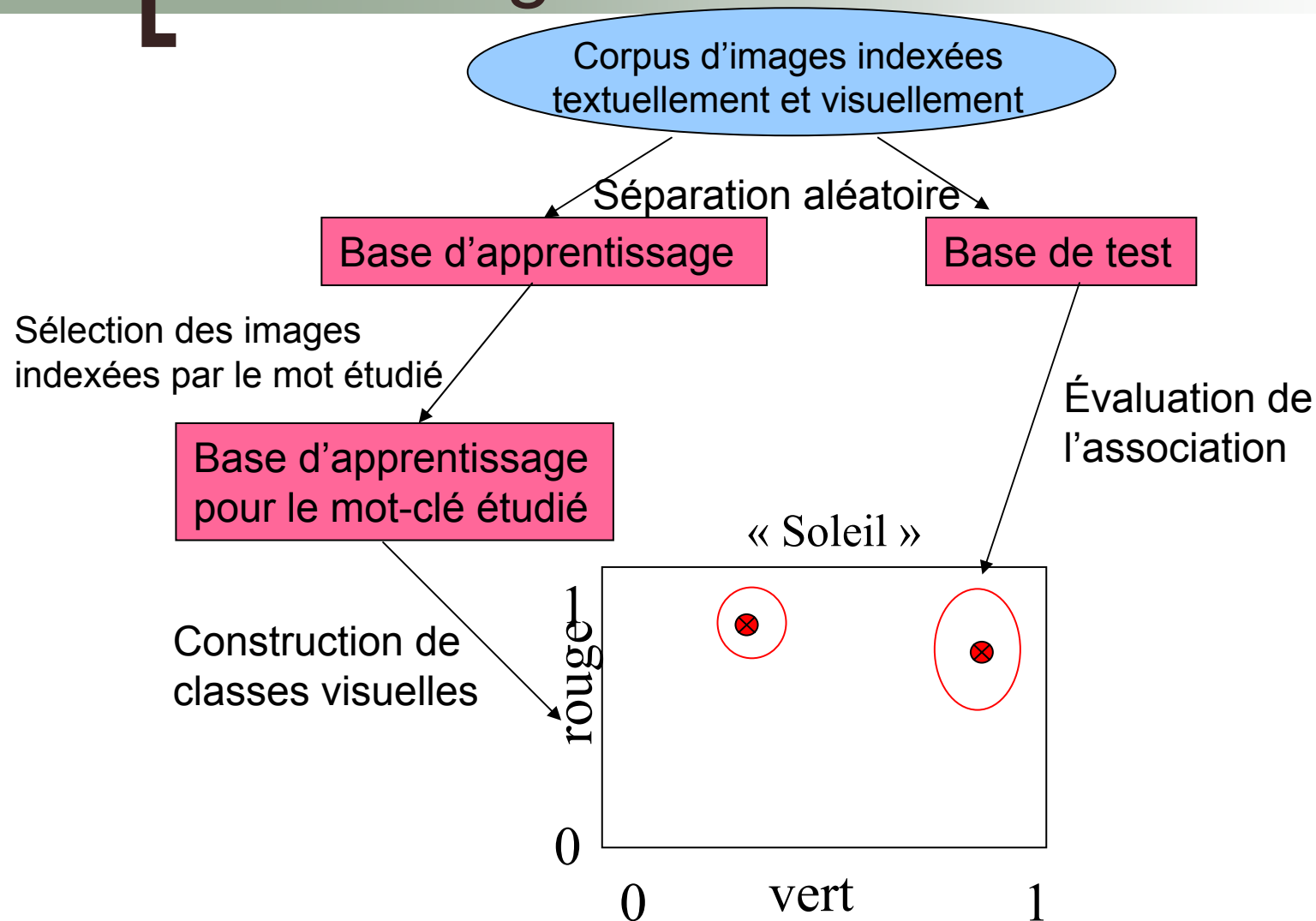
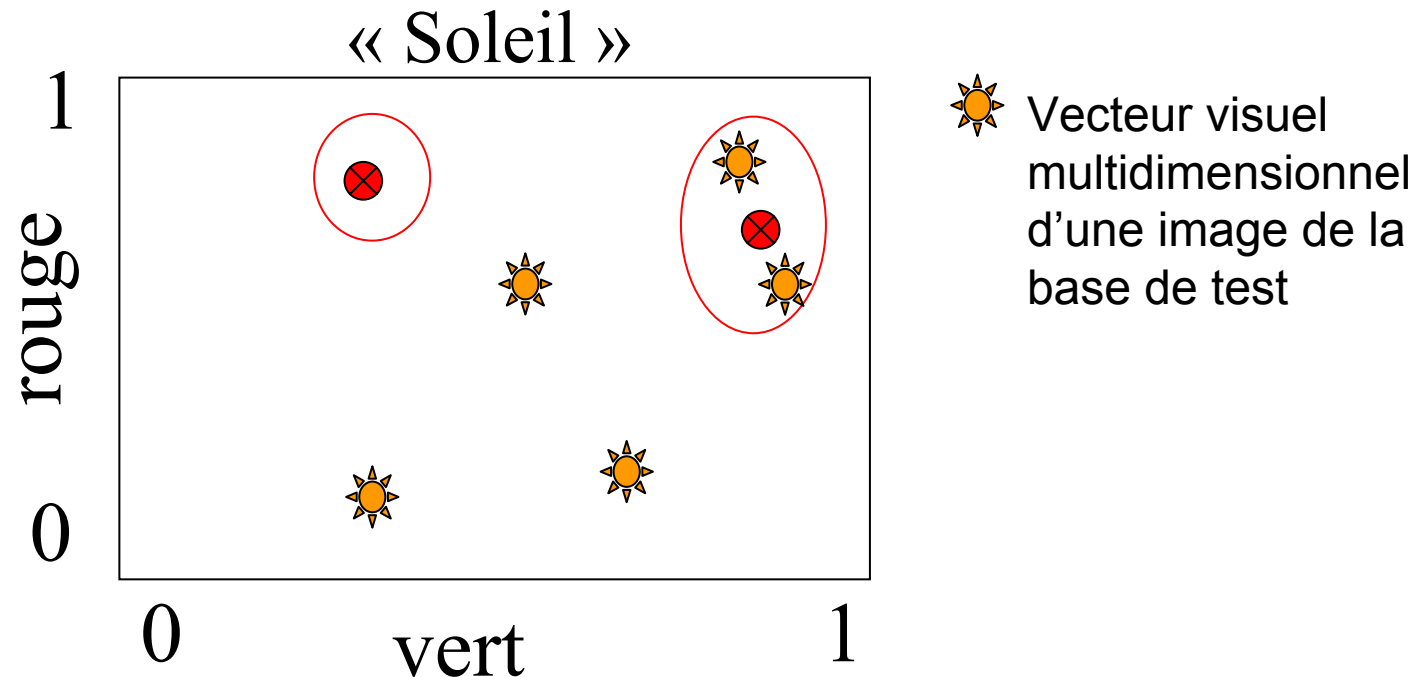


Schéma général



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



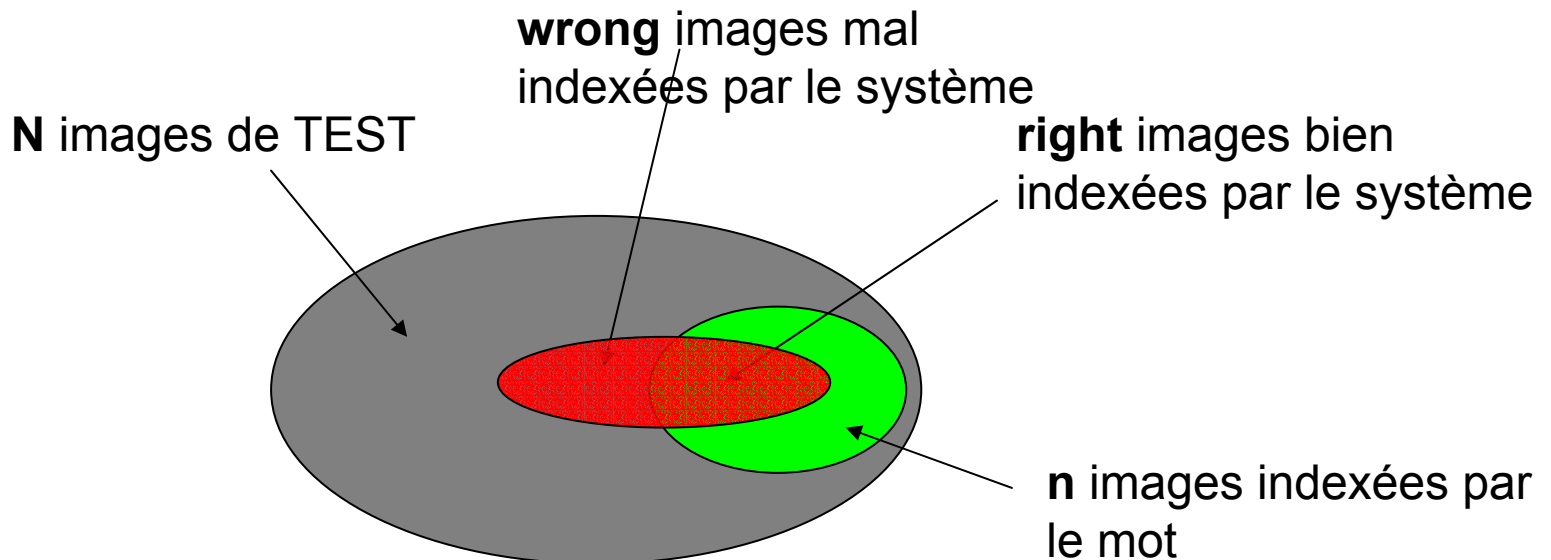
- Un mot est associé à un blob si ce blob est dans l'une des classes visuelles de ce mot.
- Un mot est associé à une image si B blobs de cette image appartiennent aux classes visuelles de ce mot.

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

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

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

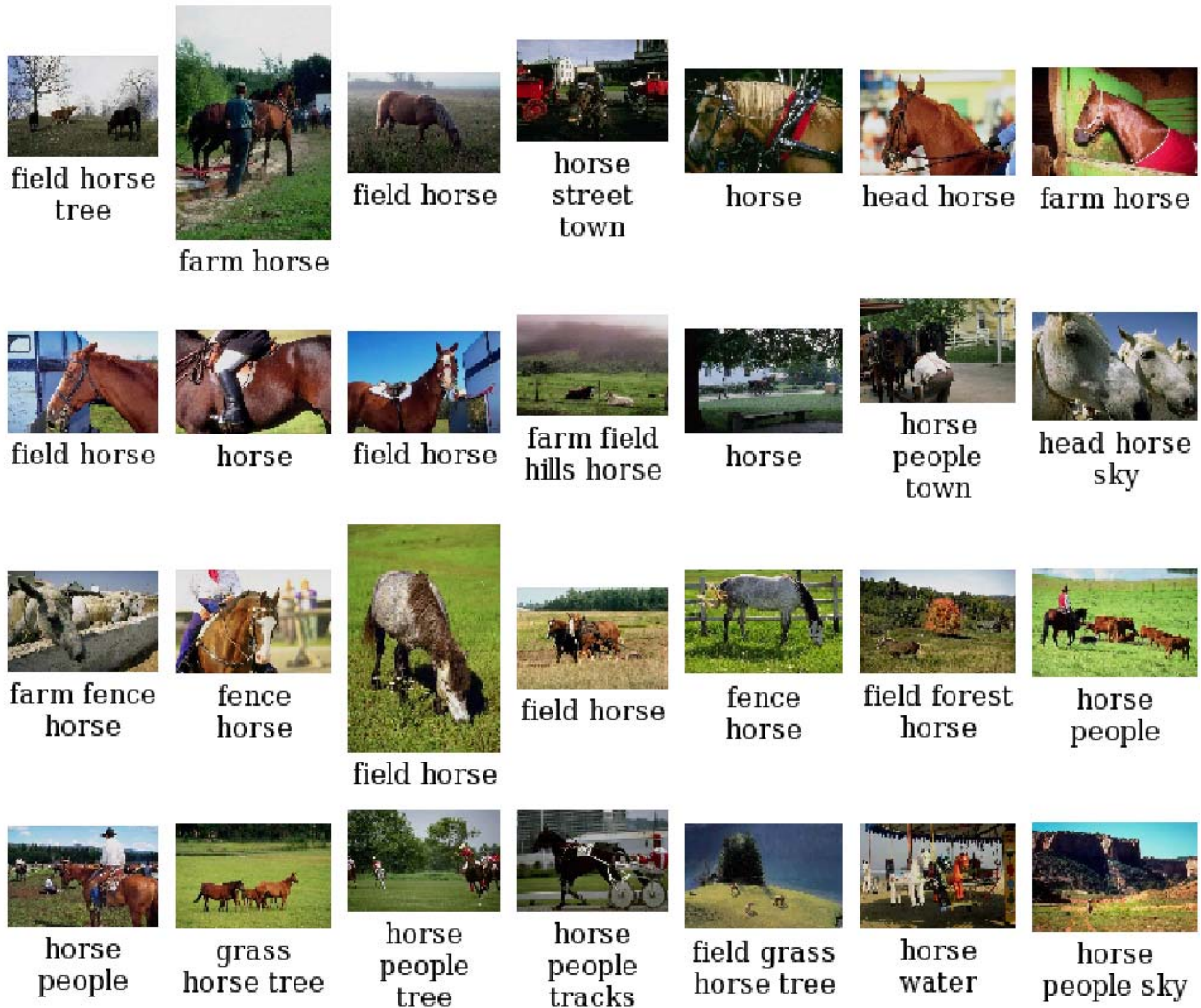
sensibilité 1-spécificité





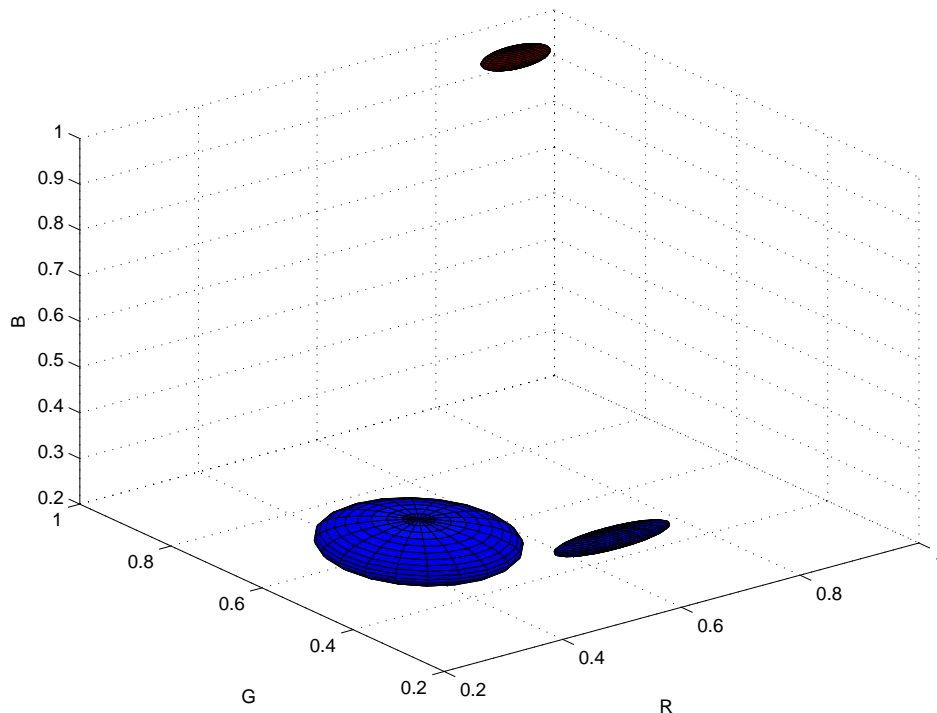
Résultats expérimentaux

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



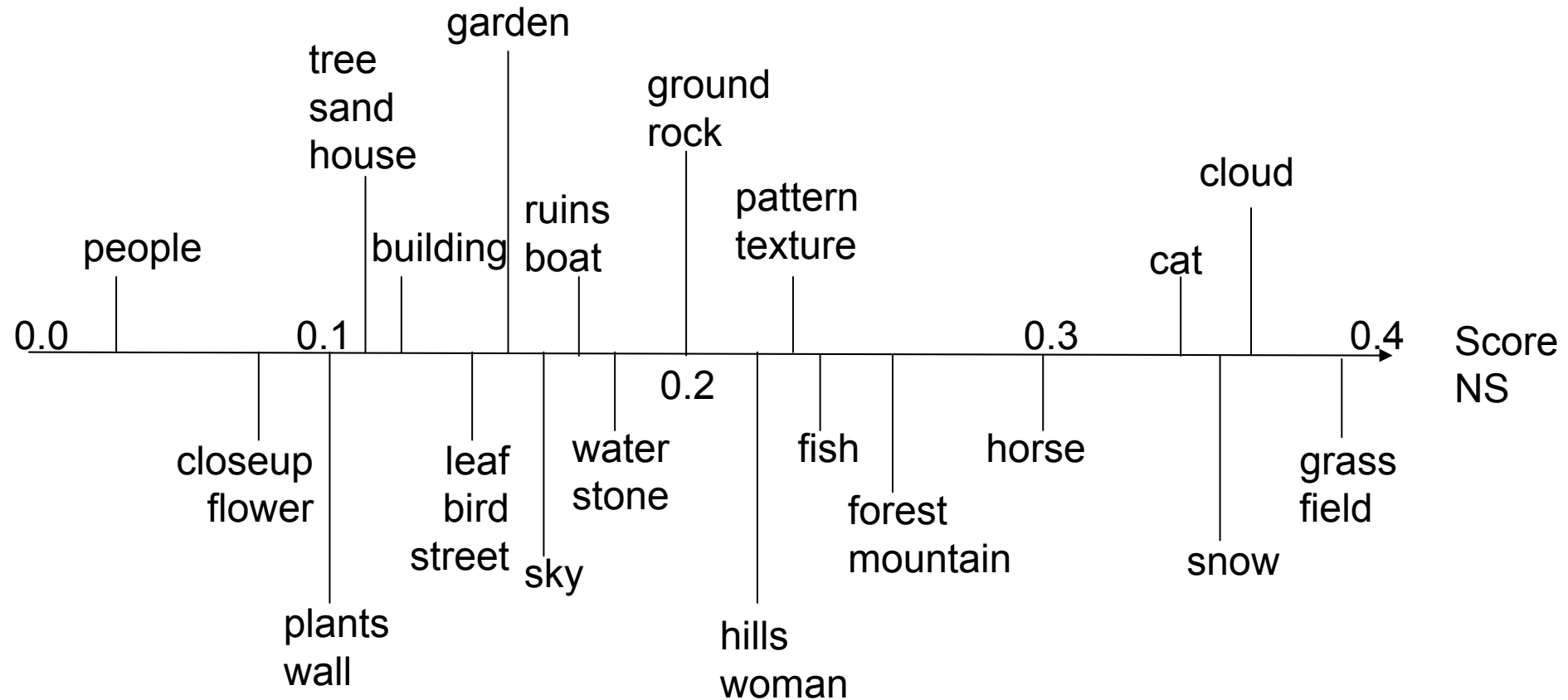
Exemple : classes visuelles construites à partir des images de la base d'apprentissage possédant le mot « horse » sur l'espace visuel complet : 40DIM

« horse »



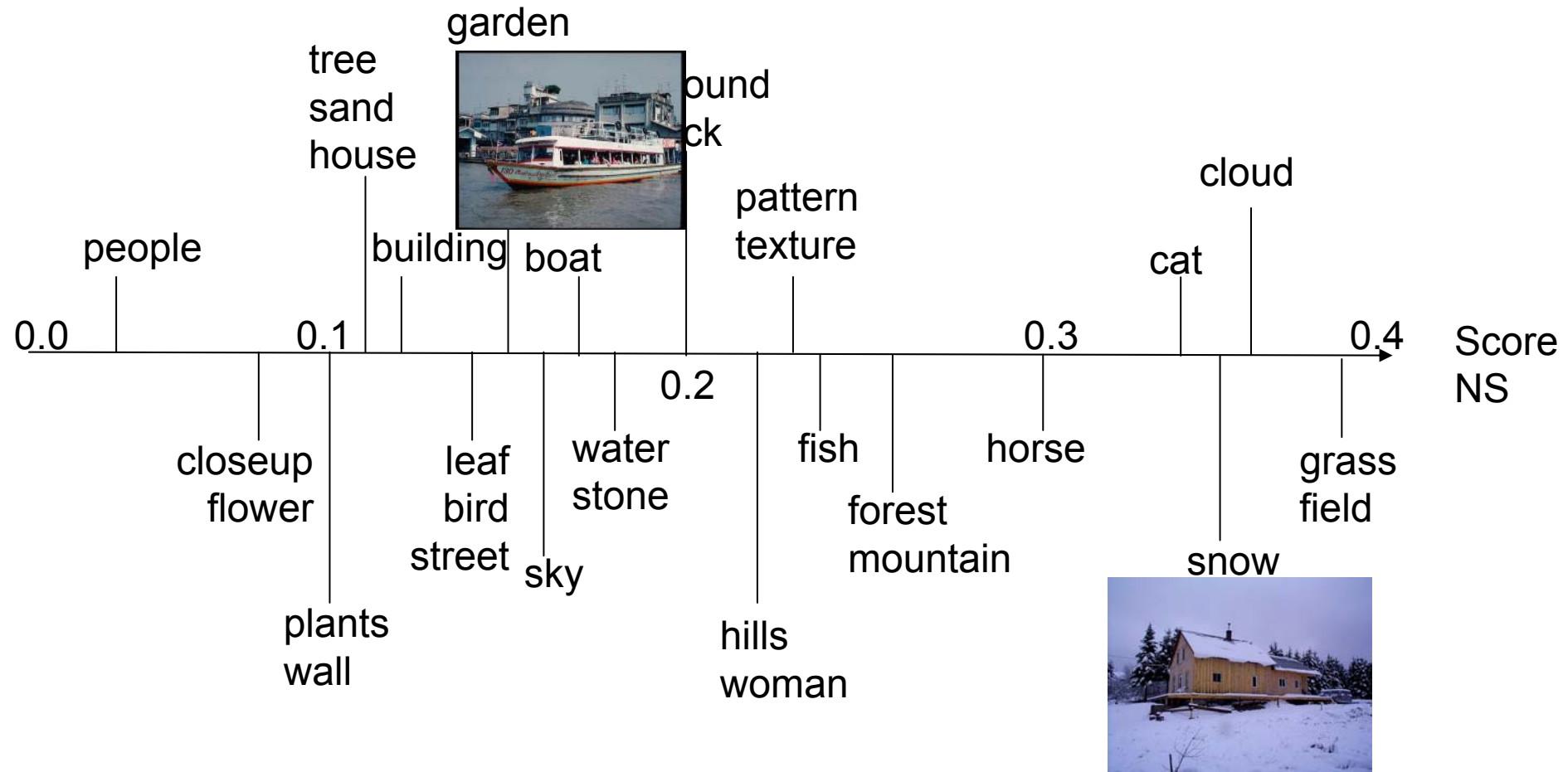
- Nombres d'images de la base d'apprentissage :
 - 4909 parmi lesquelles 133 sont indexées par « horse »
- Nombres d'images de test :
 - 2491 parmi lesquelles 78 sont indexées par « horse »
- Nombre de classes obtenues :
 - 3 classes visuelles
- Right : 37 sur 78
- Wrong : 420 sur 2413
- Sensibilité : 0.47
- Spécificité : 0.82
- Score NS : 0.30

Notion de « consistance visuelle »



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

Notion de « consistance visuelle »



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



Problème : quels traits visuels choisir ?

Motivations :

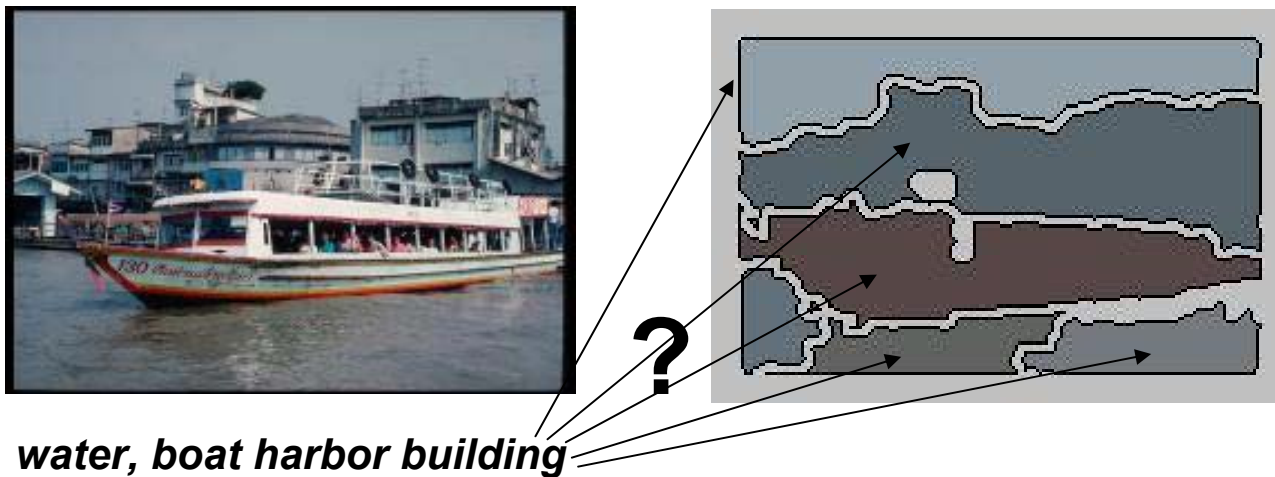
- le problème du trop grand nombre de dimensions
- l'existence de traits visuels non-informatifs
- l'existence de traits visuels anti-informatifs

Résultats pour des choix de réductions « naïfs »

	40DIM	LABSTD	COLOR
Nombre de dimensions	40	6	18
NS Moyen	0.22	0.05	0.10
NS Minimum	0.04	0.00	-0.01
NS Maximum	0.54	0.18	0.23

Difficulté

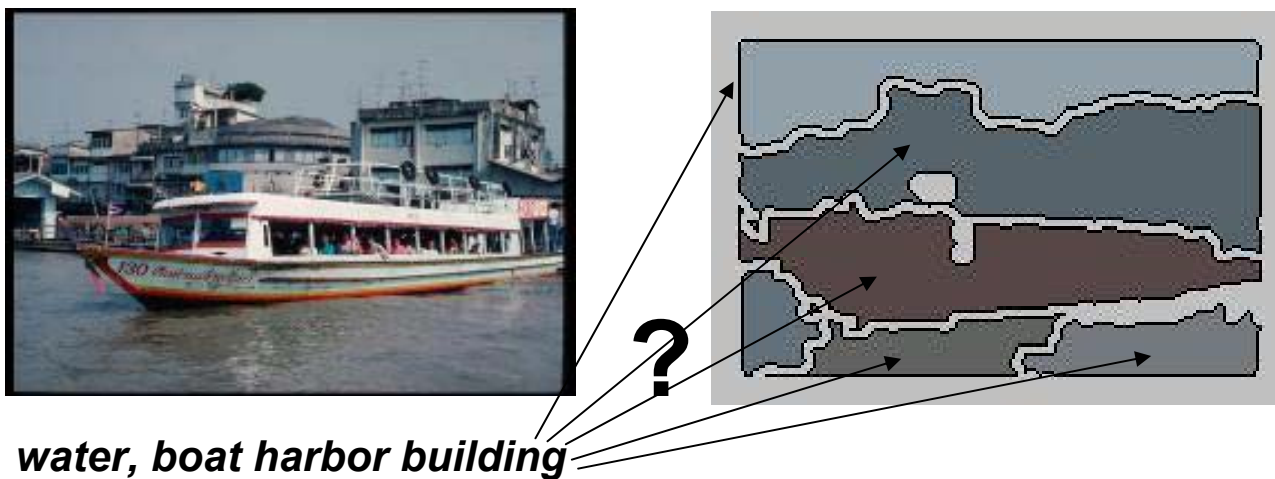
- Les bases d'images réelles (par exemple les images du web) ne sont pas étiquetées par région d'images.
- Comment apprendre les liens sémantiques entre mot-clé et région d'images en connaissant seulement les mots-clés par image ?



Par contre les régions autour de buildings seront très changeantes...

Difficulté

- Les bases d'images réelles (par exemple les images du web) ne sont pas étiquetées par région d'images.
- Comment apprendre les liens sémantiques entre mot-clé et région d'images en connaissant seulement les mots-clés par image ?



Par contre les régions autour de buildings seront très changeantes...

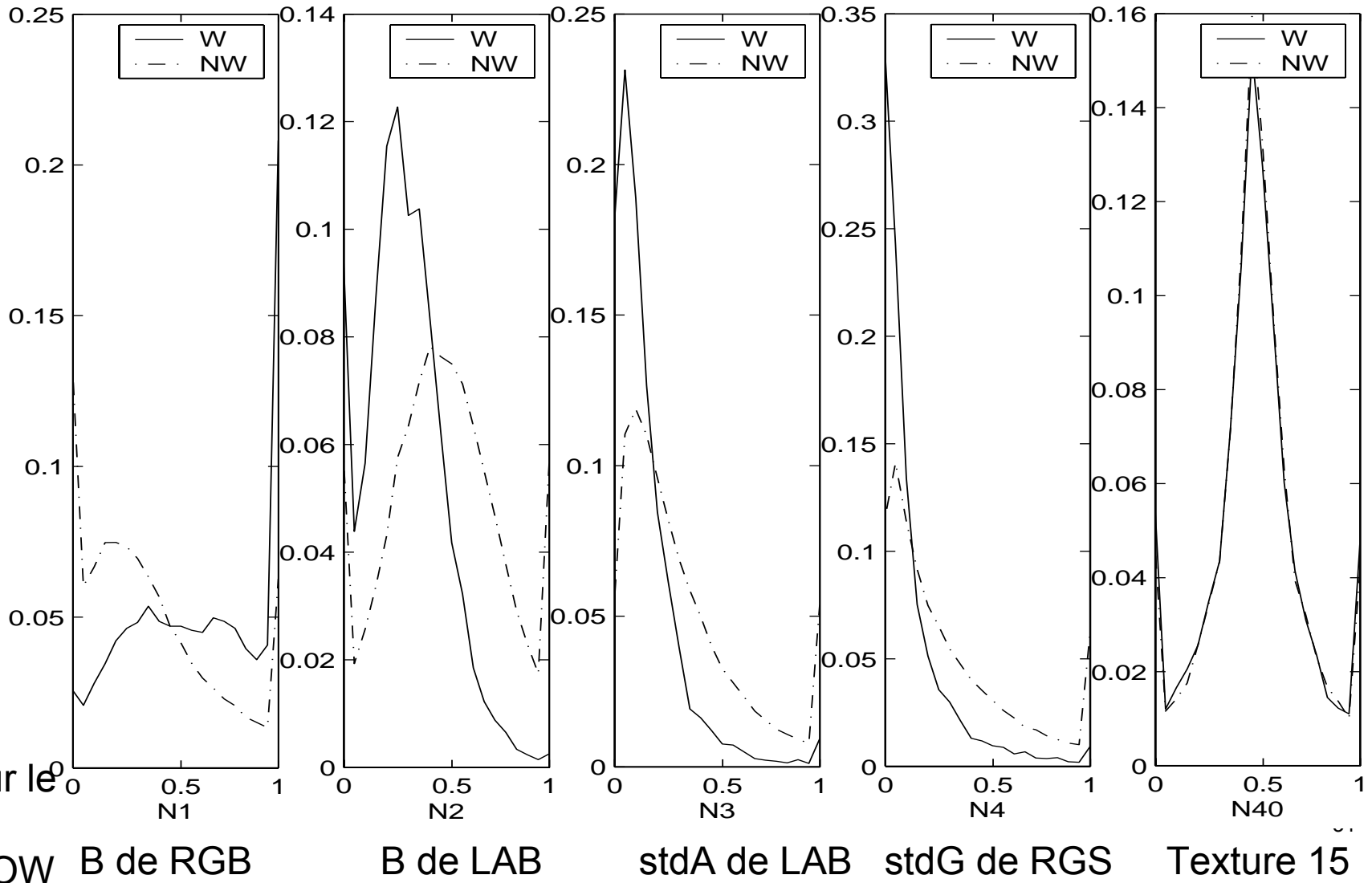
- **Proposition :**
 - utiliser une **Analyse Factorielle Discriminante (AFD)**

Utilisation de l'Analyse Factorielle Discriminante (AFD) pour déterminer les traits visuels les plus discriminants

- Pour déterminer les traits visuels les plus discriminants en fonction du mot clé
 - Séparation des images (et pas des régions) en 2 classes :
 - Celles qui sont indexées par ce mot
 - Celles qui ne sont pas indexées par ce mot
 - Calcul des variances interclasses (B) et intraclasses (W)
 - Pour chaque trait visuel v_j et pour chaque mot w_i , on calcule le pouvoir discriminant F:

$$F(v_j; w_i) = \frac{B(v_j; w_i)}{B(v_j; w_i) + W(v_j; w_i)}$$

Utilisation de l'Analyse Factorielle Discriminante (AFD) pour déterminer les traits visuels les plus discriminants



Résultats du choix pour chaque mot des traits visuels les plus discriminants par AFD

	40DIM	5SAMEBEST	10SAMEBEST	5BEST	10BEST
Nombre de dimensions	40	5	10	5	10
NS Moyen	0.22	0.08	0.10	0.28	0.27
NS Minimum	0.04	0.00	-0.01	0.02	0.04
NS Maximum	0.54	0.25	0.35	0.51	0.70

Pour tous les mots, les mêmes traits visuels les plus discriminants de toute la base.

Pour chaque mot, les traits visuels les plus discriminants.

Utilisation de l'Analyse Factorielle Discriminante (AFD) pour déterminer les traits visuels les plus discriminants

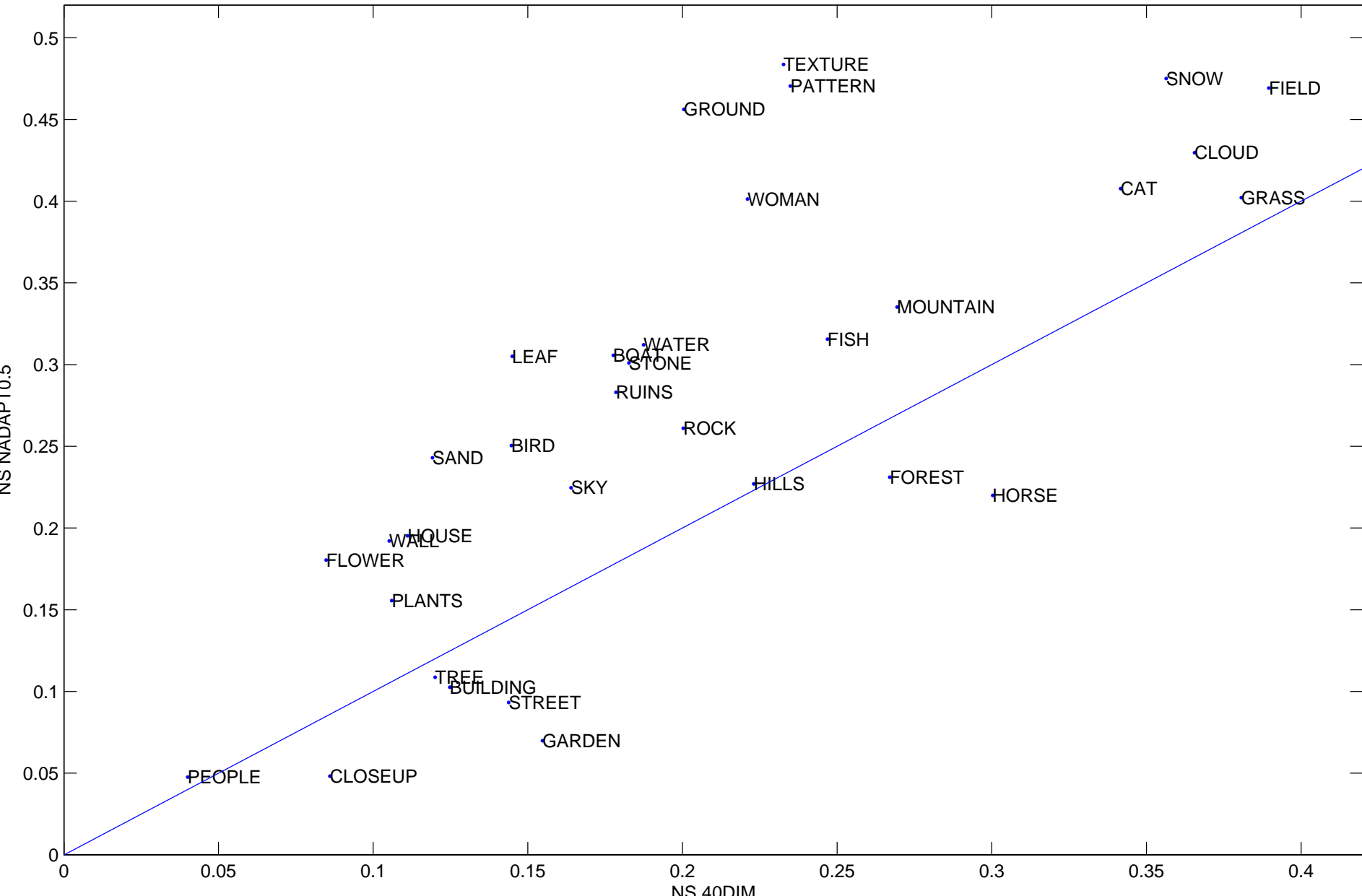
- Pour déterminer le nombre **N** de traits visuels nécessaires pour bien discriminer, nous proposons de prendre les N traits visuels qui cumulent 50% de la somme des pouvoirs discriminant de tous les traits.
 - Si l'on suppose que les traits visuels v_j sont ordonnés dans l'ordre décroissant des pouvoirs discriminant, nous choisissons N tel que :

$$\sum_{j=1}^N F(v_j; w_i) = \frac{1}{2} \sum_{j=1}^{40} F(v_j; w_i)$$

Résultats méthodes adaptatives

	40DIM	5BEST	NADAPT0.5
Nombre de dimensions	40	5	4.14 de 1 à 8 traits
NS Moyen	0.22	0.28	0.29
NS Minimum	0.04	0.02	0.04
NS Maximum	0.38	0.51	0.48
Gain moyen par rapport à 40DIM	—	+33%	+37%

Score NS 40DIM versus NS NADAPT0.5



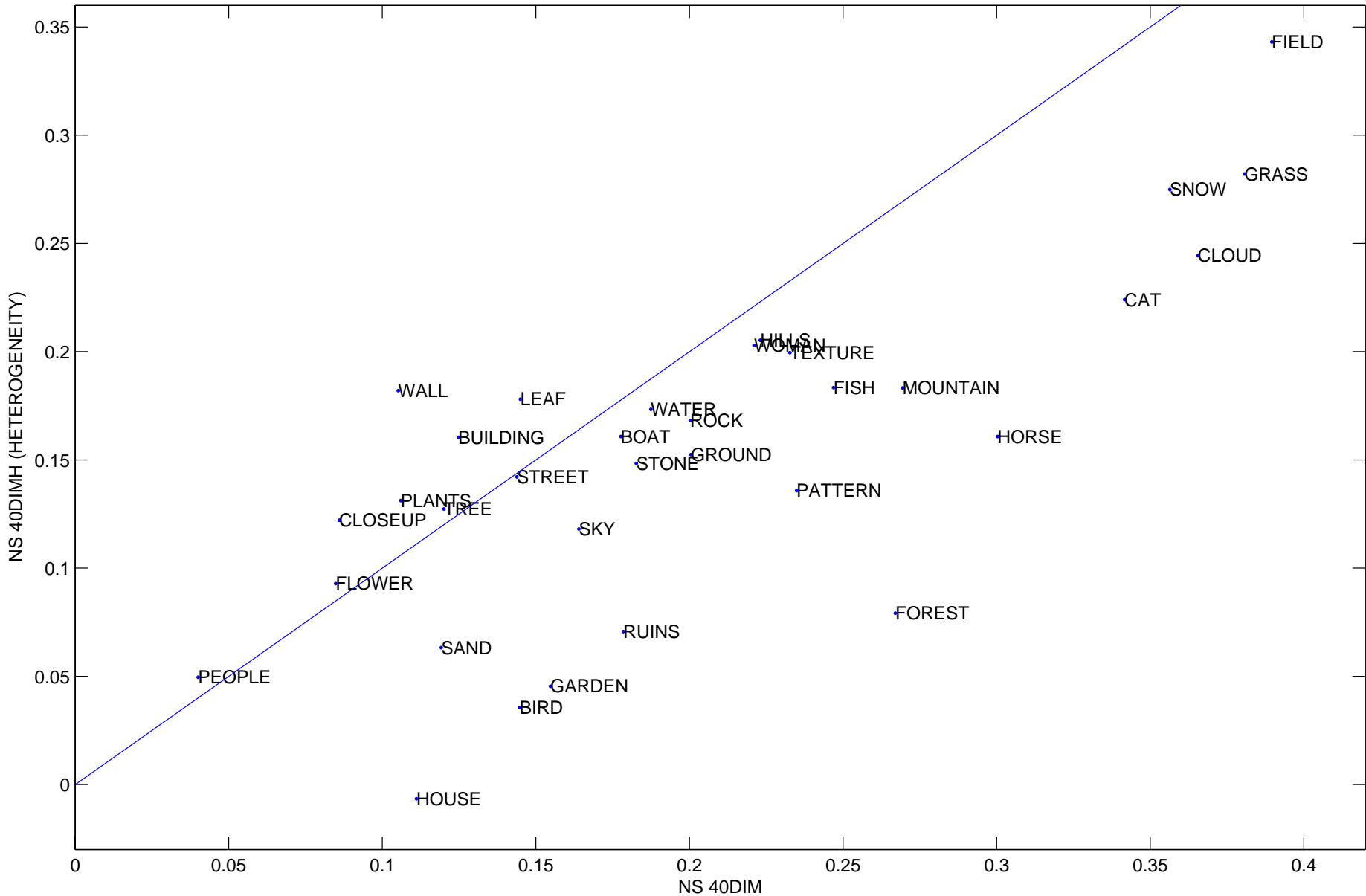
Construction de nouveaux traits visuels en utilisant des métadescripteurs

- Inspiré des travaux en psychovision
 - Le cerveau humain interprète en contexte
- La valeur de l'hétérogénéité pour le trait visuel p de l'image d est l'entropie :

$$H_p = - \sum_{b_j \in d} b_{j,p} \times \log_2(b_{j,p})$$

J. Martinet, « Un modèle vectoriel relationnel de recherche d'informations adapté aux images », Thèse de doctorat, Université Joseph Fourier, Grenoble, décembre 2004

Score NS 40DIM versus NS 40DIMH



Résultats méthodes adaptatives et hétérogénéité

	40DIM	5BEST	NADAPT0.5	40DIMH
Nombre de dimensions	40	5	4.14	40
NS Moyen	0.22	0.28	0.29	0.16
NS Minimum	0.04	0.02	0.04	-0.01
NS Maximum	0.38	0.51	0.48	0.34
Gain moyen par rapport à 40DIM	—	+33%	+37%	-17%

Conclusion et perspectives

- Nous avons montré que :
 - L'hypothèse d'application de l'AFD est validée de manière expérimentale.
 - Réduire l'espace visuel aux dimensions les plus discriminantes permet une amélioration de la recherche d'images visuo-textuelle de +37% tout en réduisant le nombre de dimensions de 90%.
 - La dérivation de nouveaux traits visuels à l'aide de l'hétérogénéité permet d'extraire une information complémentaire.
- Perspectives :
 - Utiliser la méthode de sélection des traits visuels en incluant l'hétérogénéité.
 - Fusionner les deux types de traits.
 - A long terme, développer un système de recherche d'images capable d'exploiter efficacement à la fois les informations textuelles et visuelles.



Merci de votre attention

Questions ?