
*Sélection des dimensions visuelles
par LDA ou MMD
sur des bases d'images mal-annotées*

Sabrina Tollari

UMR CNRS 6168 LSIS

Université du Sud Toulon-Var

<http://tollari.univ-tln.fr>

Séminaire Eurécom le 20 septembre 2006 à Sophia-Antipolis

Plan

Motivations :

Pourquoi a-t-on besoin d'utiliser des techniques de sélection de dimensions visuelles sur des images mal annotées ?

Rappel sur la LDA

Rappel sur la MMD

Expérimentations :

Corpus

Un système de classification des images

Comparaisons expérimentales des méthodes LDA et MMD

Conclusion

Motivations

L'augmentation du nombre de données multimédia telles que les bases d'images généralistes (le Web, les encyclopédies, les médias...) posent principalement deux problèmes :

- Le problème du fossé sémantique

- Le problème du passage à l'échelle

Motivations : Systèmes de recherche d'images sur le Web



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

Motivations :

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

Motivations :

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

Motivations :

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

Motivations :

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

- Comment vérifier simplement la pertinence d'un mot par rapport au contenu visuel de l'image ?

Problème : l'image est annotée par image entière et non par région d'images.

Problématique

Idée :

apprendre sur une base d'apprentissage les dimensions de l'espace visuel qui caractérisent le mieux un mot-clé, puis classer les images à partir uniquement de ces dimensions visuelles pour vérifier leur appartenance à la classe du mot.

Avantages :

Élimination des dimensions visuelles non discriminantes pour un mot donné (bruits)

Contournement du problème des grandes dimensions

Augmentation de la rapidité du système

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

Sélection des traits visuels

Problématique :

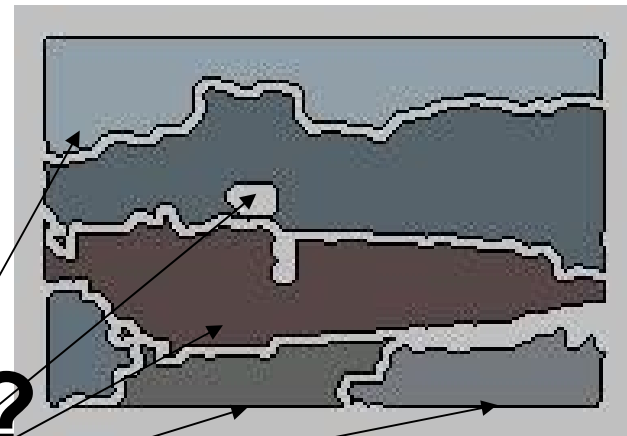
Quels traits visuels doit-on utiliser pour bien discriminer un certain concept (mot) étant donné :

- le problème des grandes dimensions et
- l'existence de traits visuels non-informatifs ?

Proposition

Utiliser la méthode classique de l'Analyse Linéaire Discriminante (LDA) ?

Difficulté : Les bases d'images réelles (ex : Web, COREL) ne sont pas étiquetées par région d'images



water boat harbor building

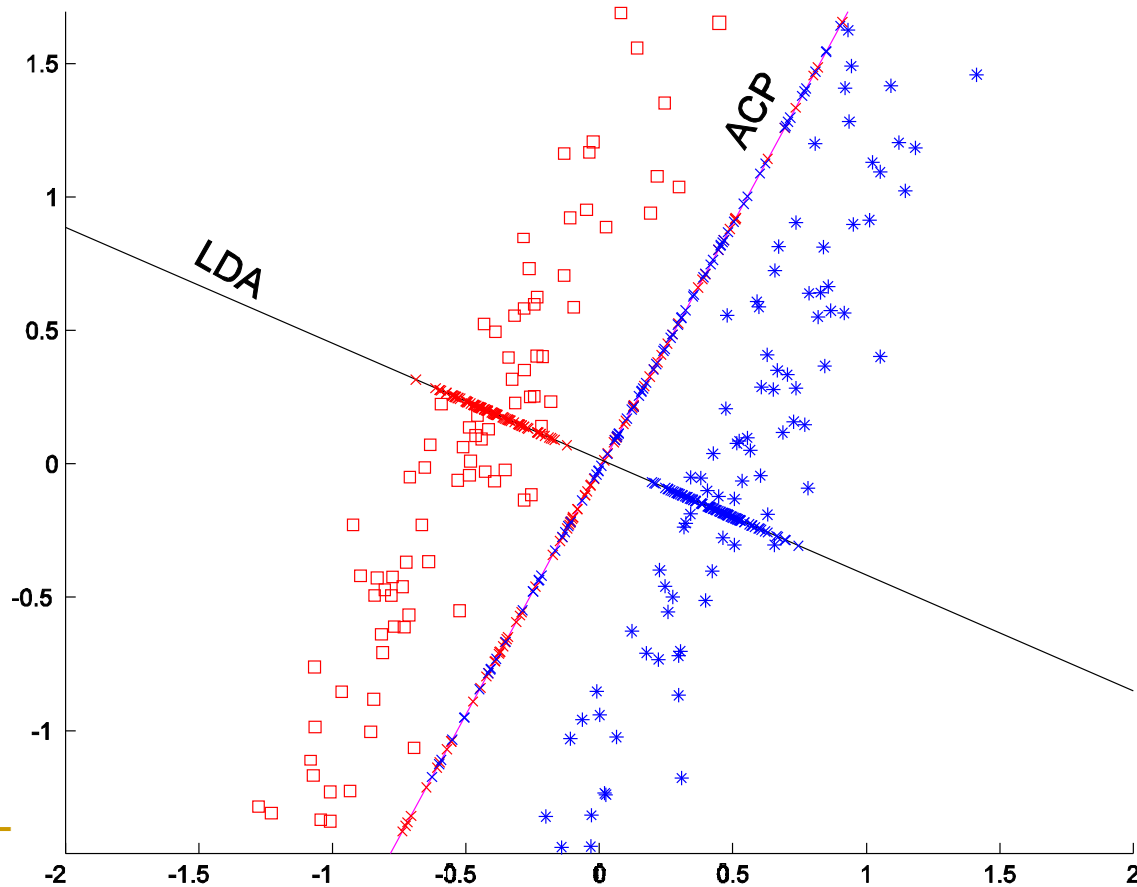
Par contre, les régions autour du « building » seront très changeantes...

=> Approximer la LDA (ALDA)

Rappel : LDA versus ACP

L'ACP recherche l'axe qui représente le mieux les données

La LDA recherche l'axe qui sépare le mieux les classes



Approximation des variances inter-/intra- classes pour estimer le critère de Fisher (LDA)

Pour déterminer les traits visuels les plus discriminants en fonction du mot w_i considéré à partir de la base d'apprentissage

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:

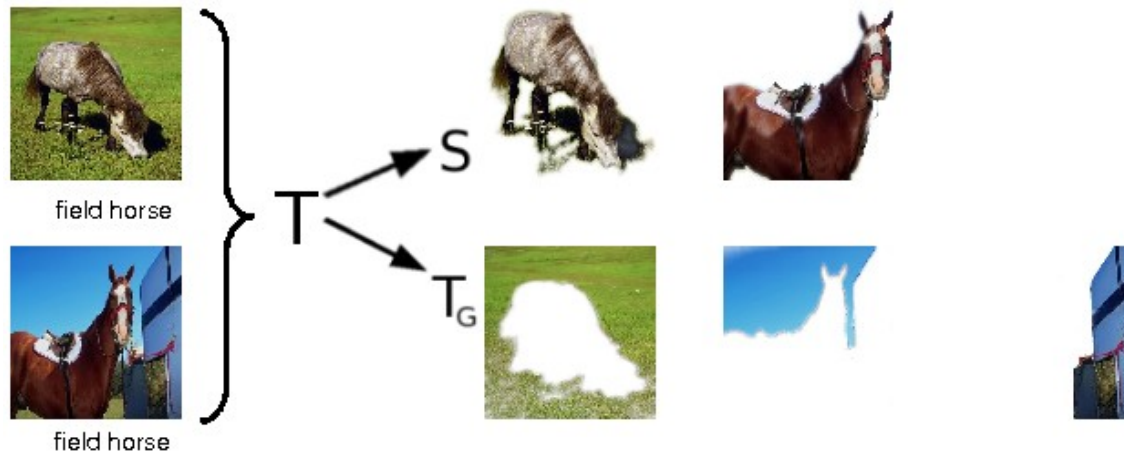
$$F(v_j; w_i) = \frac{B(v_j; w_i)}{B(v_j; w_i) + W(v_j; w_i)}$$

Nous montrons théoriquement que **l'ordre des F estimés par ALDA est le même que par LDA.**

Principe de démonstration de l'ALDA

Proposition : Le critère de Fisher sur les ensembles T et G peut être exprimé linéairement en fonction du critère de Fisher sur les ensembles S et G

Images
annotées
par le mot
« horse »



Images
non
annotées
par le mot
« horse »



Le critère de la Diversité Marginal Maximal (MMD)

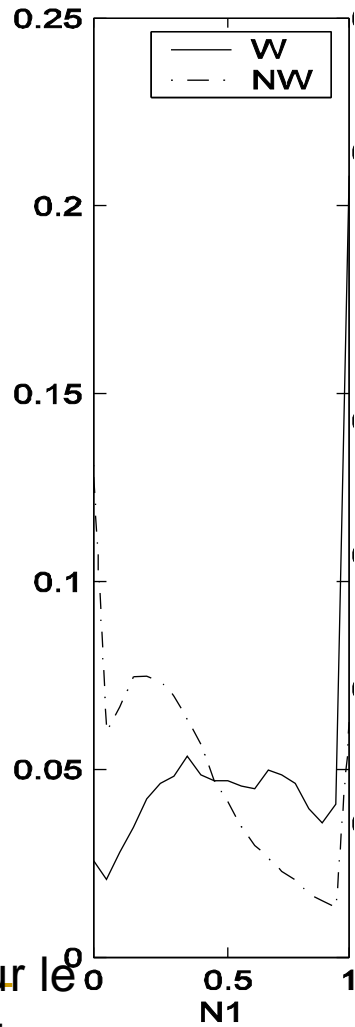
Le critère MMD consiste à choisir l'axe v_u qui maximise la diversité marginale.

La diversité maximale (MD) est la moyenne des divergences de Kullback-Leibler entre les probabilités conditionnelles des classes et leur moyenne :

$$\hat{J}_{MD}(v_u; w_j) = p(v_u | C_{w_j}) \log \frac{p(v_u | C_{w_j})}{p(v_u)} + p(v_u | C_{\neg w_j}) \log \frac{p(v_u | C_{\neg w_j})}{p(v_u)}.$$

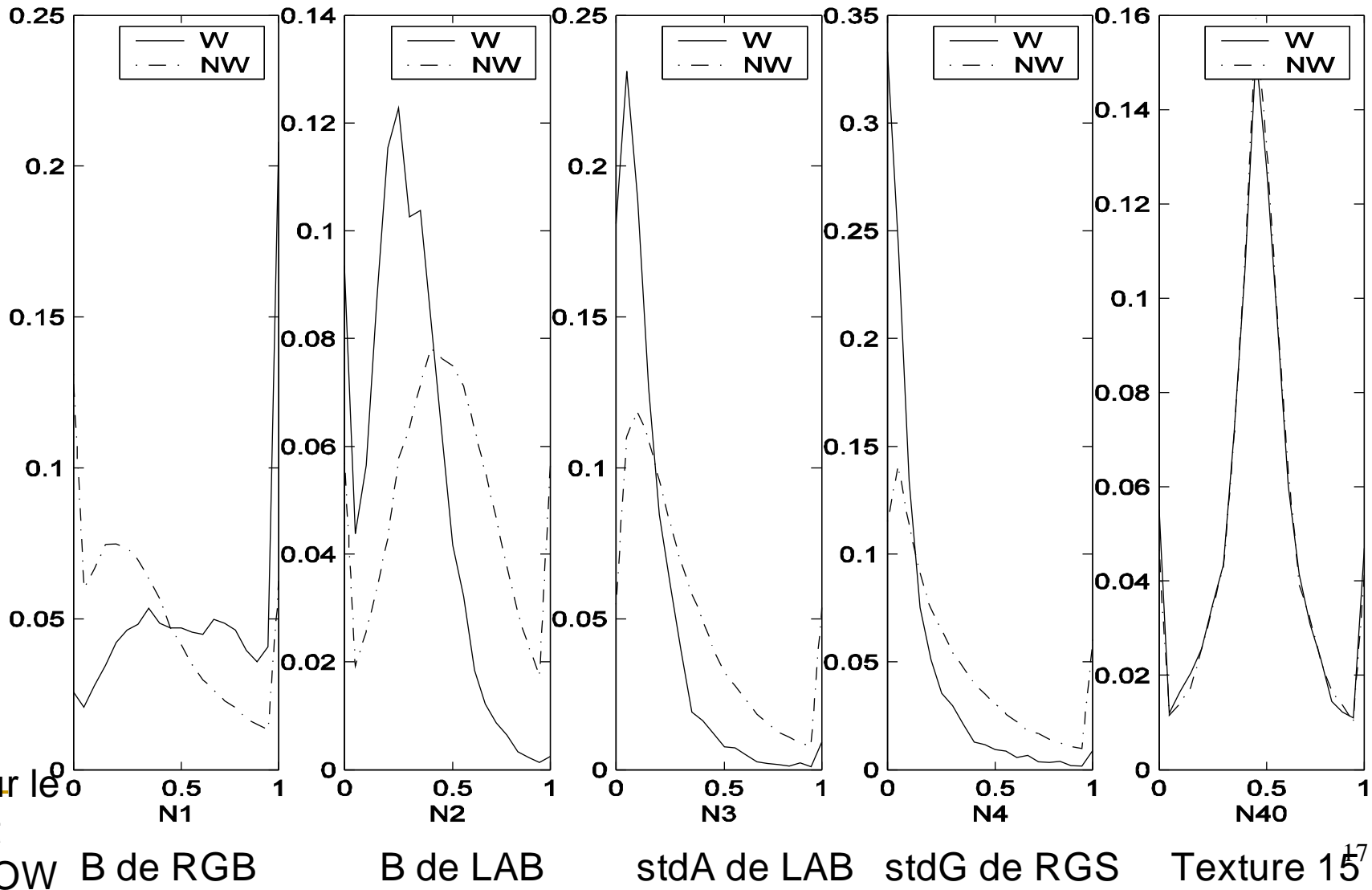
Exemple :

critère de Fisher versus critère MMD



Pour le mot
SNOW B de RGB

Exemple : critère de Fisher versus critère MMD



Détermination du nombre de traits visuels à prendre en compte pour bien discriminer un mot

Pour déterminer le nombre N de traits visuels nécessaires pour bien discriminer un mot w_i , nous proposons de prendre les N dimensions visuelles qui cumulent $\tau\%$ de la somme des pouvoirs discriminant de tous les traits.

Soient les traits visuels v_j ordonnés dans l'ordre décroissant des pouvoirs discriminants, nous choisissons N tel que :

$$\sum_{j=1}^N F(v_j; w_i) = \tau \sum_{j=1}^{\delta} F(v_j; w_i)$$

Expérimentations

Corpus

Bases d'images COREL

10 000 images :

5000 images d'apprentissage, 2500 images pour l'ensemble de développement et 2500 images pour l'ensemble de test

250 mot-clés environs en anglais,
réduit aux 52 mots les plus fréquents

Chaque image possède :

De 1 à 5 mot-clés choisis manuellement

De 2 à 10 régions « blobs » de l'image

Chaque blob est décrit par un vecteur de caractéristiques visuelles

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

Description des vecteurs visuels usuels (U)

Chaque blob de l'image possède un vecteur visuel de 40 dimensions extrait par Kobus Barnard :

- 6 dimensions de formes (aire du blob...)

- 18 dimensions de couleurs (RGB, RGS, LAB),

- 16 dimensions de textures (filtres gaussiens...).

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.

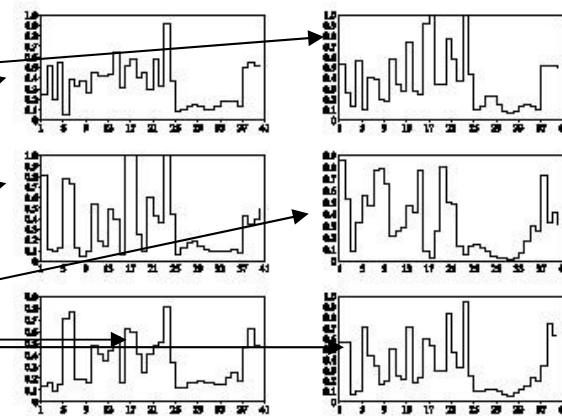
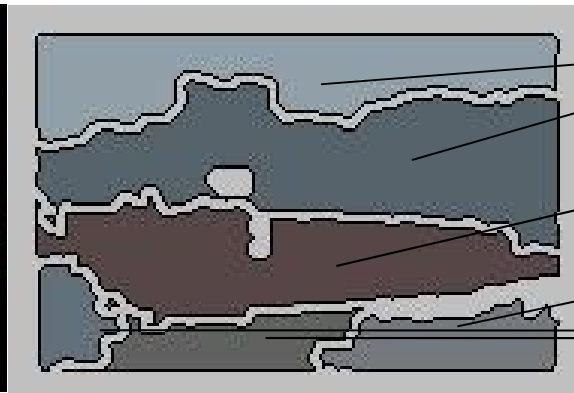
Kobus Barnard, P. Duygulu, N. de Freitas, D. Forsyth, D. Blei, and M. I. Jordan, « *Matching Words and Pictures* », Journal of Machine Learning Research, Vol 3, pp 1107-1135, 2003.

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

Corpus

Construction de nouveaux traits visuels : hétérogénéité (H) de l'image étendue à tous les traits

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

Construction de clusters visuels pour chaque mot

Objectifs :

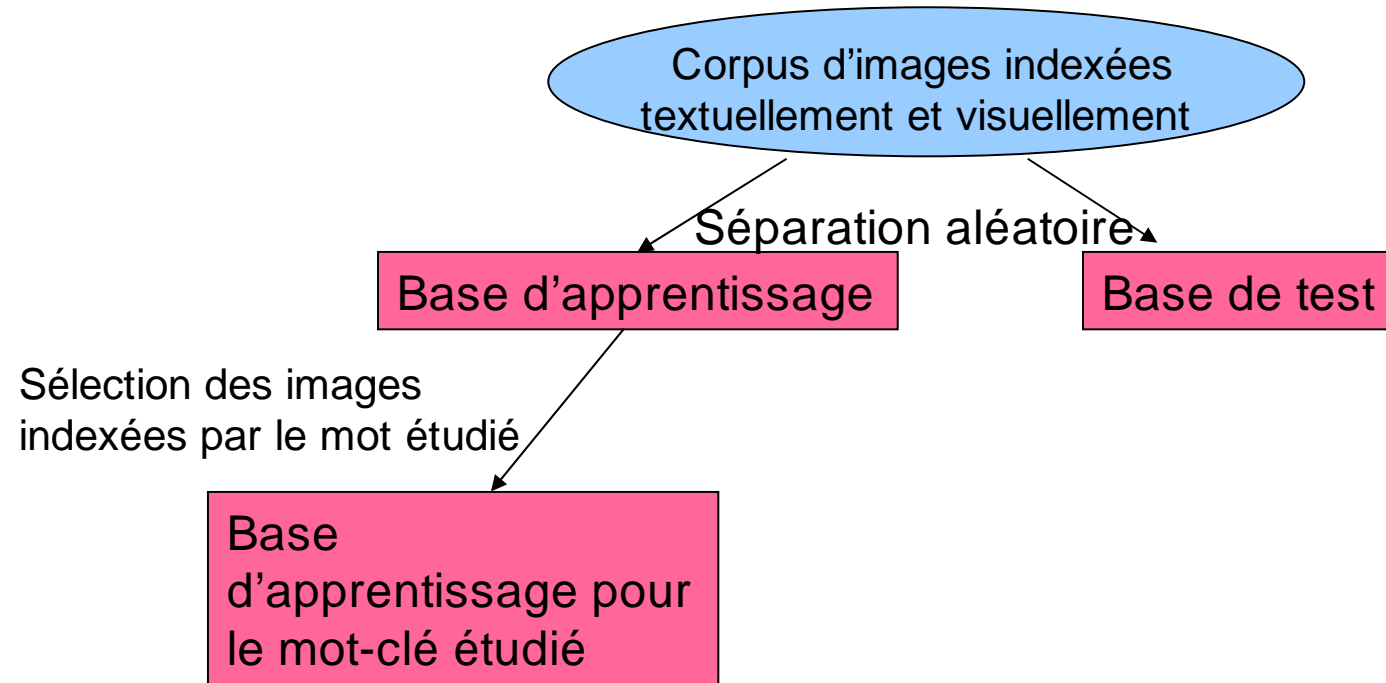
- Apprendre les zones de l'espace visuel représentant le mieux chaque mot
- Apprendre les dépendances entre les traits visuels

Construction de clusters visuels



Corpus d'images indexées
textuellement et visuellement

Construction de clusters visuels



Construction de clusters visuels pour chaque mot

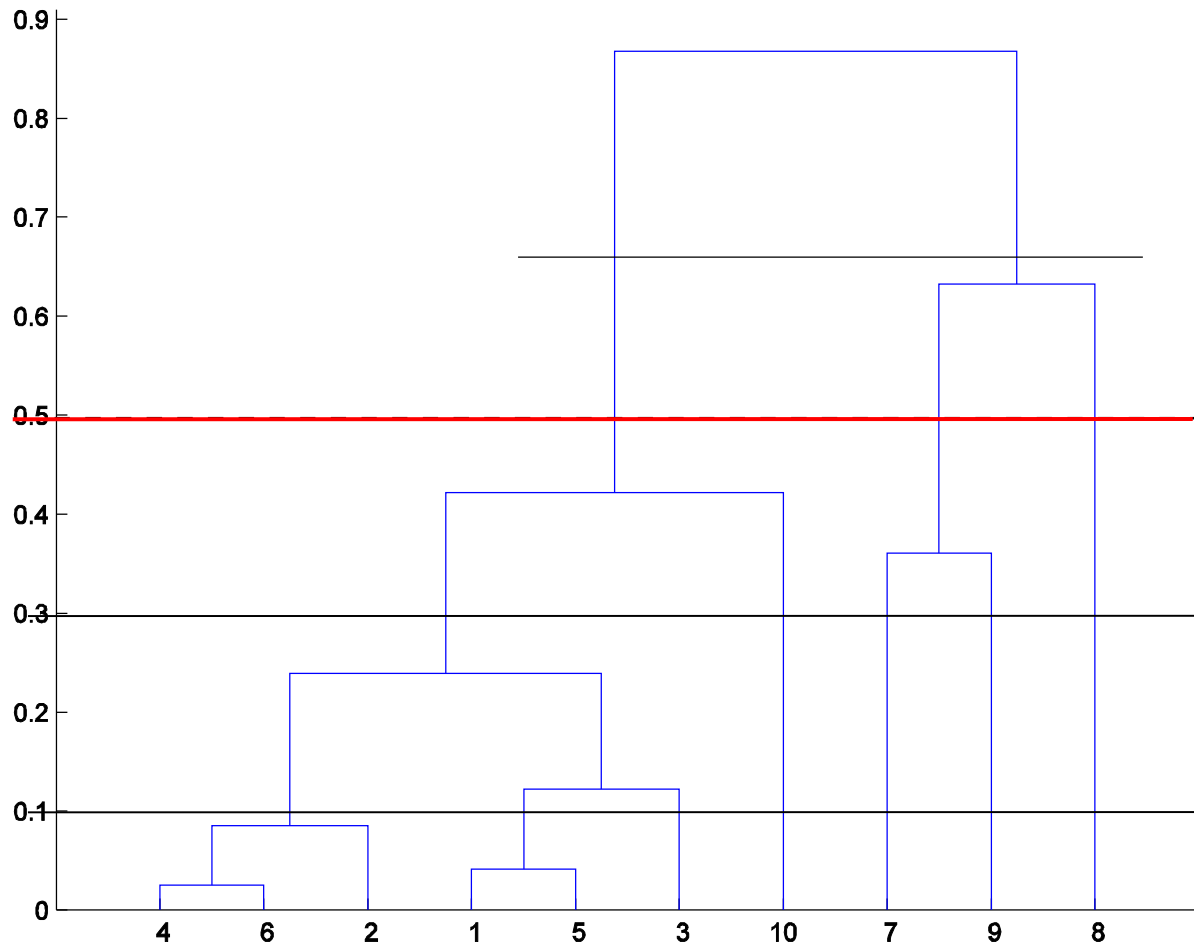
Par apprentissage non-supervisé : classification ascendante hiérarchique (CAH)

Pour chaque mot w

- 1) Construire un sous-ensemble de la base d'apprentissage avec les images possédant le mot w
- 2) Initialiser pour chaque blob un cluster visuel
- 3) Regrouper les clusters visuels les plus proches
- 4) Évaluer les clusters obtenus sur l'ensemble de développement
- 5) Recommencer les étapes 3) et 4) jusqu'au regroupement de toutes les clusters visuels
- 6) Garder les clusters qui maximisent le score sur l'ensemble de développement
- 7) Évaluer les clusters obtenus sur l'ensemble de test

Un cluster visuel est un hyper-rectangle dans l'espace visuel.

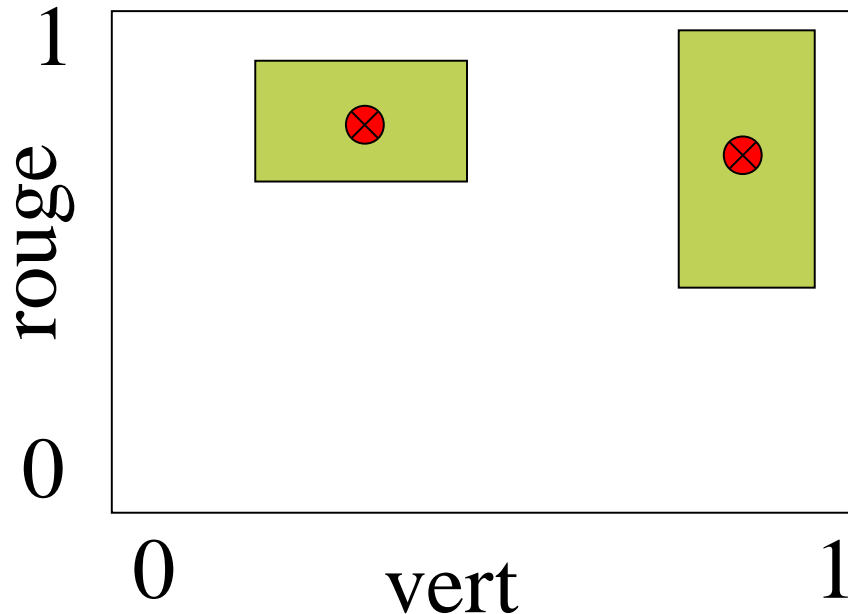
Construction de clusters visuels pour chaque mot : exemple



Blobs des images d'apprentissage du mot « Soleil »

Construction de clusters visuels pour chaque mot : exemple

« Soleil »

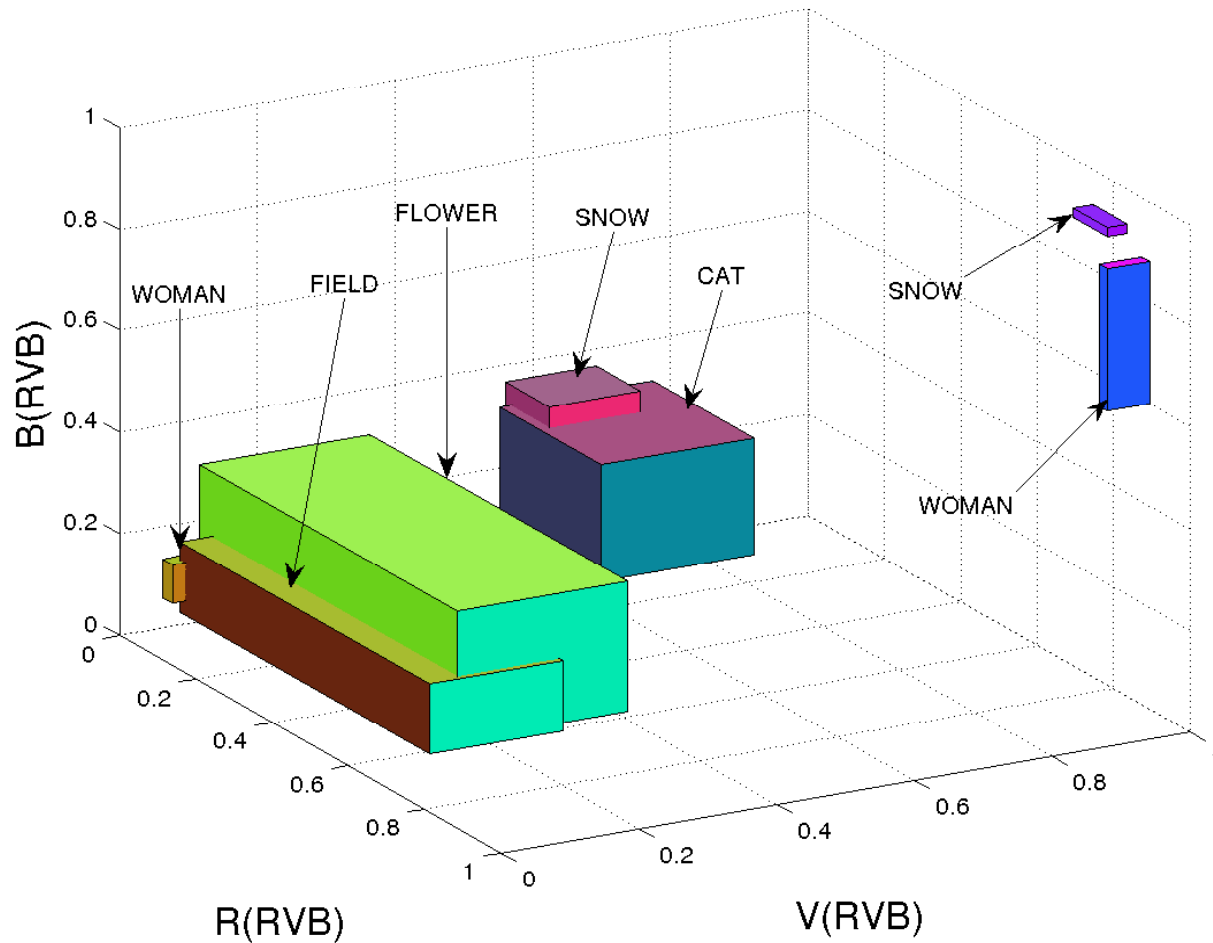


Pour chaque cluster de chaque mot, on garde seulement :

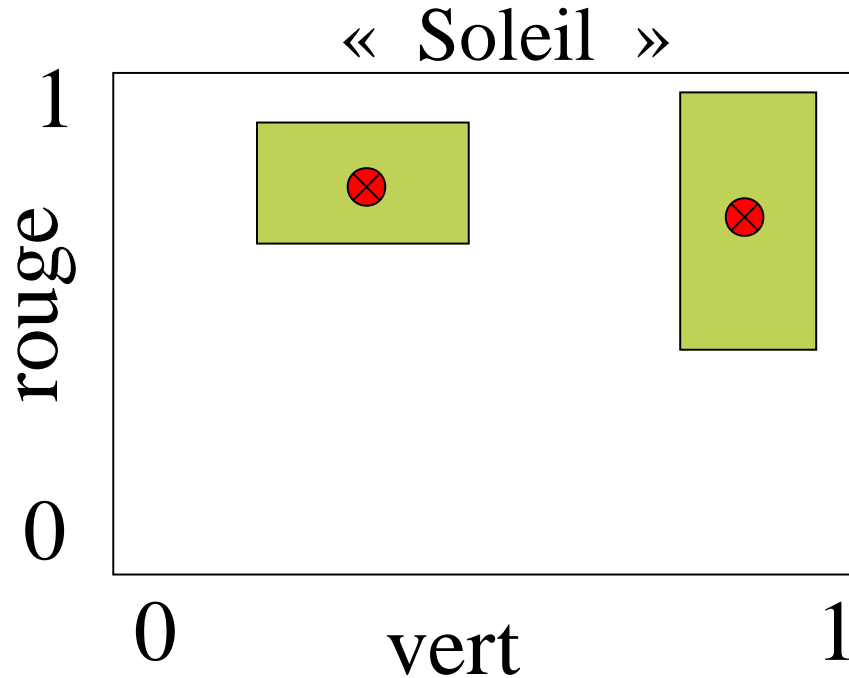
Le vecteur centroïde

Les valeurs des écarts-types du cluster pour chaque dimension de l'espace

Exemple de clusters visuels de plusieurs mots dans l'espace visuel



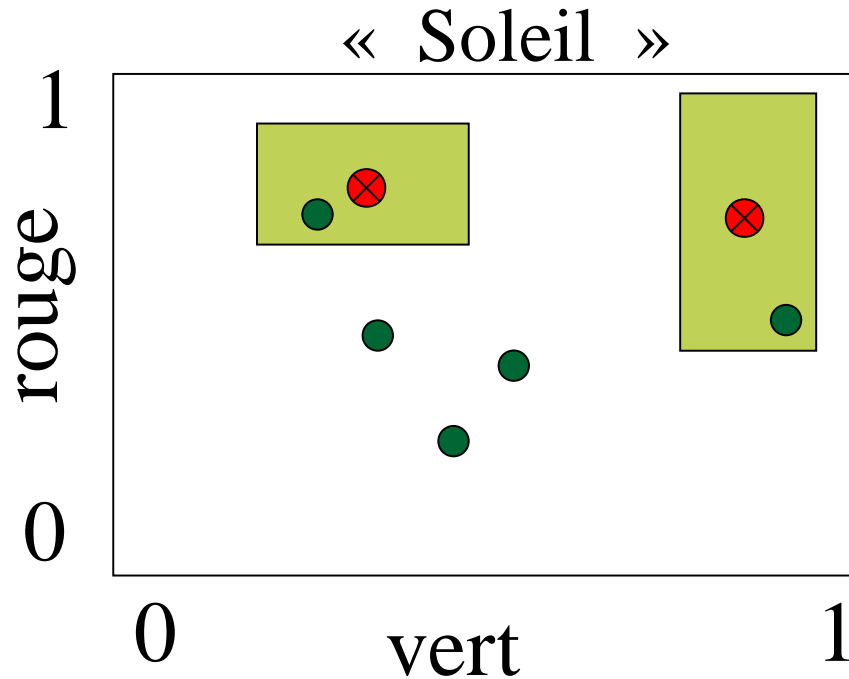
Construction de clusters visuels pour chaque mot : exemple



Un mot est associé à un blob si ce blob est dans l'un des clusters visuels de ce mot.

Un mot est associé à une image si B blobs de cette image appartiennent aux clusters visuels de ce mot.

Construction de clusters visuels pour chaque mot : exemple



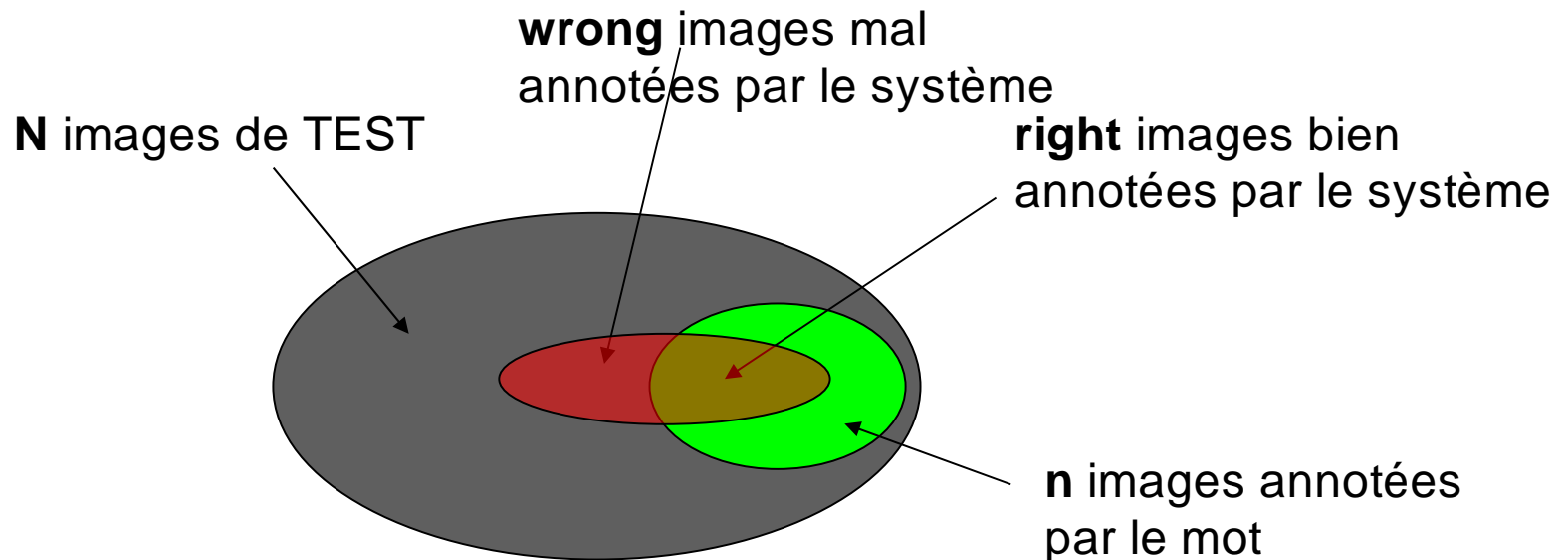
Un mot est associé à un blob si ce blob est dans l'un des clusters visuels de ce mot.

Un mot est associé à une image si B blobs de cette image appartiennent aux clusters visuels de ce mot.

Évaluation des clusters

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

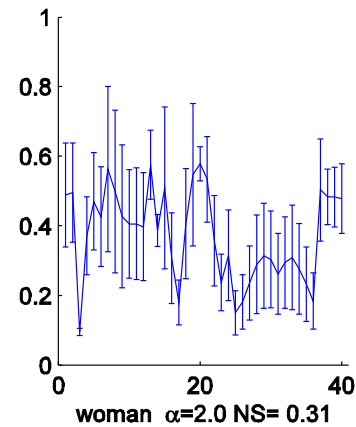
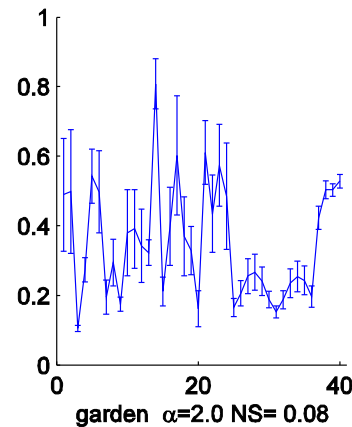
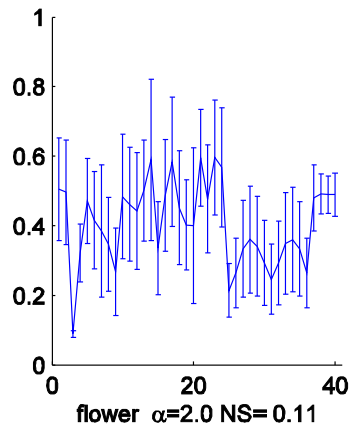
$$\text{NS} = \frac{\text{right}/n}{\text{sensibilité}} - \frac{\text{wrong}/(N-n)}{1-\text{spécificité}}$$



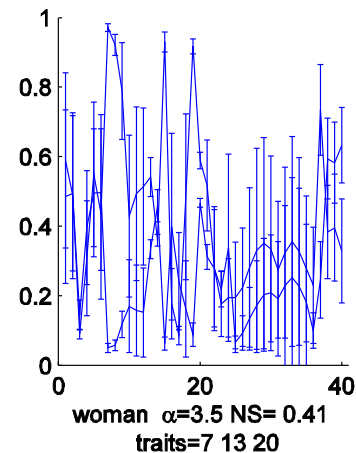
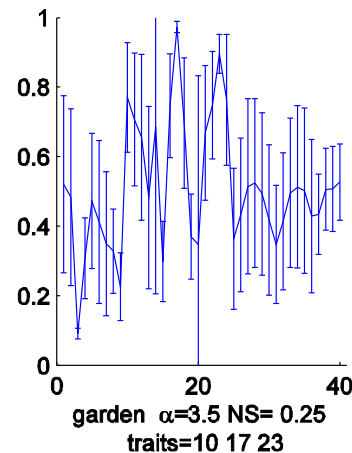
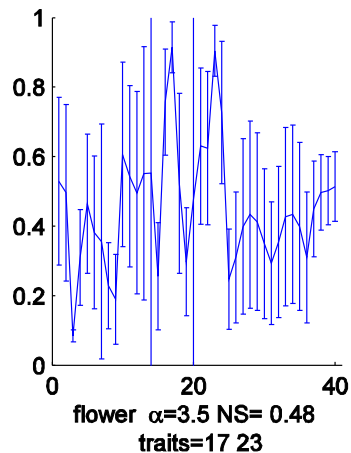
Sélection des traits visuels

Exemples de clusters visuels dans l'espace usuel à 40 dimensions avec ou sans sélection par ALDA

Sans
sélection

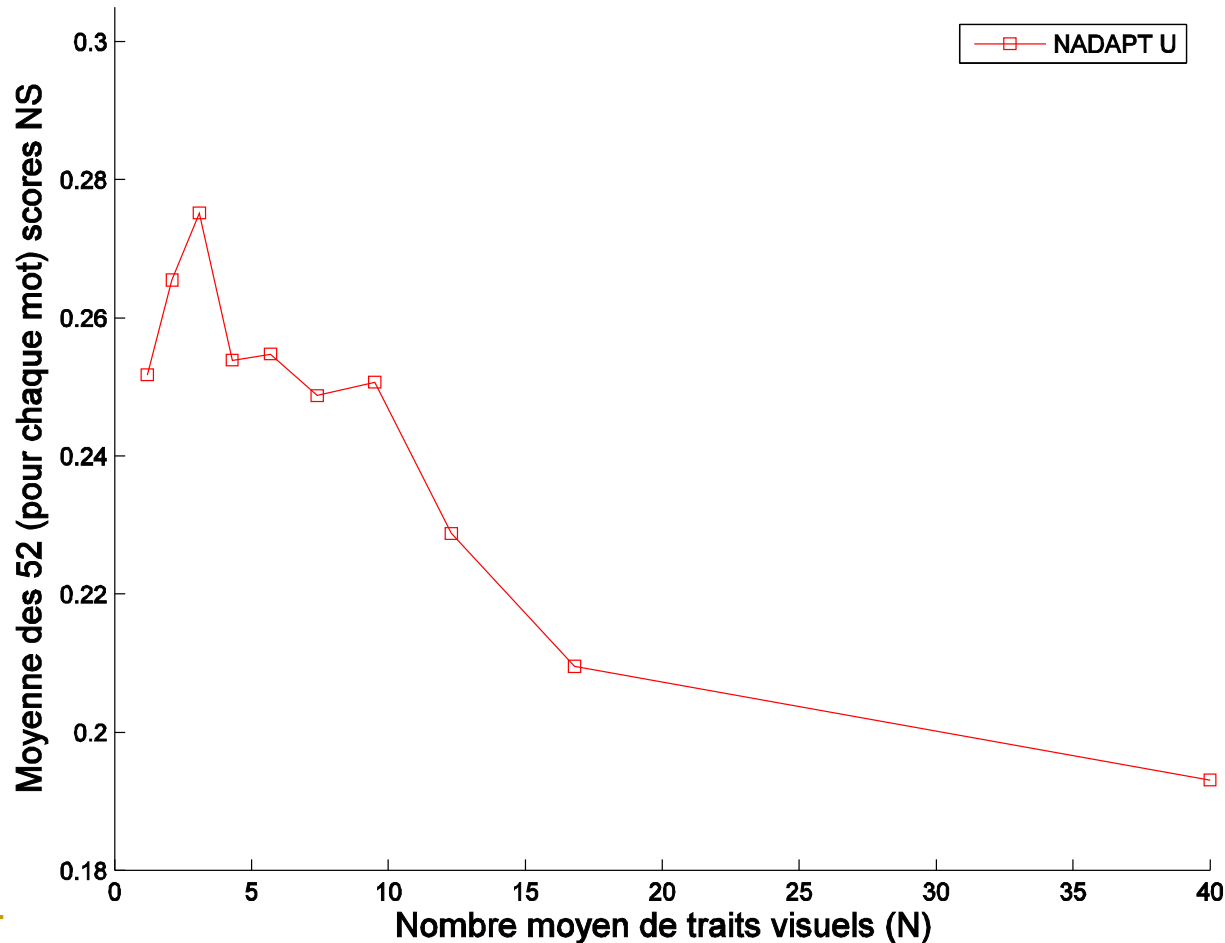


Avec
sélection



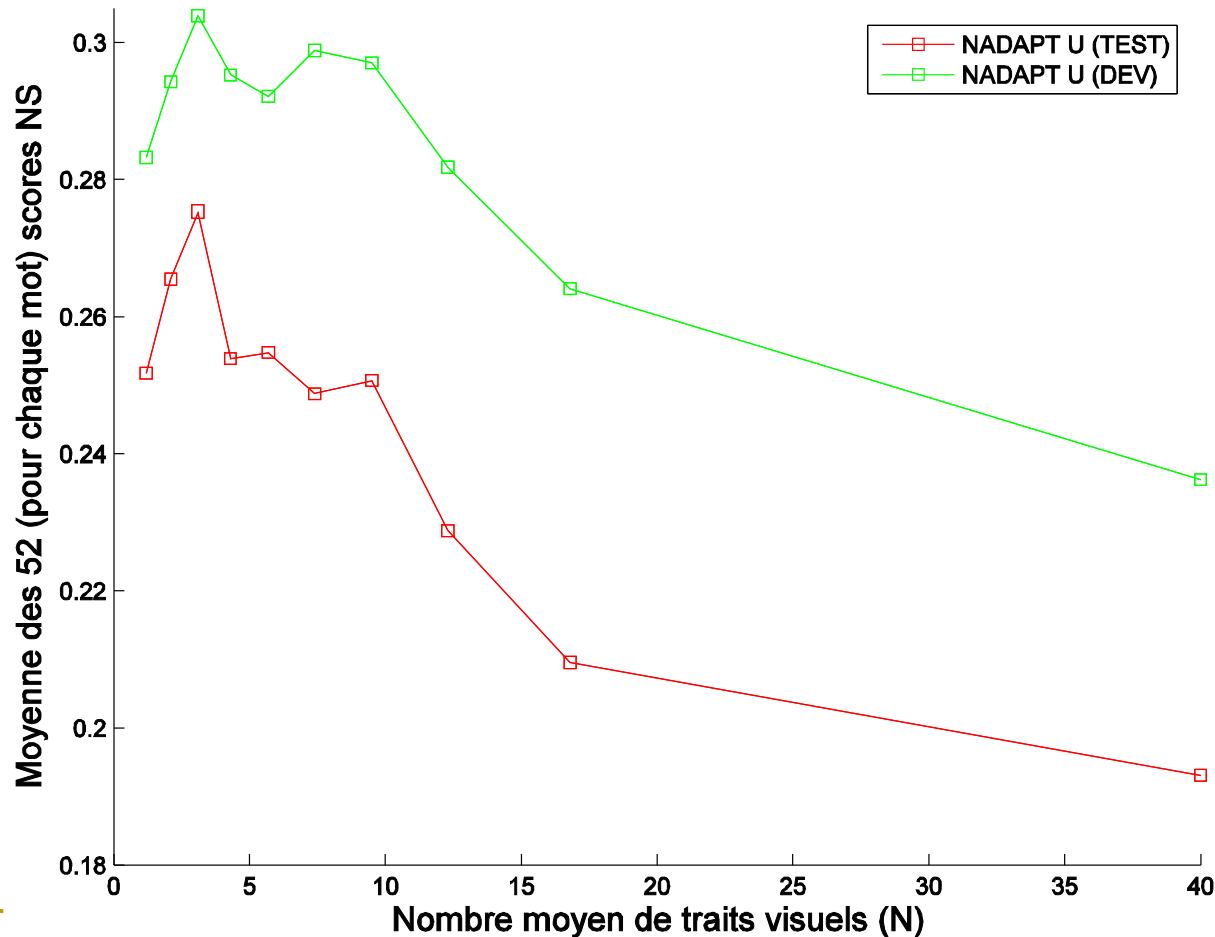
Scores moyens des classifications de 52 mots en fonction du nombre moyen de dimensions utilisées

Expériences avec les traits visuels usuels (U)



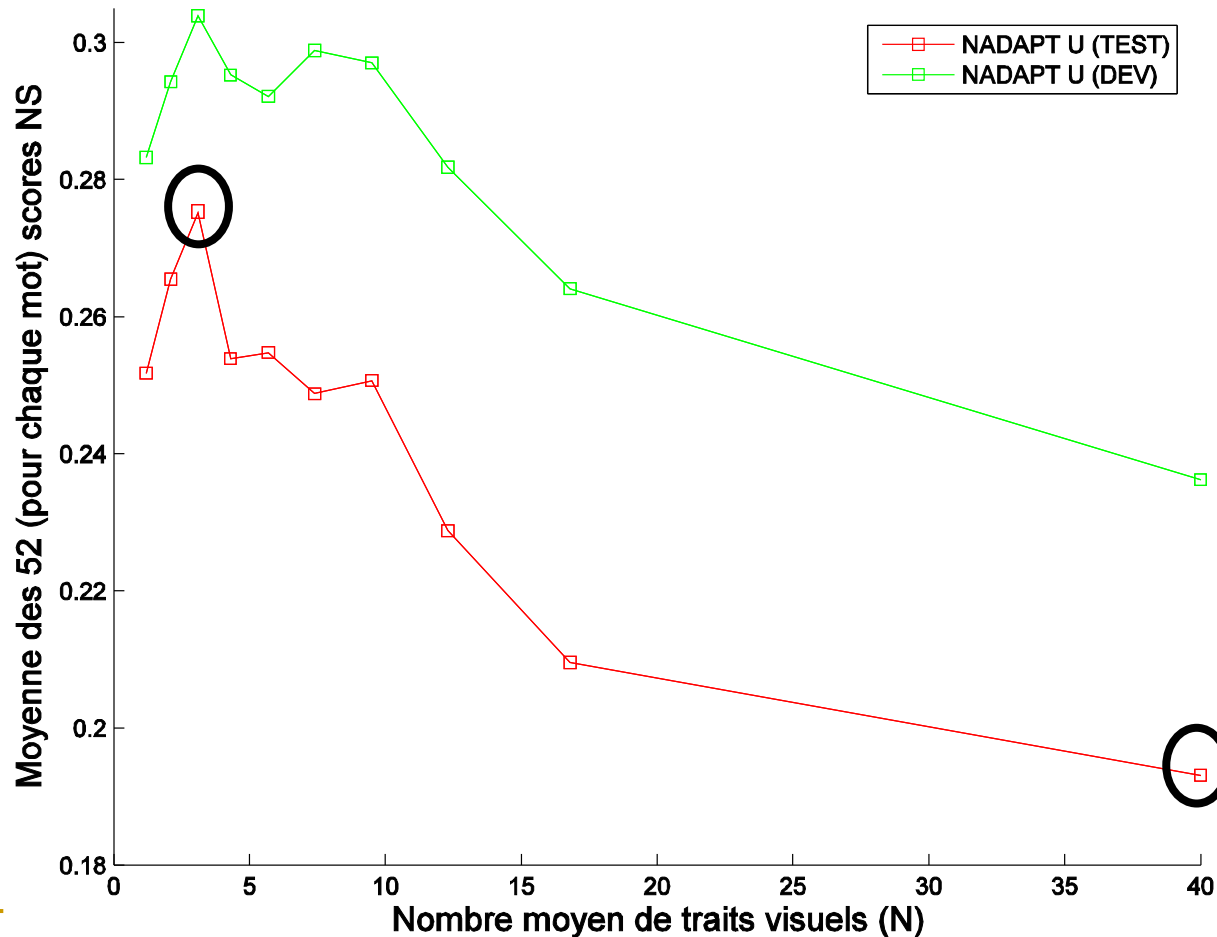
Scores moyens des classifications de 52 mots en fonction du nombre moyen de dimensions utilisées

Expériences avec les traits visuels usuels (U)

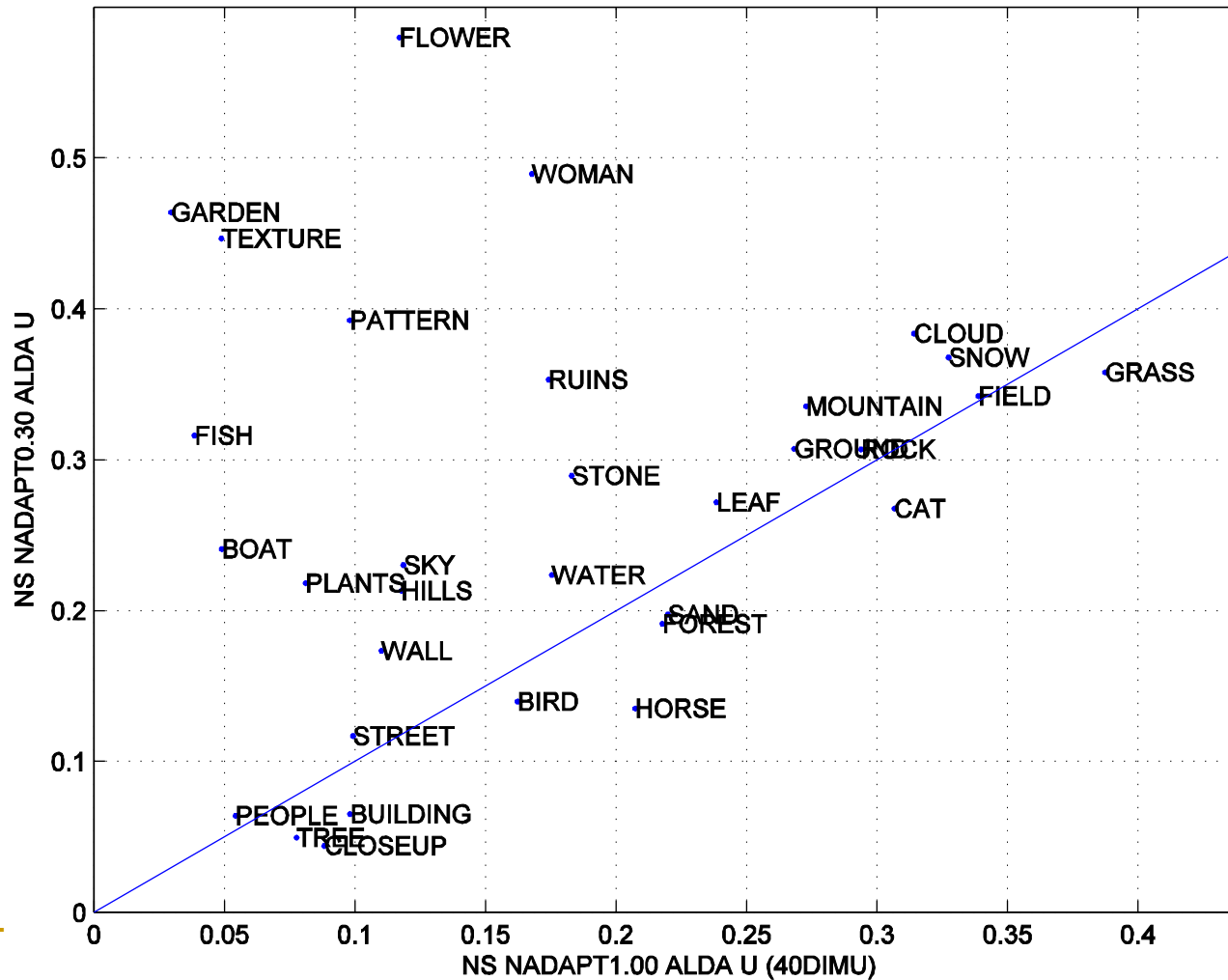


Scores moyens des classifications de 52 mots en fonction du nombre moyen de dimensions utilisées

Expériences avec les traits visuels usuels (U)

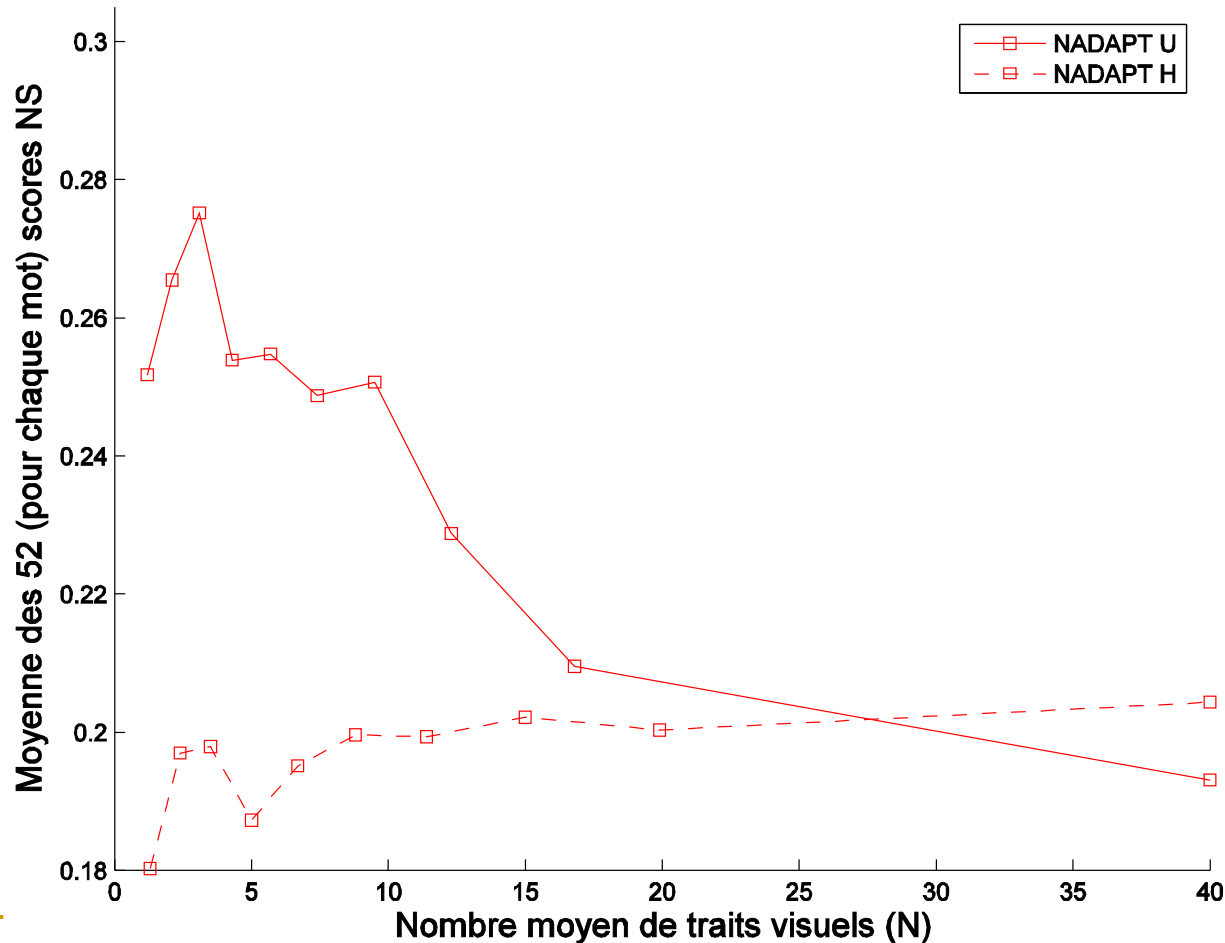


Scores de classification de chaque mot sans/avec sélection de traits visuels

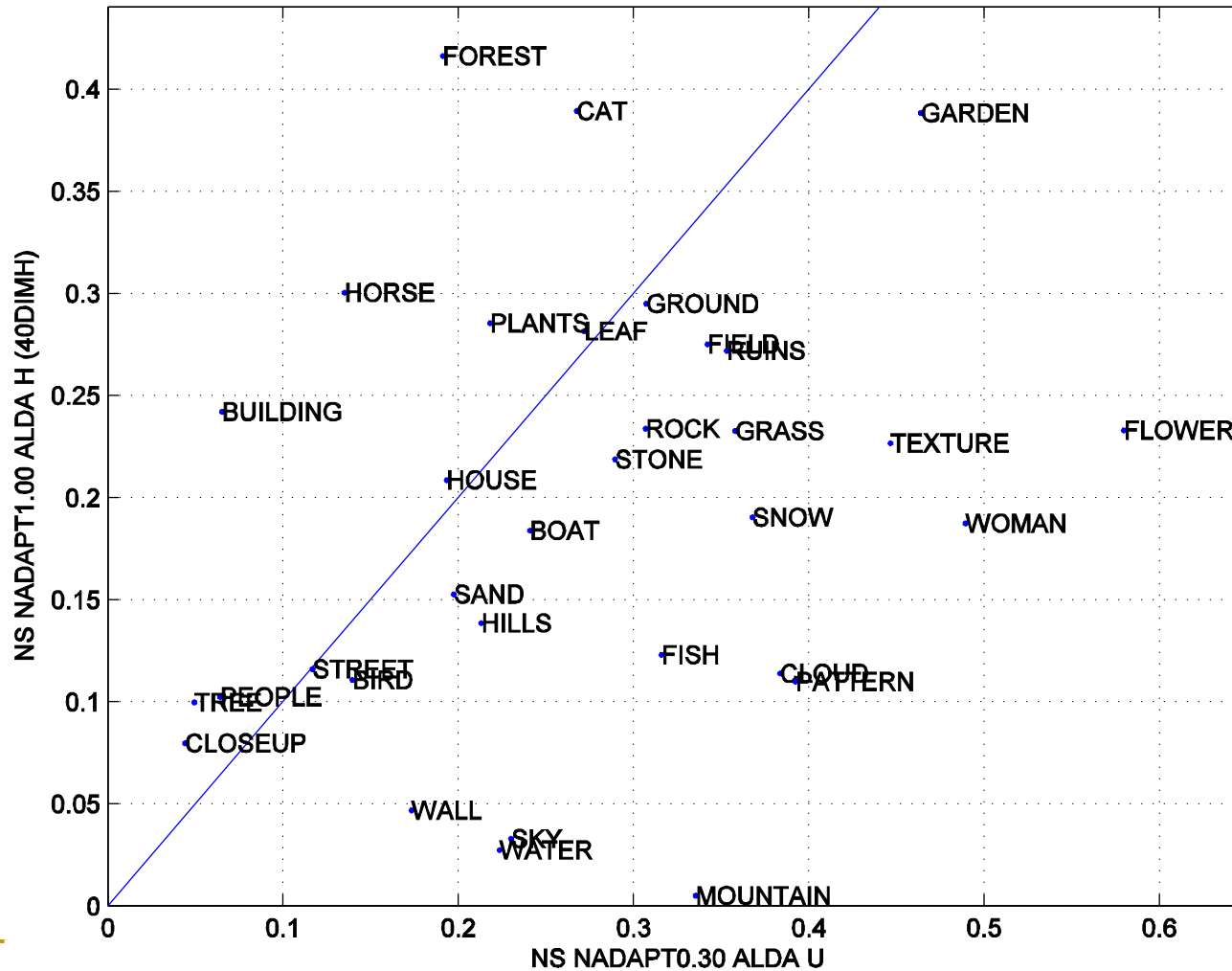


Scores moyens des classifications de 52 mots en fonction du nombre moyen de dimensions utilisées

Expériences avec les traits d'hétérogénéité (H)

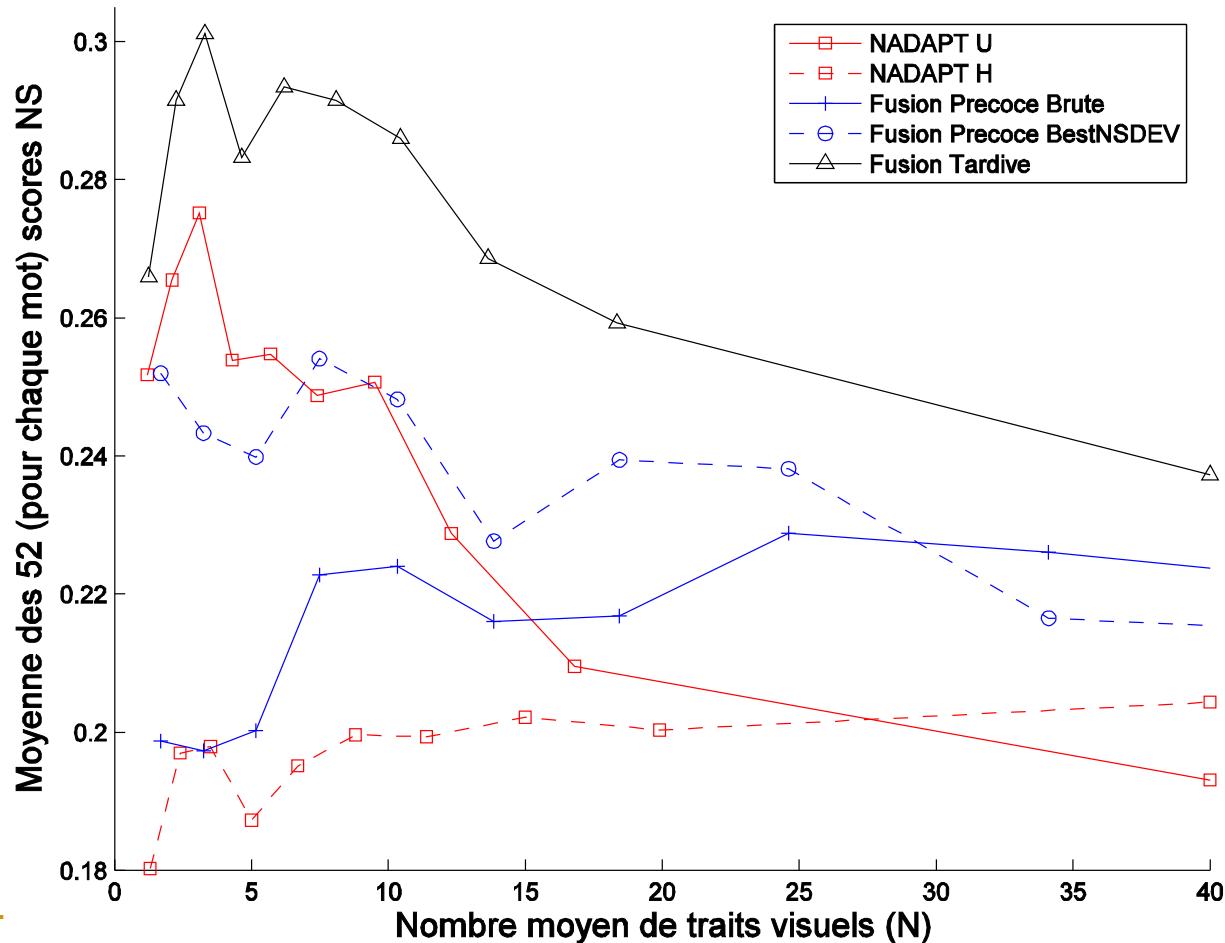


Scores de classification de chaque mot H versus U

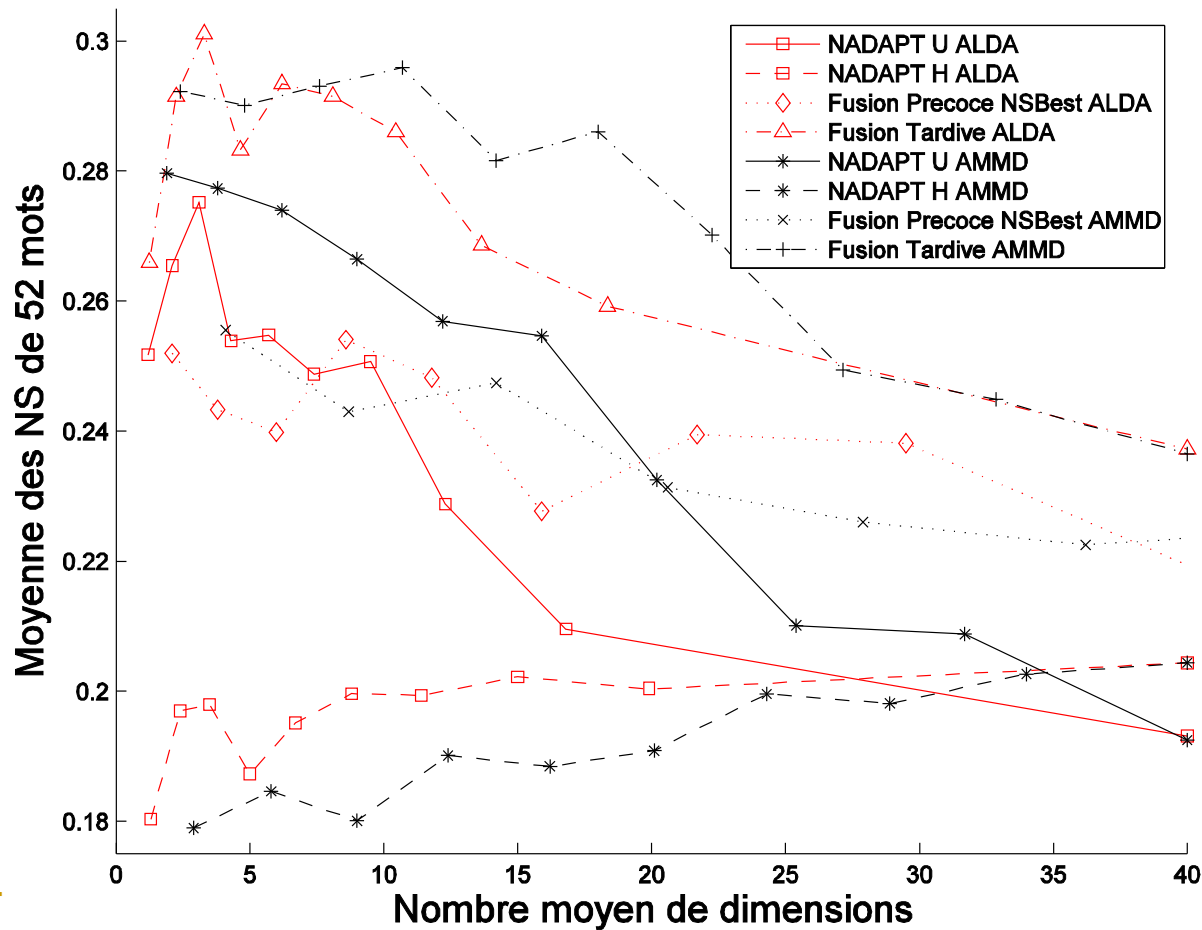


Scores moyens des classifications de 52 mots en fonction du nombre moyen de dimensions utilisées

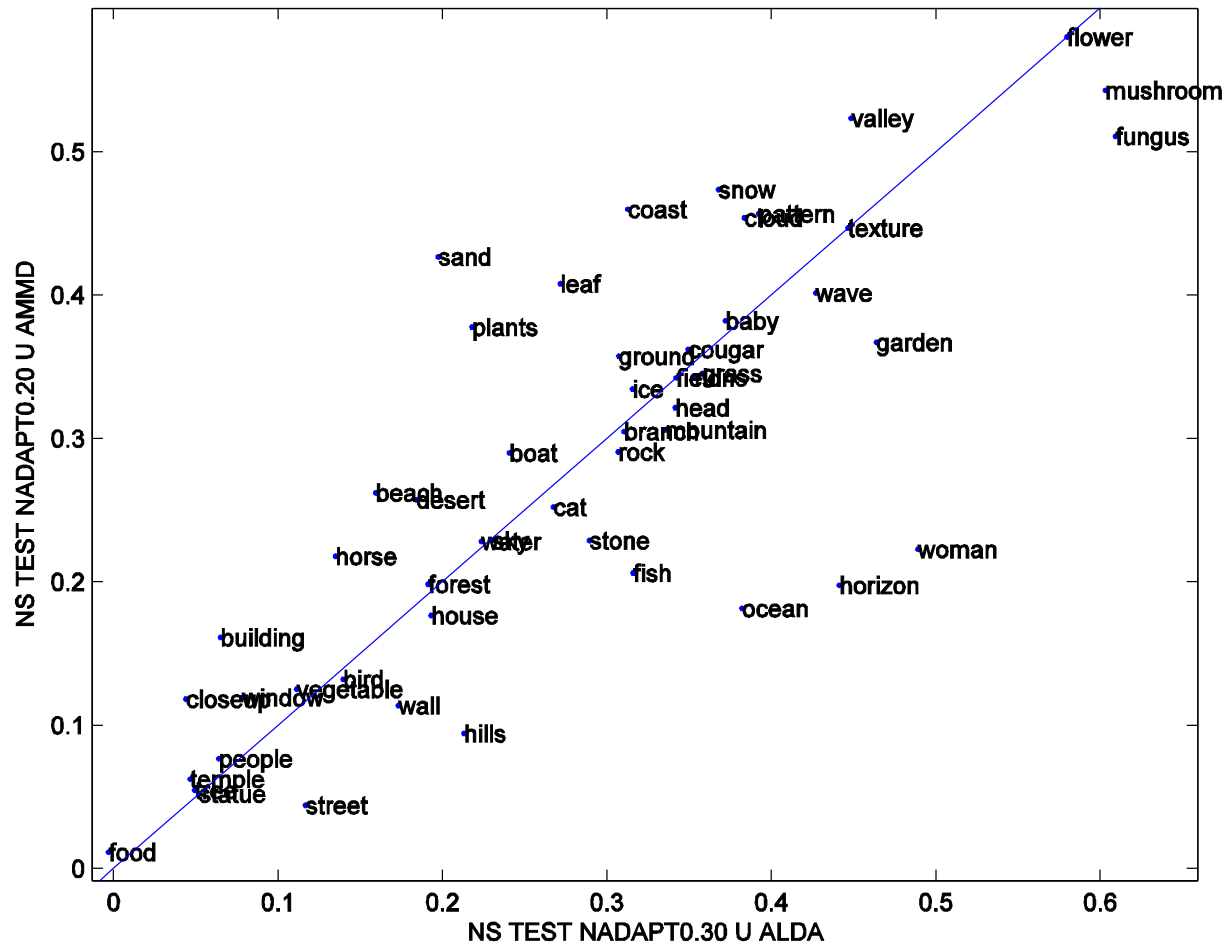
Expériences fusions tardives de U et de H



Scores moyens des classifications ALDA versus AMMD



Scores de classification pour chaque mot ALDA versus AMMD



Résultats

| | U sans sélection | U avec sélection par ALDA | U avec sélection par AMMD | H sans sélection | ACP U |
|----------------------------|------------------|---------------------------|---------------------------|------------------|-----------------|
| τ | 1.0 | 0.3 | 0.2 | 1.0 | 0.7 |
| Nombre moyen de dimensions | 40 - | 3.1 (-92%) | 3.8 (-91%) | 40 (+0%) | 6 (-85%) |
| NS Moyen Classification | 0.192 - | 0.275 (+43%) | 0.277 (+44%) | 0.204 (+6%) | 0.211 (+10%) |

Score aléatoire NS=0.02

Conclusion

Nous avons montré théoriquement et expérimentalement que le critère de l'ALDA peut être utilisé efficacement sur des images mal annotées

Le critère de l'ALDA est plus facile à calculer et moins sensible aux bruits que le critère AMMD

Toutes les expériences montrent :

Qu'il est préférable de choisir les dimensions visuelles les plus discriminantes pour chaque mot, plutôt que de prendre pour tous les mots toutes les dimensions visuelles

Merci de votre attention
