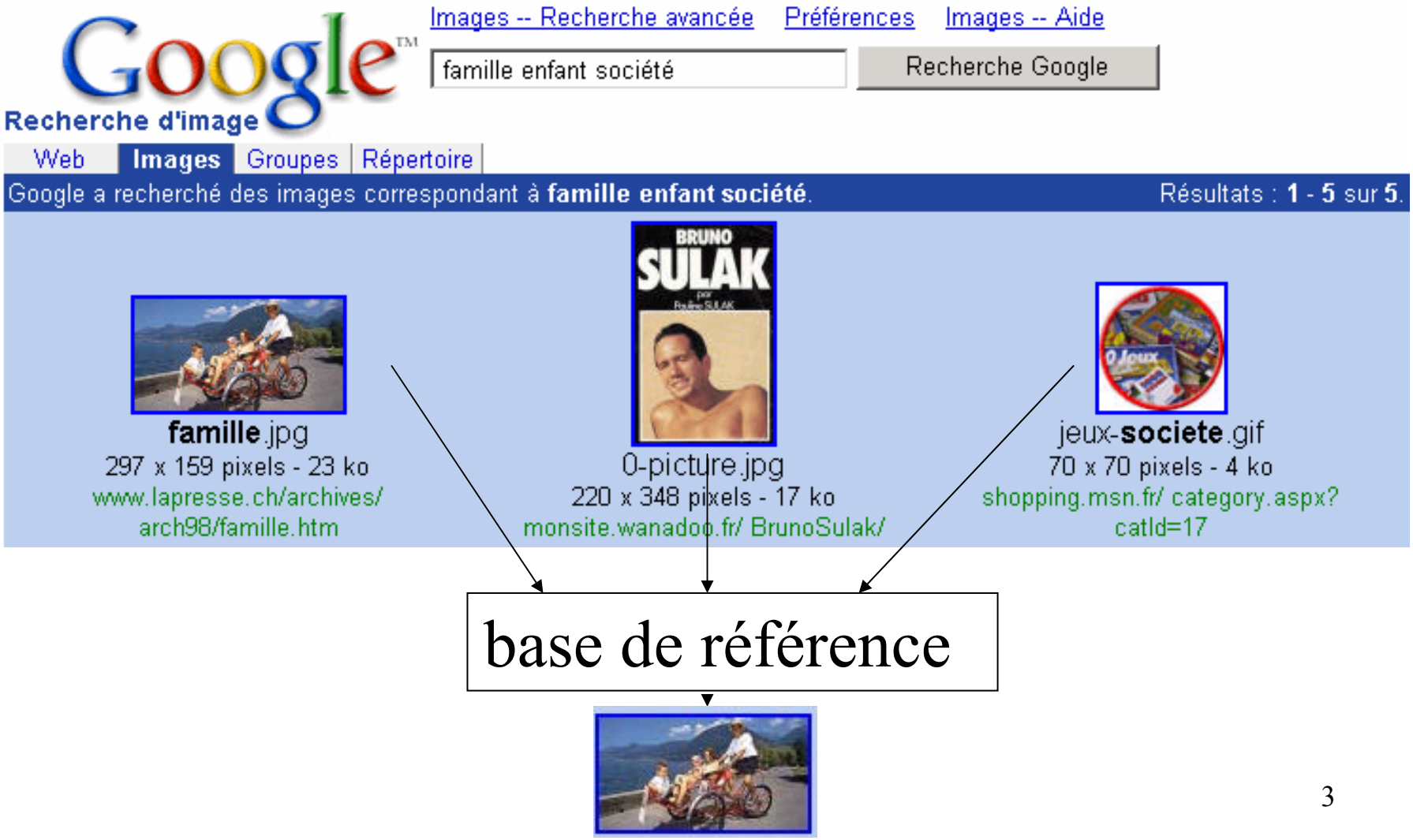
Rehaussement de la classification textuelle d'une base de données photographiques par son contenu visuel

Sabrina Tollari
Sous la direction de Hervé Glotin et Jacques Le Maitre
Laboratoire SIS - Équipe Informatique
Juin 2003

Plan

- Problématique
- Présentation du corpus
- Protocole du système visuo-textuel
- Expérimentations
- Discussion
- Conclusion et perspectives

Comment raffiner une requête textuelle d'images ?



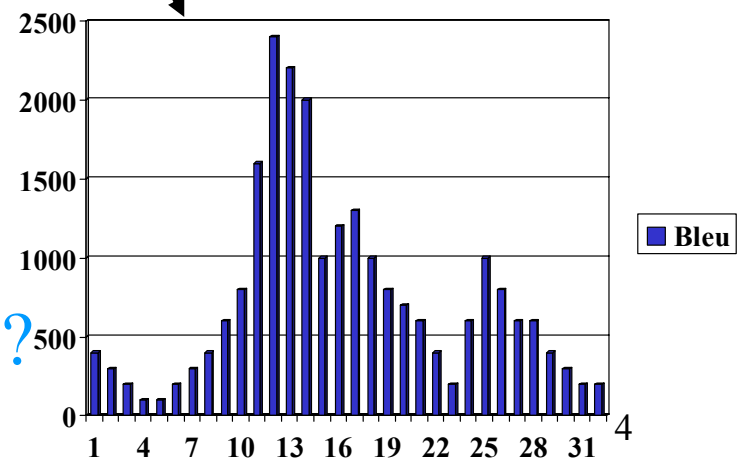
Problématique

Indices textuels

Indices visuels

Paysage Cameroun
Agriculture

Cohérence ?



Nature des indices

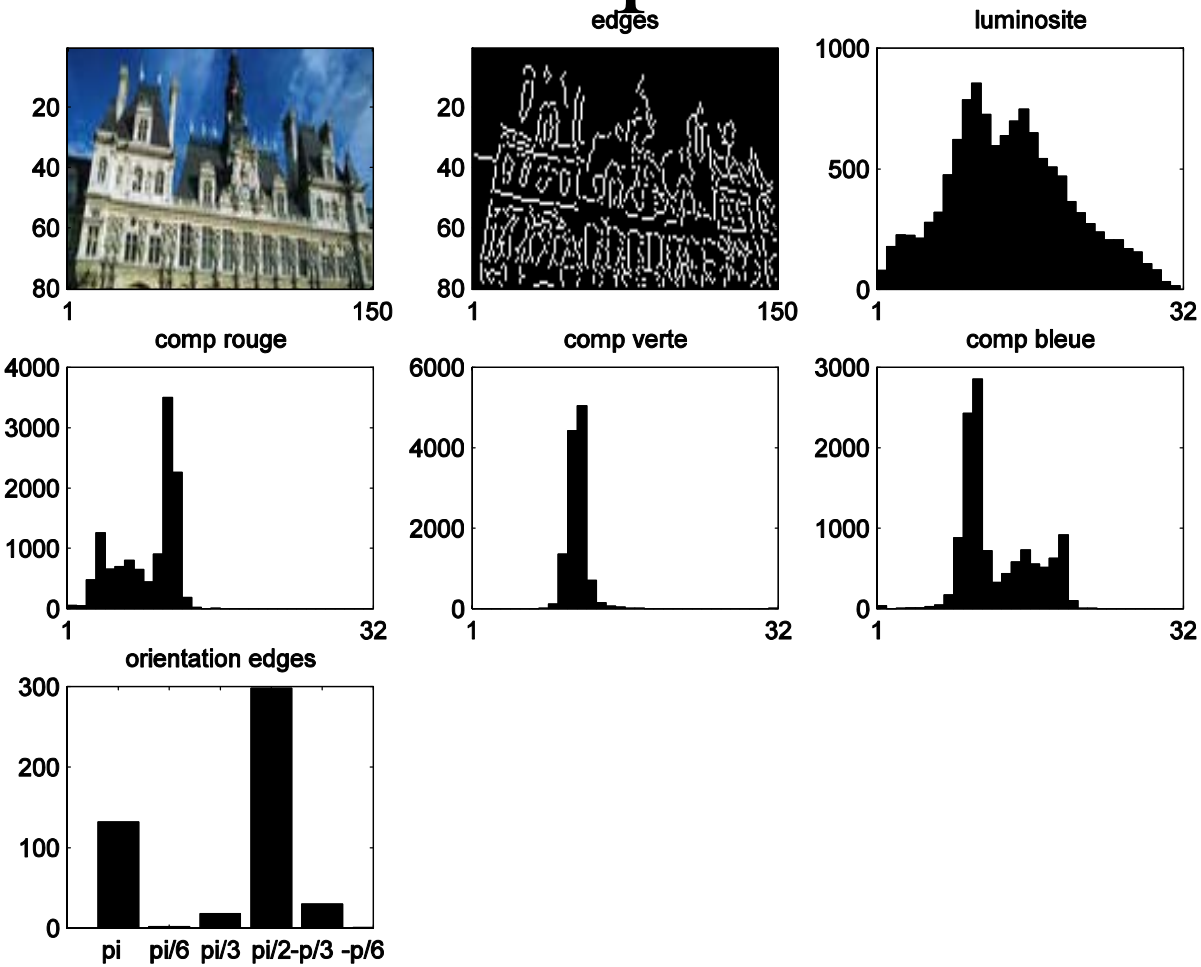
- Indices textuels :
  - Indexation manuelle : mot-clés, metadata, annotation…
  - Indexation automatique : mots clés de la légende, du texte entourant l'image…
- Indices visuels :
  - Forme : contour, surface, transformée en ondelettes, transformée de Fourrier…
  - Couleur : espaces RGB, HSV…
  - Texture : grossièreté, contraste, directionnalité…
  - Localisation, segmentation en zones d'intérêt…

Systèmes de recherche d'images

Indices visuels uniquement	**Indices visuels et/ou textuels**
Virage(1996) NeTra(1997) SurfImage(INRIA,1998) IKONA(INRIA, 2001)	Chabot(1995) QBIC(IBM,1995) VisualSeek(1996) MARS(1997)

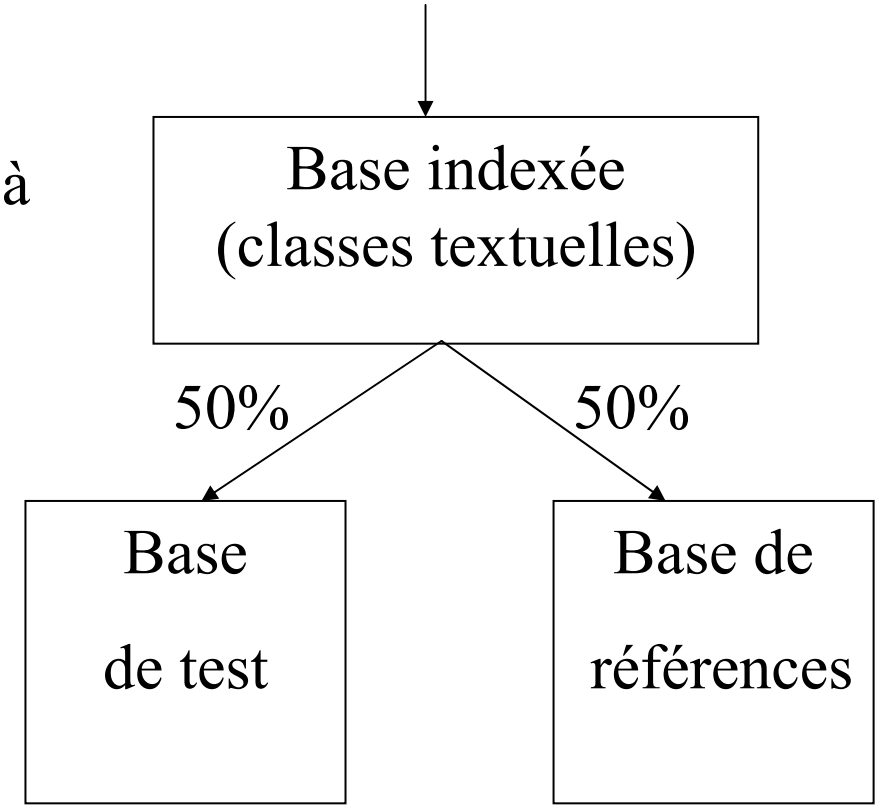
Présentation du corpus

- 665 photos de presse
- Indexées textuellement par une iconographe à partir des mot-clés extraits d'un thésaurus
- Indexées visuellement par les histogrammes rouge, vert, bleu, luminance et direction



Protocole

Corpus d'images



Étape A

Classer les images à partir des indices textuelles

Étape B

Diviser aléatoirement en deux bases

Étape C

Reclasser les images de la base de test par rapport aux indices textuels, aux indices visuels et par fusion des classifications visuelle et textuelle

Construction de la base indexée par classification ascendante hiérarchique (CAH) des indices textuelles

- Lance et Williams, 1967
- Principe : regrouper ensemble des images proches
- Intérêt : cette méthode peut être mise en œuvre sur des images n'ayant pas de lien sémantique apparent
- Objectif : obtenir des classes sémantiquement et numériquement significatives

Algorithme de la CAH

Algorithme Classification ascendante hiérarchique

Données :
- E : ensemble des n images à classer
- Tableau $n \times n$ des distances entre images

Variables :
- C : ensemble des c classes

Début

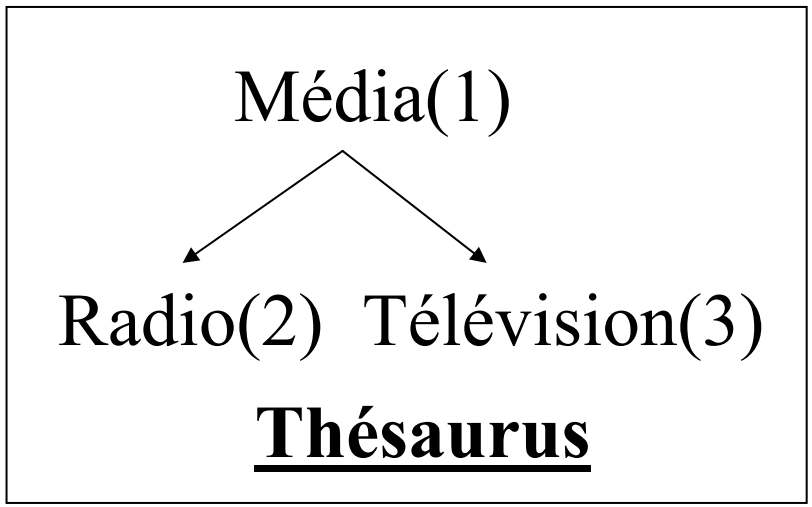
Pour chaque image e de E **faire**
Créer une classe dans C contenant e
fin pour

Tant que C a plus d'une classe **faire**
Agréger les deux classes C_p et C_q les plus similaires relativement aux distances de leurs images repectives
fin tant que

Fin

Représentation textuelle des images : le modèle vectoriel

- Salton, 1971
- Une image est représentée par un vecteur des mot-clés
- Exemple :



Soit une image I telle que Terme(I)={Radio}

- Vecteur(I)=(0,1,0)
- Vecteur_etendu(I)=(1,1,0)

Mesure de la similarité : le cosinus

Soit $\vec{x}$ et $\vec{y}$ les vecteurs des images X et Y

$$cos(\vec{x}, \vec{y}) = \frac{\sum_{j=1}^{n} x_j \times y_j}{\sqrt{\sum_{j=1}^{n} x_j^2} \times \sqrt{\sum_{j=1}^{n} y_j^2}}$$

La distance entre deux images X et Y est :

dist(X,Y) = 1- $cos(\vec{x}, \vec{y})$

Critère d'agrégation

- Critères classiques
  - Plus proche voisin
  - Diamètre maximum (ou voisin le plus éloigné)
  - Distance moyenne
  - Critère de Ward…
- Les classifications obtenues sur notre corpus par ces critères n'étaient pas significatives

Nouveau critère d'agrégation

- La distance par « Diamètre maximum contraint » de contrainte CT entre une classe C_p et une classe C_q est définit par :

$$D(C_p, C_q) =$$

- $max\{dist(I_r, I_s) < CT; I_r \in C_p,\ I_s \in C_q\}$
  si $\exists I_a \in C_p,\ I_b \in C_q$ telles que $dist(I_a, I_b) < CT$
- $min\{dist(I_r, I_s); I_r \in C_p,\ I_s \in C_q\}$
  sinon.

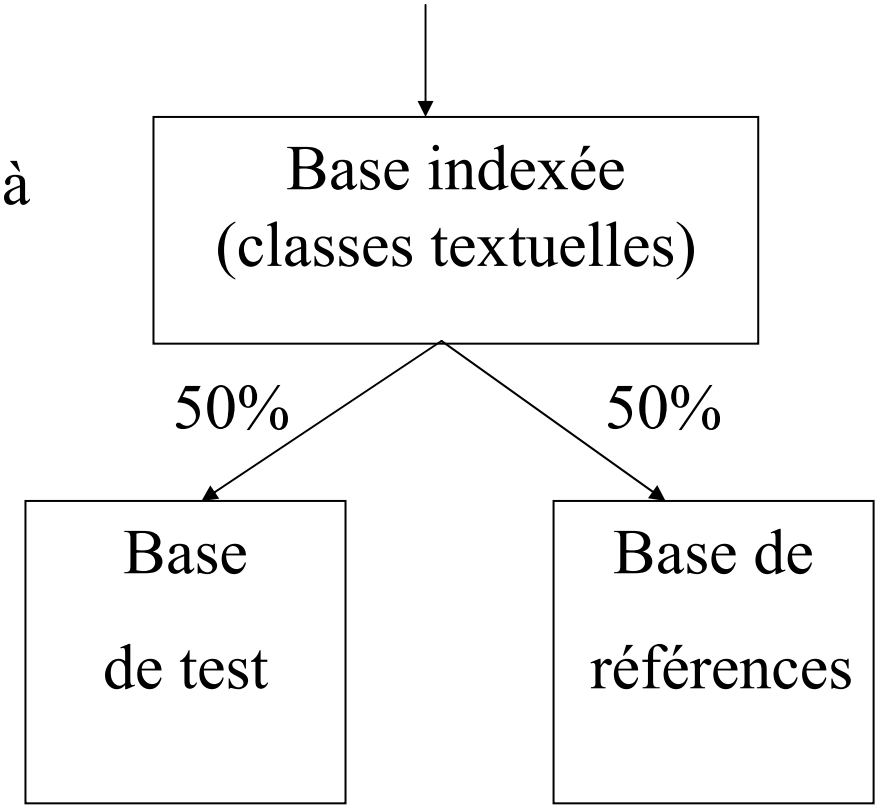
Classification obtenue

- 24 classes
  - contenant de 8 à 98 images
  - sémantiquement homogènes

Classe	Fréquence 1	Fréquence 2	Fréquence 3
1	Femme	Ouvriers	Industrie
2	Cameroun	Agriculture	Paysage
3	Constructeurs	Transport	Automobile
4	Contemporaine	Portrait	Rhône
5	Société	Famille	Enfant

Protocole

Corpus d'images



Étape A

Classer les images à partir des indices textuelles

Étape B

Diviser aléatoirement en deux bases

Étape C : déterminer la classe d'une image de la base de test

Image de la base de test (classe d'origine **Co**)

C_1

Paysage, agriculture, Cameroun

C_2

Femme, Ouvrier, Industrie

Base de références

Classe estimée **Ce**

(obtenue par distance minimale)

Si **Co**≠**Ce** alors erreur

Les classifications

1. Classification textuelle pure
2. Classification visuelle pure
3. Classification par fusion des classifieurs visuels et textuels

Distance de Kullback-Leibler(1951)

Soit x et y deux distributions de probabilité

Divergence de Kullback-Leibler :

$$KL(x,y) = \sum_{j=1}^{n} x_j \log \frac{x_j}{y_j}$$

Distance de Kullback-Leibler :

$$DKL(x,y) = KL(x,y) + KL(y,x)$$

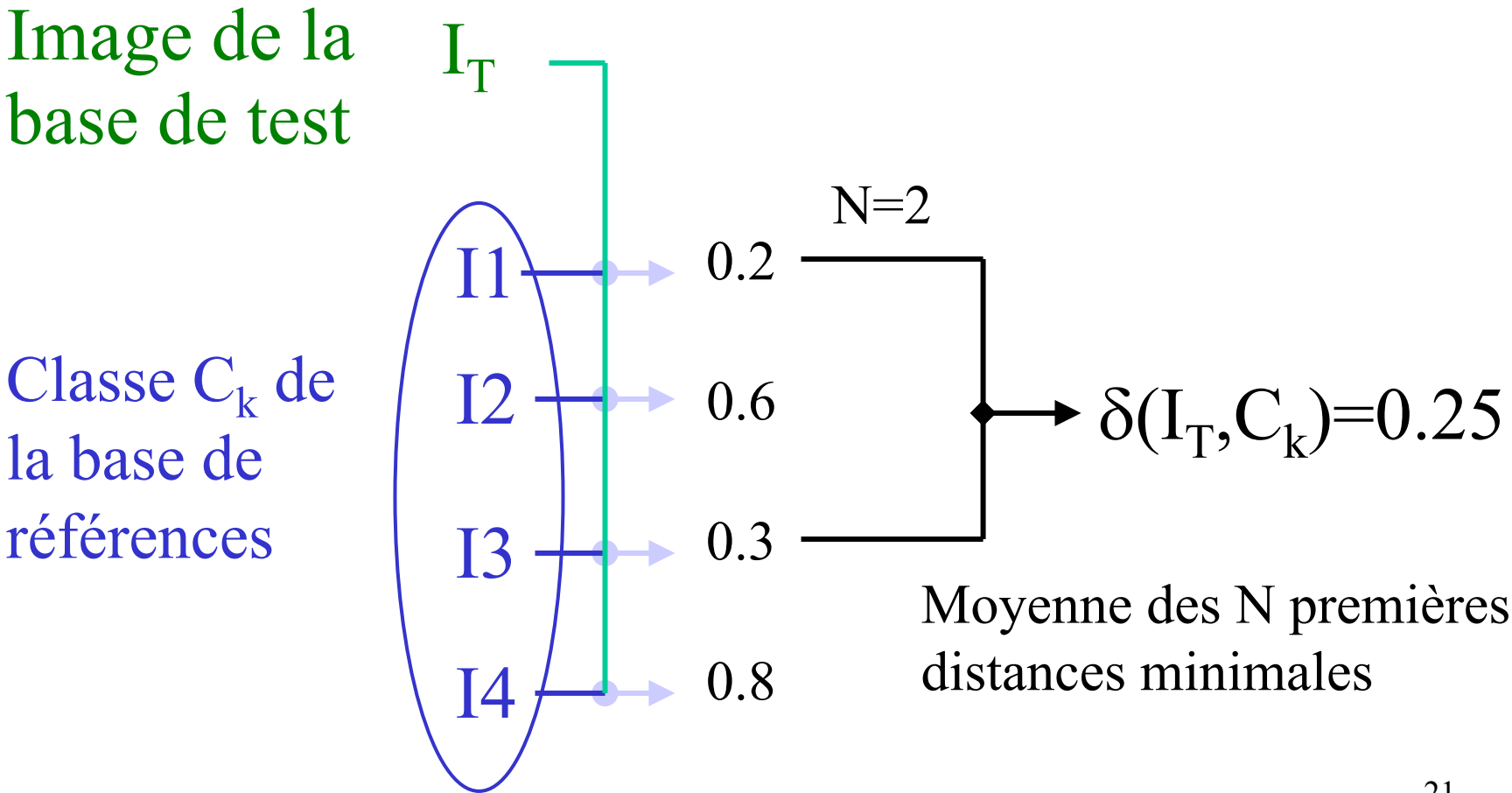
1.Classification textuelle pure

- Vecteur moyen pour chaque classe $\vec{C}_k^t$
- Classe textuelle de l'image I_T:

$$C^t(I_T) = \text{argmin}_{k\in\{1,2,\ldots,c\}} DKL(\vec{I}_T^t, \vec{C}_k^t)$$

Résultats	Textuelle avec thésaurus (vecteur étendu)	Textuelle sans thésaurus (vecteur non-étendu)
Taux d'erreur	1.17 %	13.72 %

2. « Fusion précoce » des indices visuels



2. Classification visuelle pure

$$C^v(I_T) = \text{argmin}_{k\in\{1,2,\ldots,c\}}\delta(I_T, C_k)$$

N	1	2	3	4
Rouge*	75.68	74.50	**71.76**	**71.76**
Vert*	79.60	78.03	76.86	**76.07**
Bleu*	78.03	77.64	78.03	**77.25**
Luminance*	79.21	78.03	**76.07**	77.64
Direction*	84.70	78.03	**76.86**	**76.86**

* Taux d'erreur en % Taux d'erreur théorique : 91.6%

3. « Fusion tardive » visuo-textuelle

- Probabilité d'appartenance de l'image I_T à la classe C_k par fusion des probabilités textuelles et visuelles :

$$P_{I_T}^{v \vee t}(C_k) = \sum_{j=1}^{5} P_{I_T}^{v}(C_k|A_j) \times \omega'(A_j) + P_{I_T}^{t}(C_k) \times \omega'(A_6)$$

$$C^{v \vee t}(I_T) = \underline{\text{argmax}}_{k \in \{1,2,...,c\}} P_{I_T}^{v \vee t}(C_k)$$

3. Définitions des probabilités d'appartenance d'une image à une classe

$$P^t_{I_T}(C_k) = 1 - \frac{DKL(\vec{I_T}, \vec{C_k})}{\sum_k DKL(\vec{I_T}, \vec{C_k})}$$

$$P^v_{I_T}(C_k|A) = 1 - \frac{\delta_A(I_T, C_k)}{\sum_k \delta_A(I_T, C_k)}$$

A $\in$ {Rouge, Vert, Bleu, Luminance, Direction}

3. Définitions des pondérations

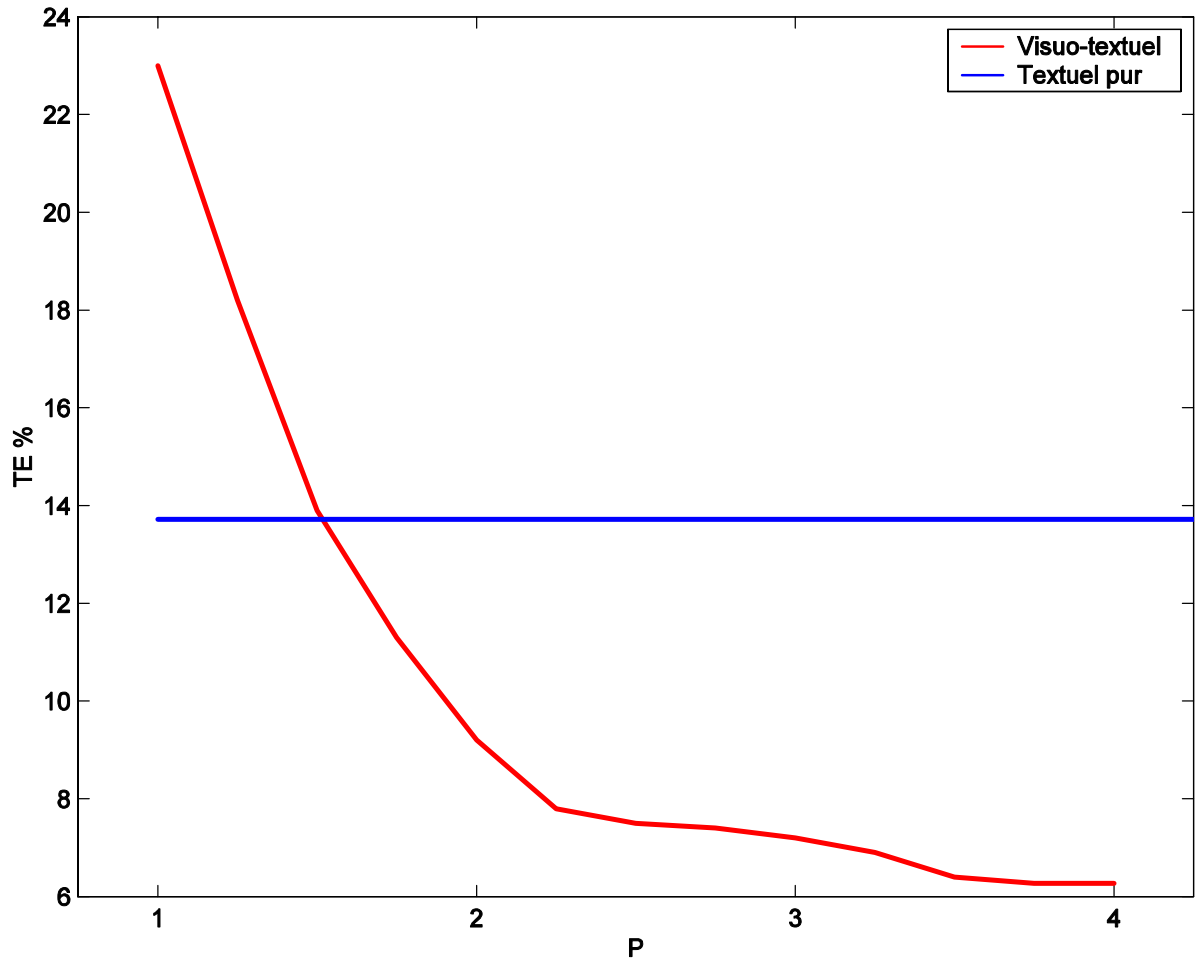
- Soit TE(j) le taux d'erreur du classifieur utilisant les attributs A_j

$$\omega(A_j) = \frac{1 - TE(j)}{\sum_{i=1}^{6} 1 - TE(i)}$$

- Élévation à la puissance p pour contraster les poids

$$\omega'(A_j) = \frac{\omega(A_j)^p}{\sum_{i=1}^{6} \omega(A_i)^p}$$

3. Influence du paramètre p



Rappel : taux d'erreur visuel 71 %

Résultat : rehaussement visuo-textuel

Résultats	Textuelle sans thésaurus	Fusion visuo-textuelle	Gain
Taux d'erreur	13.72%	6.27%	+54.3%

Discussion

- Ces résultats doivent être affinés sur une base de données plus grande
- La méthode de pondération doit être comparée à d'autres (entropie des distributions…)
- Les poids devraient être optimisés sur une base de développement

Conclusion

- Il existe une cohérence entre l'indexation textuelle et visuelle
- Cette cohérence permet le rehaussement d'une recherche par mot-clés d'images par leur contenu
- Méthode simple et automatique, donc utilisable sur le web
- Ce système peut être utilisé avec n'importe qu'elle type d'indices visuels

Perspectives

- Utilisation pour raffiner les recherches textuelles sur le Web (Google, Altavista…)
- Inversion du système pour corriger des erreurs d'indexation textuelle des images sur le Web (base de références visuelles)