Toward Cross-Language and Cross-Media Image Retrieval

Carmen Alvarez, Ahmed Id Oumohmed, Max Mignotte, Jian-Yun Nie

DIRO (Département d'Informatique et de Recherche Opérationelle)
University of Montreal

summer 2004

RALI group

- Recherche Appliquée en Linguistique Informatique
- Carmen Alvarez, Jian-Yun Nie

Image processing group

- Ahmed Id Oumohmed, Max Mignotte

1. Plan

1. Introduction

2. Image processing-based learning procedure

3. Cross-language text retrieval

4. Results and discussion

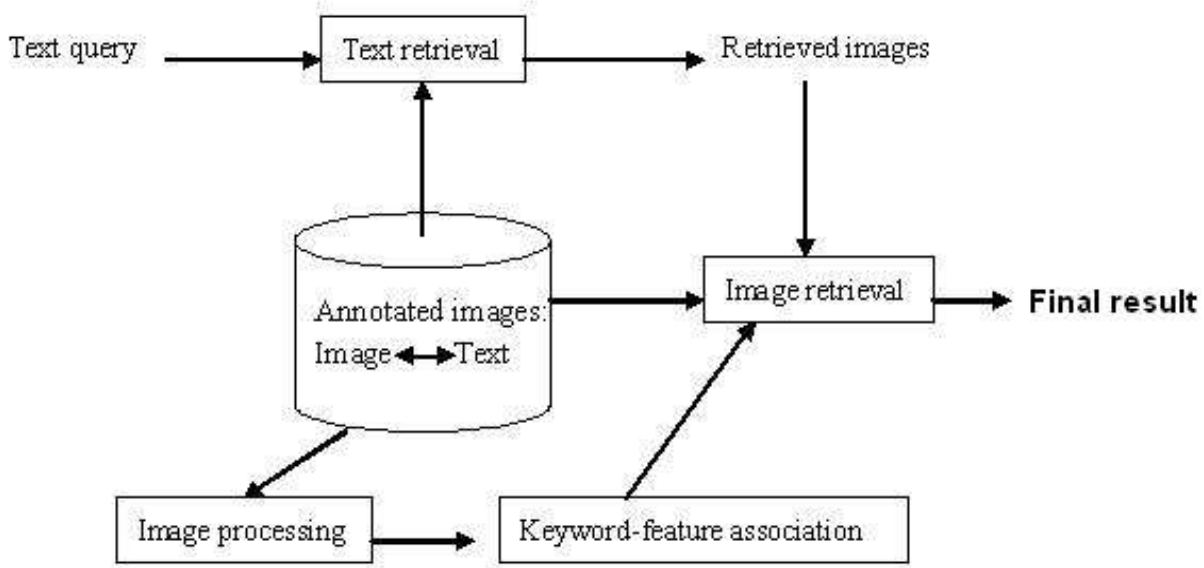
1. Introduction
1.1. Main approach

French to English translation
- translation model trained on parallel web pages

Translation from other languages
- bilingual dictionaries

Image retrieval from text queries
- text query using image captions
- image query using different features
- semantic query (image or keyword)

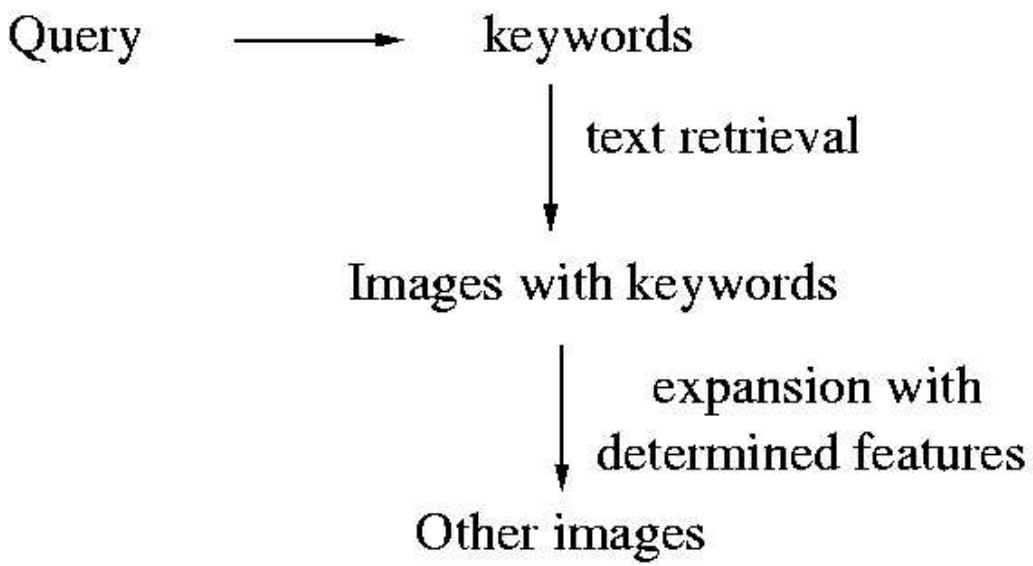
1. Introduction
1.2. Work flow of image retrieval



1. Introduction
1.2. Learning principle

- keyword $w \longrightarrow$ images with in annotation
- create centroid(s) according to some features
- retrieve similar images to each centroid in the database
- how many images retrieved are with w in annotation
- retain relevant features and associated centroids

1. Introduction
1.3. expansion notion



2. Image processing-based learning procedure

2.1. Overview

2.2. Edge class

2.3. Texture class

2.4. Shape class

2.5. Learning procedure

2.6. Results

2. Image processing-based learning procedure
2.1. Overview

for each keyword w

1. high level visual feature estimation

- consider the 3 high level visual features :
  - ▶ texture
  - ▶ edge
  - ▶ shape
- discriminant measure
- challenge

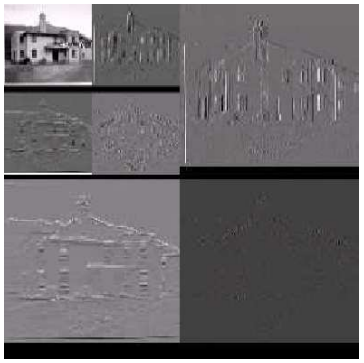
2. representative images and combination

- choose 1, 2 or 3 high level visual features
- normalize the similarity measures
- $\longrightarrow$ set of representative images
- finally : cross-language Information $\longrightarrow$ refine the retrieval process

2. Image processing-based learning procedure
2.2. Edge class - Haar wavelet transform

- multi resolution analysis image
- local texture and distribution of of edges at multiple scale
- recursive numeric filtering and sub sampling
- mean and standard deviation of the energy distribution
- 3 level decomposition $\longrightarrow$ 20 wavelet coefficients

(a)



(b)

a) The original image STAND03_1028/STAND03_26737_BIG.PS of size 238×308.
b)The Haar wavelets coefficients after the image a) is adjusted to size 256×256.

2. Image processing-based learning procedure
2.3. Texture class - algorithm 1 / 2

- $\forall$ pixel (x, y) in the image
- for $k = 1, 2, \ldots, 6$
- $A_k(x, y)$ is the moving average over $2^k \times 2^k$ window

$$A_k(x,y) = \sum_{i=(x-2^{k-1})}^{i=(x+2^{k-1}-1)} \sum_{j=(y-2^{k-1})}^{j=(y+2^{k-1}-1)} \frac{I(i,j)}{2^{2k}}$$

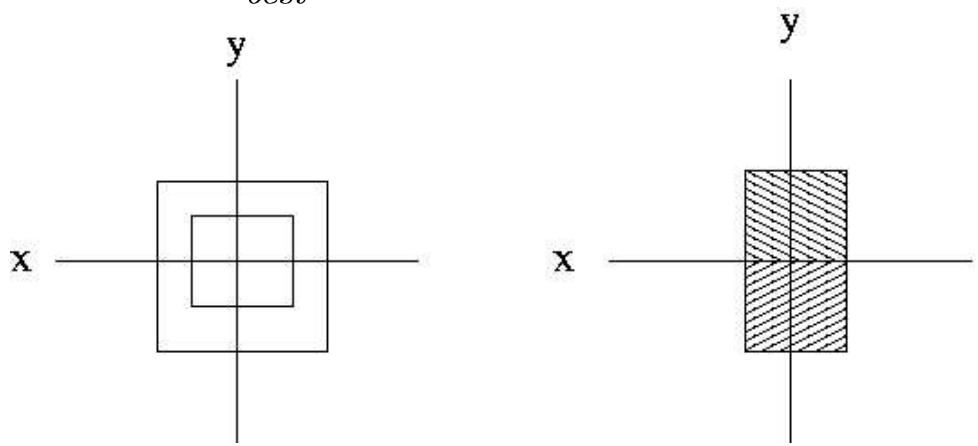
- $I(i, j)$ is the intensity pixel of the image at pixel (i, j)

- $\forall$ pixel (x, y) in the image
- for horizontal and vertical directions
- for non-overlapping windows just on opposite sides
- compute the differences $E_{k,horizontal}$ and $E_{k,vertical}$

$$\begin{aligned} E_{k,horizontal} &= |A_k(x + 2^{k-1}, y) - A_k(x - 2^{k-1}, y)| \\ E_{k,vertical} &= |A_k(y + 2^{k-1}, y) - A_k(y - 2^{k-1}, y)| \end{aligned}$$

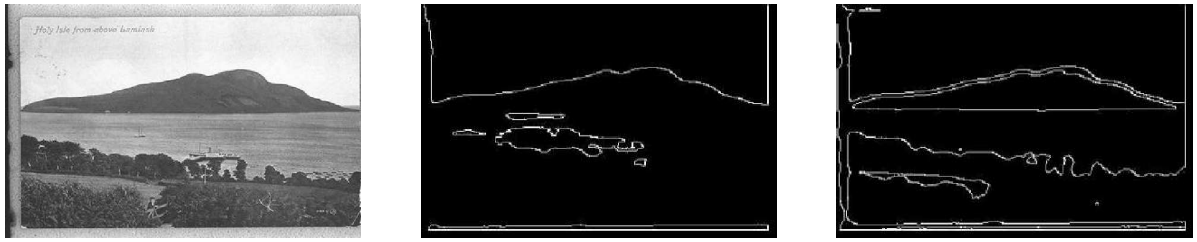
2. Image processing-based learning procedure
2.3. Texture class - algorithm 2 / 2

- $\forall$ pixel (x, y) in the image
- without considering direction
- for k which maximize $E_k(i, j)$
- best size of texture resolution is $S_{best} = 2^k$



2. Image processing-based learning procedure
2.4. Shape class - Method

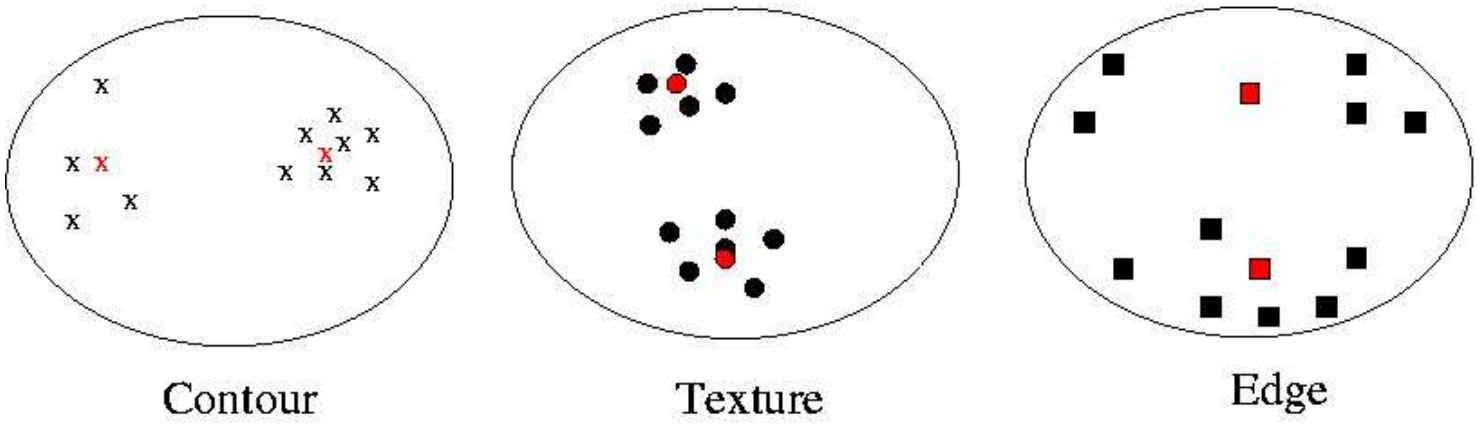
- several possible approaches
- $K \times K$ image blocks $\longrightarrow$ set of vectors of dimension K^2
- cluster into R classes $\longrightarrow$ R regions $\longrightarrow$ contours
- horizontal, vertical, first and second diagonal
- 4 bins histogram



(a) **(b)** **(c)**

a) The original image STAND03_2093/STAND03_7363_BIG.JPG
b) 4×4 *pixel blocks and clustering result into* 2 *regions*
c) 4×4 *pixel blocks and clustering result into* 3 *regions*

2. Image processing-based learning procedure
2.5. Learning procedure - Overview of procedure



Clustering for the three vector spaces (clusters = 2) and the centroid vectors

2. Image processing-based learning procedure
2.5. Learning procedure - Notations

- $\mathbf{I}_w$: images annotated by keyword w
- 3 high level estimation (texture, edge and shape)
  $\Downarrow$
  3 vector spaces $[D^{\mathsf{t}exture}_{I_w}, D^{\mathsf{e}dge}_{I_w}, D^{\mathsf{s}hape}_{I_w}]$
- $D^{class}_{I_w}$ either $D^{texture}_{I_w}$ or $D^{edge}_{I_w}$ or $D^{shape}_{I_w}$
- k : cluster id $(k = 1, \ldots, K)$
- $\mathbf{I}^{class}_{k,w}$: representative images for w according to class $class$ and to centroid of cluster k

2. Image processing-based learning procedure
2.5. Learning procedure - Choosing representative images

- $\forall\ k \in \{1, \ldots, K\}$, $\forall\ class \in \{texture, edge, shape\}$

  $\Longrightarrow \{\overrightarrow{D_{1,w}^{class}}, \ldots, \overrightarrow{D_{K,w}^{class}}\}$: prototype vectors

- $\mathbf{I}_{k,w}^{class}$: closest images to $\overrightarrow{D_{k,w}^{class}}$
- $N_{k,w}^{class}$: first top level T after retrieval over the entire database
- if $N_{k,w}^{class} > \xi$, retain $class$ and $\overrightarrow{D_{1,w}^{class}} \Longrightarrow \mathbf{I}_{k,w}^{class}$

2. Image processing-based learning procedure
2.6. Results - keywords $\longrightarrow$ important visual features 1/2

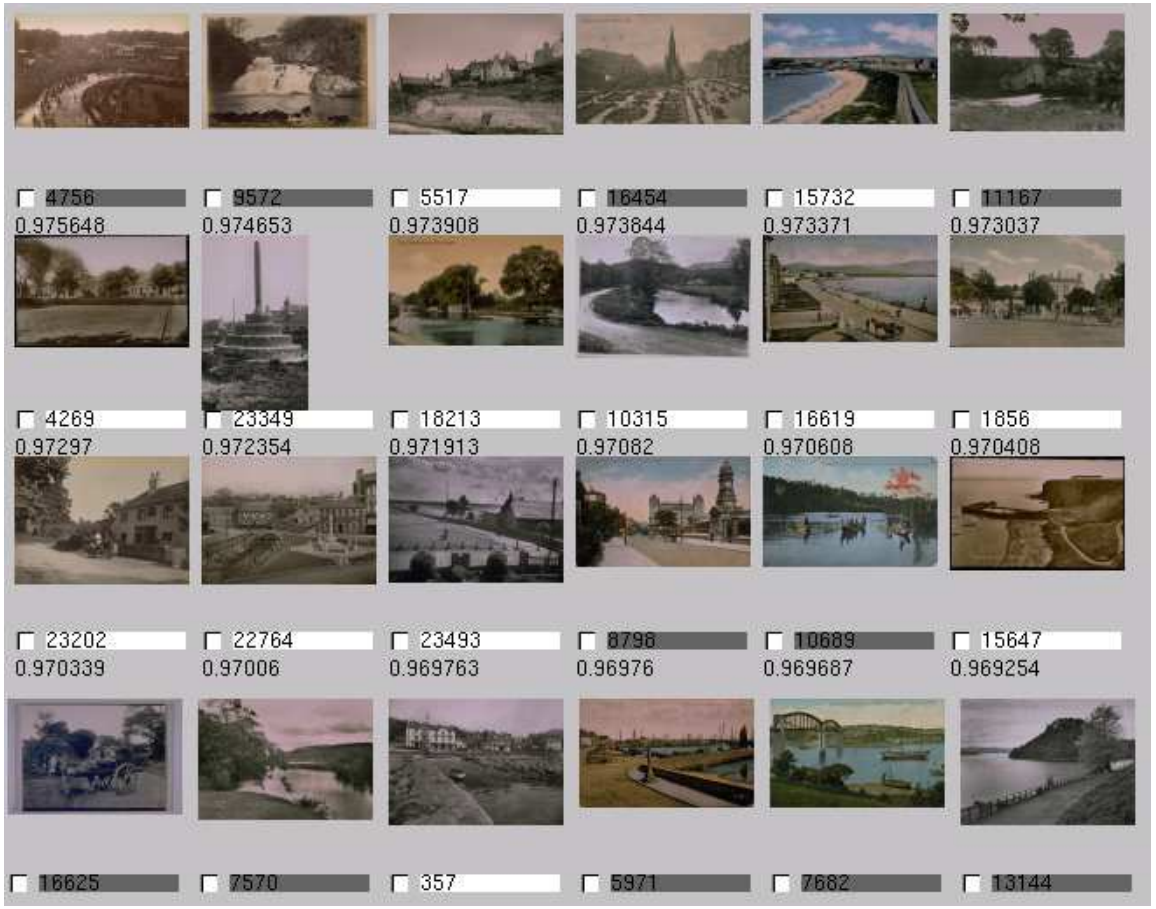
mountain	$\longrightarrow$	shape	
tree	$\longrightarrow$	wavelets	
street	$\longrightarrow$	wavelets	
interior	$\longrightarrow$	shape	
plant	$\longrightarrow$	shape	
hous	$\longrightarrow$	wavelets	
wood	$\longrightarrow$	wavelets	
portrait	$\longrightarrow$	shape	
stone	$\longrightarrow$	shape	wavelets

2. Image processing-based learning procedure
2.6. Results - keywords $\longrightarrow$ important visual features 2/2

cathedr	$\longrightarrow$	shape
tower	$\longrightarrow$	wavelets
garden	$\longrightarrow$	shape
boat	$\longrightarrow$	wavelets
lake	$\longrightarrow$	texture
land	$\longrightarrow$	wavelets
rive	$\longrightarrow$	wavelets

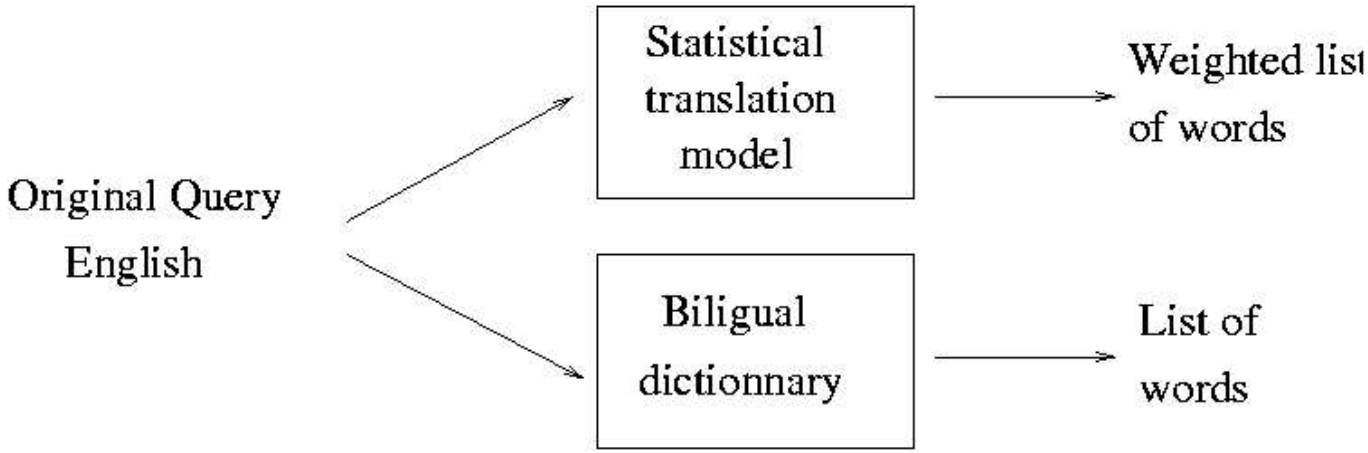
2. Image processing-based learning procedure
2.6. Results - keyword garden

2. Image processing-based learning procedure
2.6. Results - keyword hous



2. Image processing-based learning procedure
2.6. Results - keyword plant

3. Cross-language text retrieval
3.1. Query translation



- English $\longrightarrow$ French : Statistical translation model (web corpus)
- English $\longrightarrow$ Other languages : bilingual dictionary only

3. Cross-language text retrieval
3.2. Text and image retrieval combination

1. **text retrieval**
   - $\longrightarrow R_{text}(q)$
   - $\lambda_{text} = 0.8$
2. **learning procedure list**
   - $\forall\ \mathbf{I}_{k,w}^{class}$, normalize scores between 0 and 1
   - same weight for selected classes
   - $\longrightarrow R_{cluster}(q)$
   - $\lambda_{cluster} = 0.1$
3. **query image lists (texture, edge and shape)**
   - $\lambda_{class} = 0.033$
4. **final combination**

$$R(i,q) = \lambda_{text} R_{text}(q) + \lambda_{cluster} R_{cluster}(q)$$
$$+\lambda_{edge} R_{edge}(i) + \lambda_{texture} R_{texture}(i) + \lambda_{shape} R_{shape}(i)$$

4. Results and discussion
ImageCLEF 2004 adhoc retrieval task

	Group	Submission ID	MAP	%monolingual	Rank
Monolingual					
	Montreal	UmenTNFBTI	0.56	Na	5
Dutch					
	Montreal	UmnlTFBTI	0.4	68.27	7
Finnish					
	Montreal	UmfiTFBTI	0.23	40.02	1
French					
	Montreal	UmfrTFBTI	0.51	87.4	1
Italian					
	Montreal	UmitTFBTI	0.36	61.34	8
Spanish					
	Montreal	UmesRevTFBTI	0.45	76.82	7
Swedish					
	Montreal	UmsvTFBTI	0.34	57.98	1

4. Results and discussion
Example

```
mysql> select  word, attribut,nClusters,idClass,top10
from word_attribut where word like "stone%";
+-------+----------+-----------+---------+-------+
| word  | attribut | nClusters | idClass | top10 |
+-------+----------+-----------+---------+-------+
| stone | texture  |         2 |       1 |     4 |
| stone | texture  |         3 |       1 |     4 |
| stone | texture  |         6 |       1 |     4 |
| stone | wavelets |         3 |       1 |     4 |
| stone | wavelets |         4 |       0 |     6 |
| stone | wavelets |         5 |       0 |     6 |
| stone | wavelets |         6 |       0 |     4 |
| stone | wavelets |         6 |       1 |     6 |
| stone | wavelets |         6 |       3 |     4 |
| stone | shape    |         2 |       1 |     4 |
| stone | shape    |         3 |       2 |     5 |
| stone | shape    |         4 |       0 |     5 |
| stone | shape    |         4 |       1 |     5 |
| stone | shape    |         4 |       2 |     5 |
| stone | shape    |         5 |       0 |     7 |
| stone | shape    |         5 |       1 |     5 |
| stone | shape    |         5 |       2 |     8 |
| stone | shape    |         6 |       2 |     9 |
| stone | shape    |         6 |       3 |     5 |
+-------+----------+-----------+---------+-------+
19 rows in set (0.01 sec)
```