UPMC/LIP6 at ImageCLEFphoto 2008: on the exploitation of visual concepts

Sabrina Tollari, Marcin Detyniecki, Ali Fakeri Tabrizi,
Massih-Reza Amini, Patrick Gallinari

Université Pierre et Marie Curie – Paris 6
UMR CNRS 7606-LIP6

AVEIR PROJET : ANR-06-MDCA-002

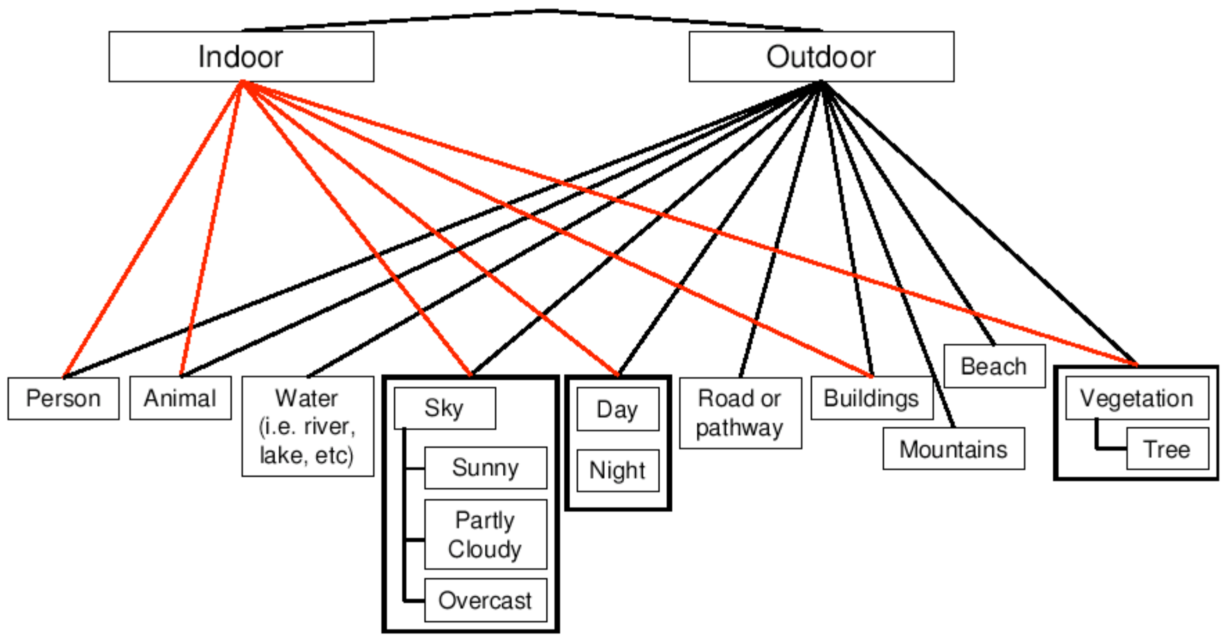
Motivation

- How to solve the challenge of image retrieval?
- It would be great to have an "image to text" translator.
- General dictionary seems impossible
  - LSCOM: 449 visual concepts over 61901 shots (TrecVID'05)
- but… maybe a simpler dictionary could help to improve the classical text-based image-retrieval.

ImageCLEF 2008

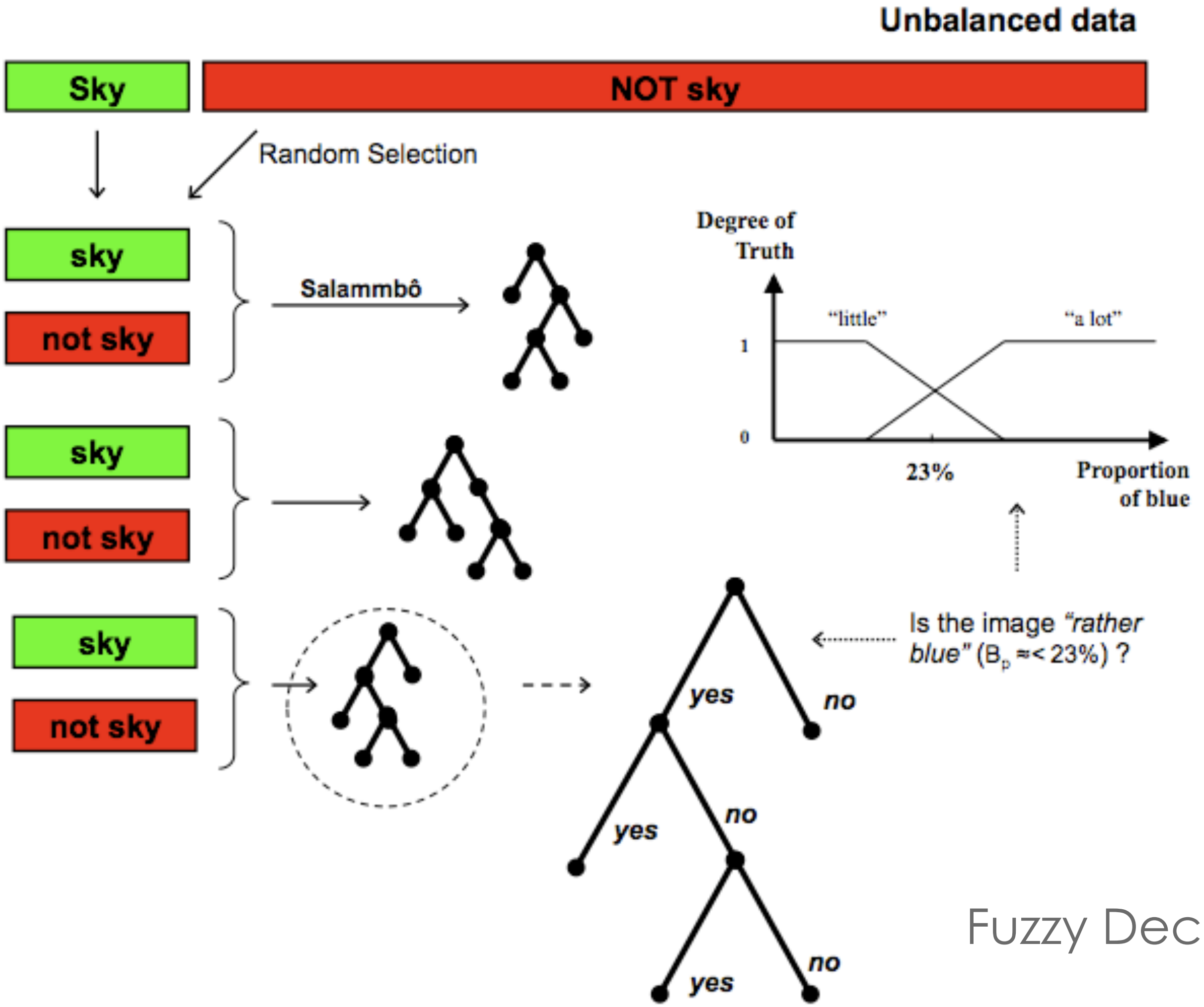
- Visual Concept Detection Task (VCDT)
  - **How to “translate” an “image” into “text” ?**
  - Forest of Fuzzy Decision Trees (FFDT)
- Photo Retrieval Task (PHOTO)
  - **How to exploit a simple translator ?**
  - Text based retrieval filtered by VCDT
- WikipediaMM Task (Wikipedia)

Visual Concept Detection Task



- 17 concepts (classes) - 2K training images – 1K for test
- Learning perspective: multi-class and multi-label problem.
- Concepts are presented in a simple hierarchy

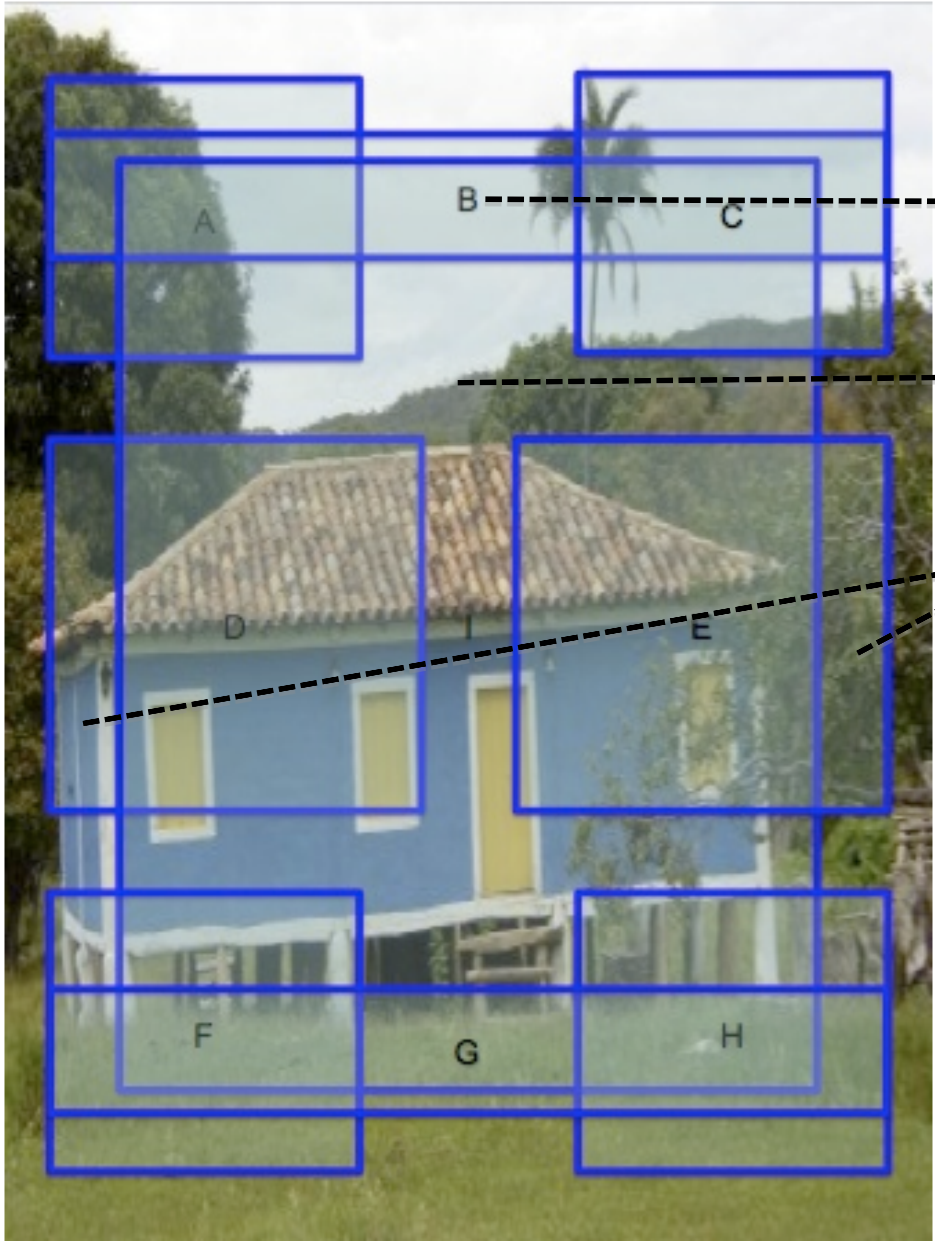
FFDT Learning diagram



Fuzzy Decision Tree

Salammbô by Christophe Marsala, 1998

Visual descriptors:
image split into 9 zones



colors: HSV 4x3x3

colors: HSV 8x3x3

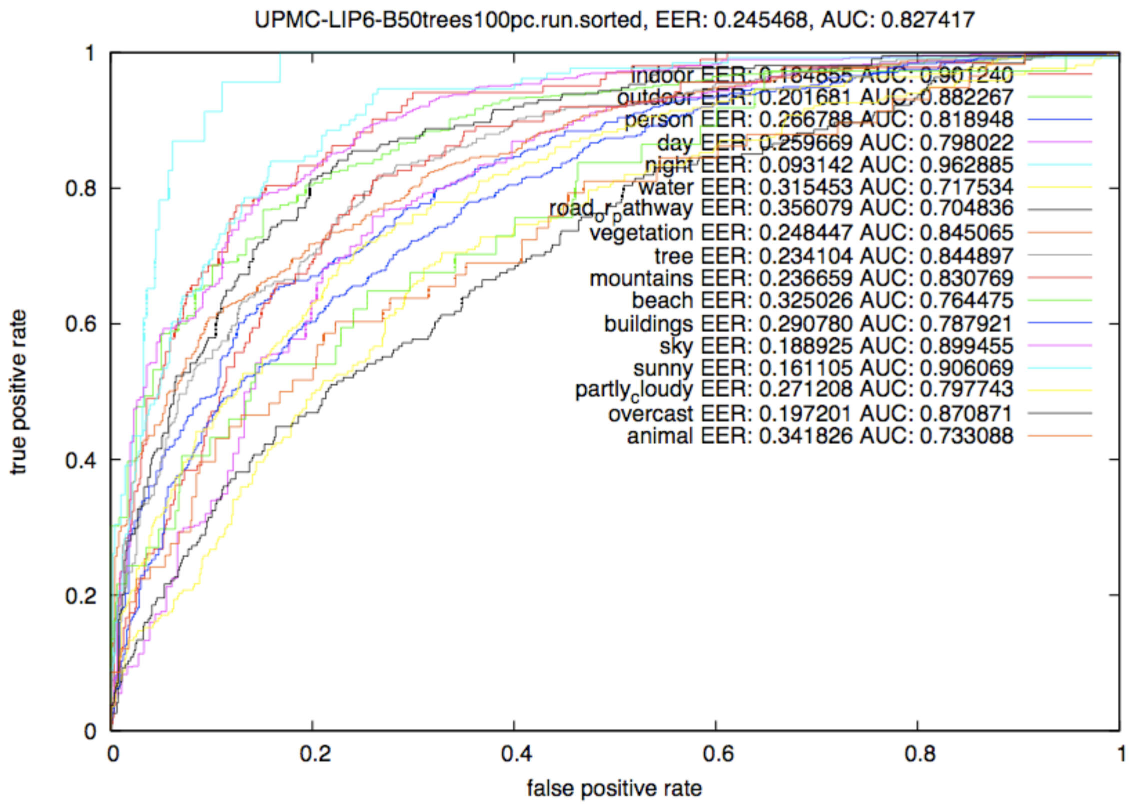
Color difference :
HSV 4x3x3

unique
Histogram

Fuzzy Decision Trees

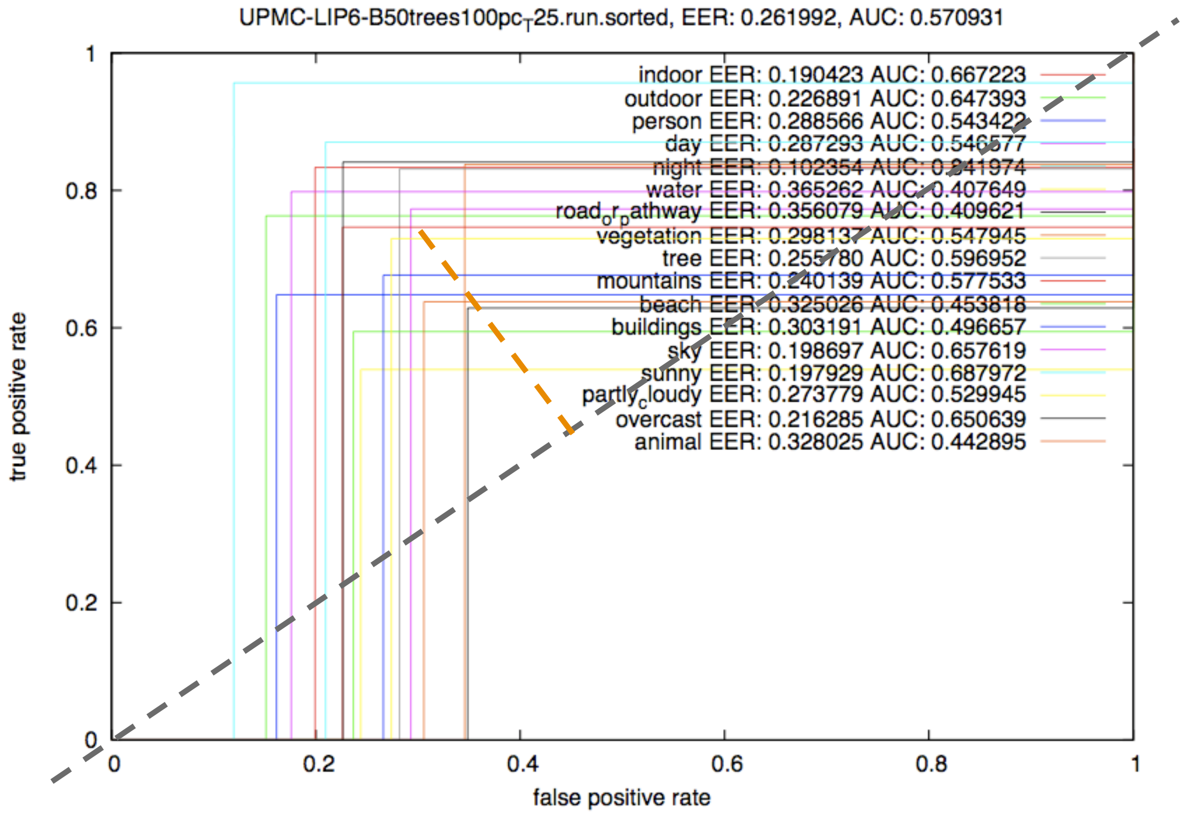
Official Results VCDT

- 53 runs for 11 teams (LIP6, LSIS, CEA-LIST, XRCE, IPAL…)
- Equal Error Rate 24.6% - AUC: 82.7% => 3rd of 11

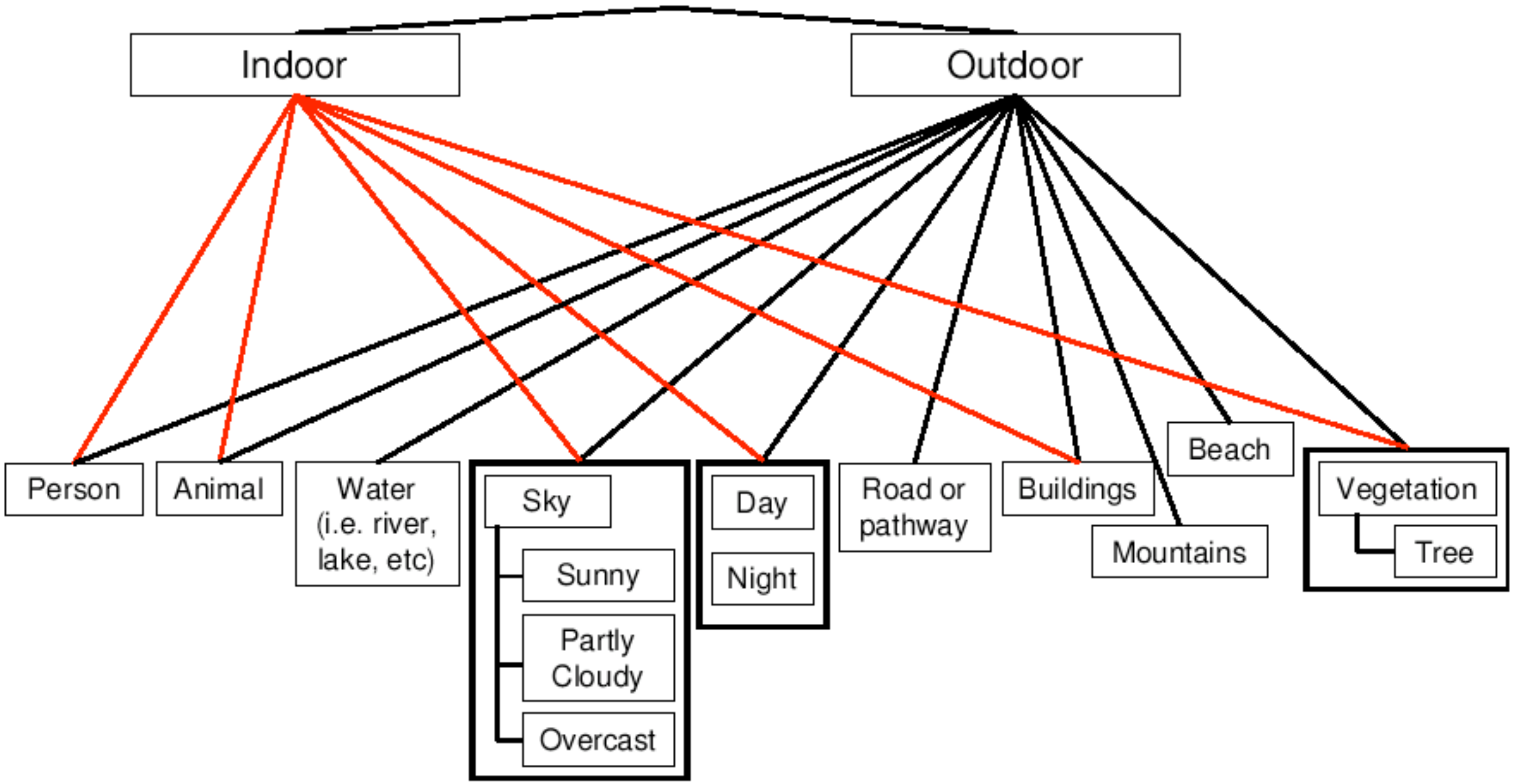


EER => Normalized Score (NS)

- EER is not adapted when we have a class decision
- Normalized Score (NS)



Concept are related (theory)

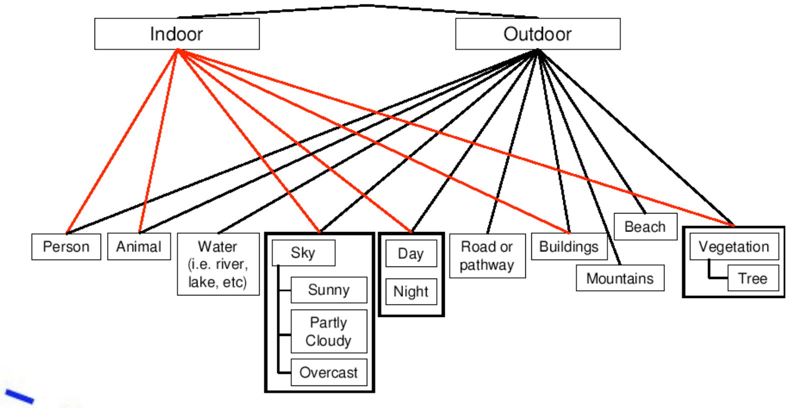


How we discover the relations ?

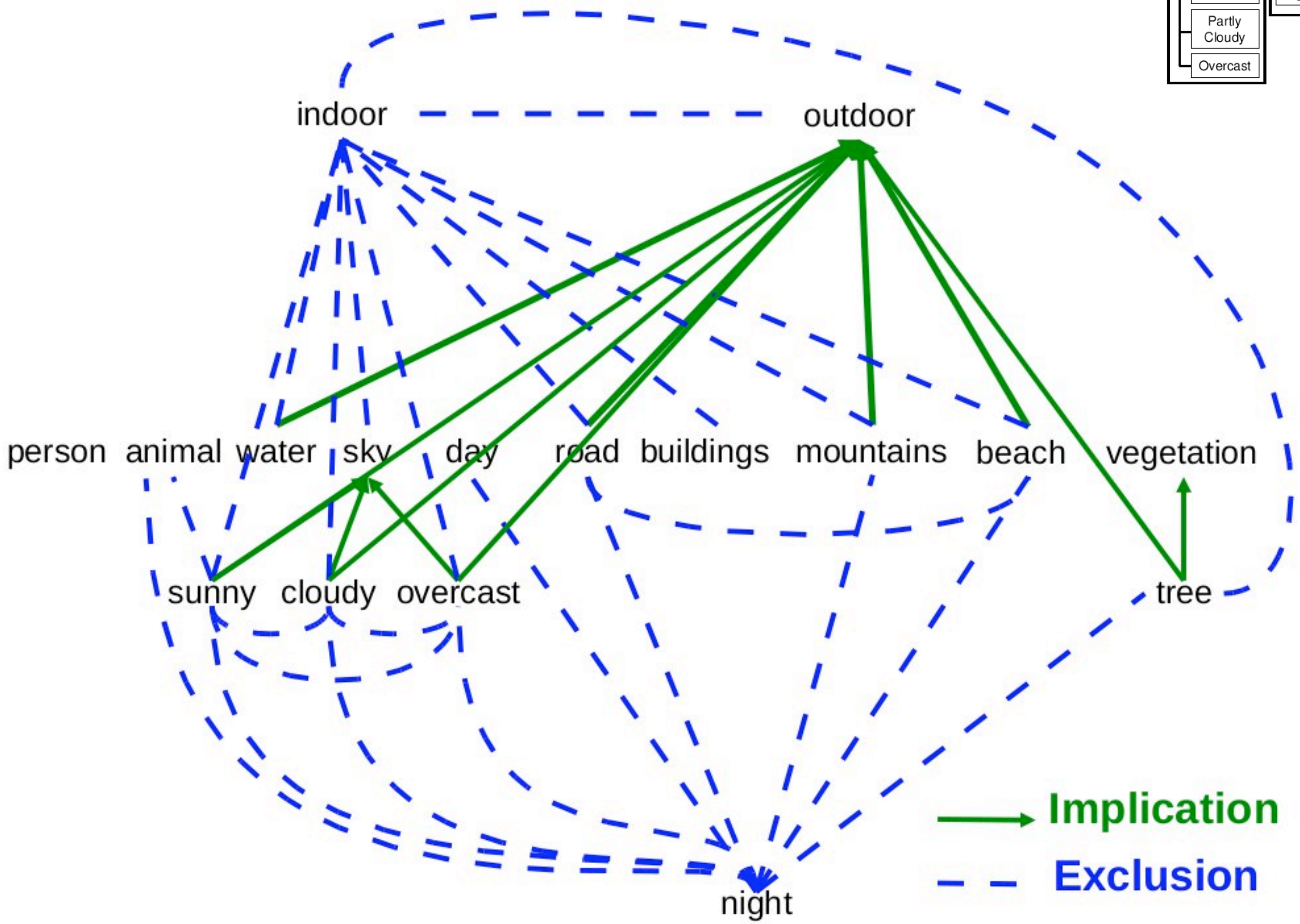
- EXCLUSION
  - We study the co-occurrence matrix
  - When two concepts **never** appear together -> exclusion

- IMPLICATION (or necessity)
  - (A=>B) is equivalent to (not B or A)
  - We build a matrix of presence vs. absence (COOCNEG)
  - When **never** (B and not A) -> (not B or A) -> A => B

We found …



Indoor
Outdoor
Person
Animal
Water
(i.e. river,
lake, etc)
Sky
Sunny
Partly
Cloudy
Overcast
Day
Night
Road or
pathway
Buildings
Beach
Mountains
Vegetation
Tree



indoor
outdoor
person
animal
water
sky
day
road
buildings
mountains
beach
vegetation
sunny
cloudy
overcast
tree
night
Implication
Exclusion

How to use exclusion and implication ?

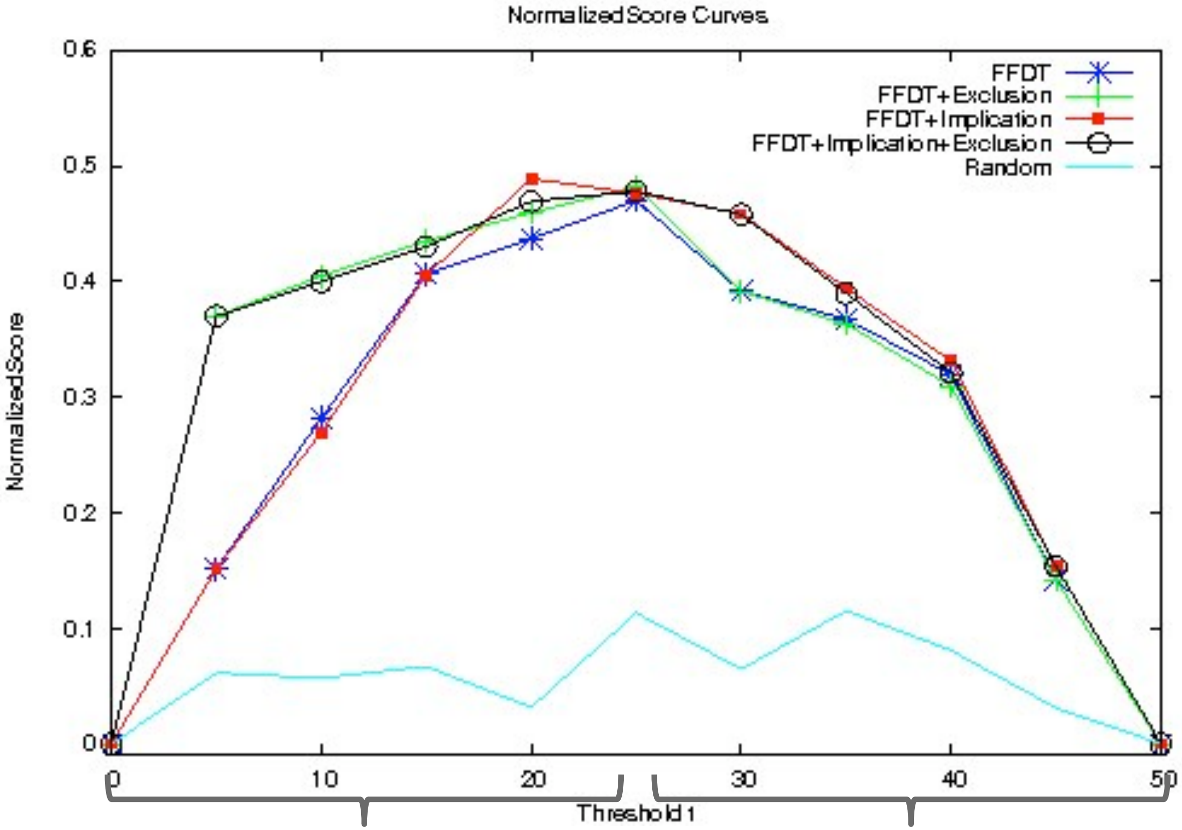
- Exclusion
  - COOC(A,B)≈0
  - before: Outdoor=0.8 and Indoor=0.5
  - after: Outdoor=0.8 and Indoor=0
  - IF score(I,A)>score(I,B) THEN score(I,B)=0 ELSE score(I,A)=0
- Implication
  - COOC(A,non B)≈0
  - before: Cloudy=0.8 and Sky=0.5
  - after: Cloudy=0.8 and Sky=0.8
  - score(I,B)=max(score(I,A),score(I,B))

Results (Normalized Score)

	Exclusion	Implication	NS
FFDT			0.470
	X		0.482
		X	0.476
	X	X	0.478

- Exclusion and Implication **improve only slightly** the NS ?!

General behavior



Exclusion is better

Implication is better

Exclusion + Implication is always better

ImageCLEF 2008

- Visual Concept Detection Task (VCDT)
  - **How to “translate” image into text ?**
  - Forest of Fuzzy Decision Trees (FFDT)
- Photo Retrieval Task (PHOTO)
  - **How to exploit a simple translator ?**
  - Text based retrieval filtered by VCDT

Photo retrieval task

- 20K Images
  - the same collection as for VCDT
  - associated to a textual description
  - semi-structured : title, location, date, visual description, ...
- 39 Topics
  - Semi-structured textual description: <title>, <narr>, <cluster>
  - Image examples
- 2008 edition
  - focus on diversity
  - the measures were P20 and CR20

Photo retrieval : our approach

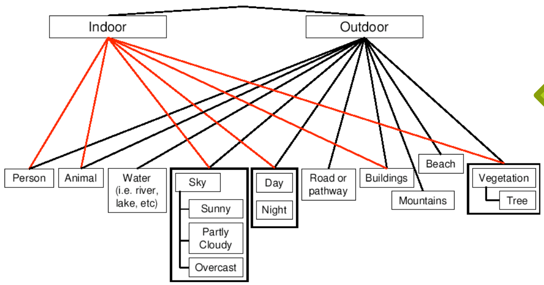
- Text retrieval
  - tdf-idf, language model (LM)
- From topics to queries
  - textual queries using title + narr + narr-"not".
- Filter the resulting ranked list by using the FFDT (VCDT)

Topic 58: “seals near water”

VCDT (FFDT): “water”

TOPIC 58 : seals near water
CLUSTER BY : country
TOPIC NARRATIVE : Relevant images will show seals (or more specific: fur seals, ear seals and sea lions) at a body of water (sea, lake, etc.). The water has to be visible for an image to be relevant. Images of seals with no water visible in the image are not relevant. Images of water but without seals are not relevant either.

Which VC should we use?



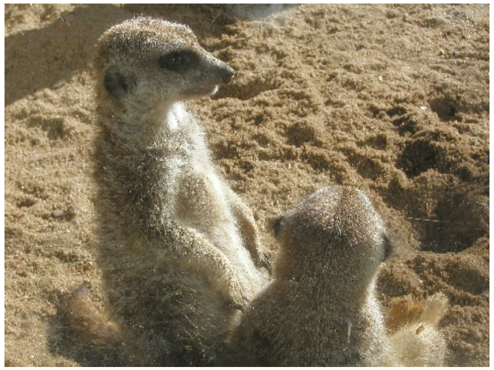
?

<title>church with more than two towers</title>
<cluster>city</cluster>
<narr>Relevant images will show a church, cathedral or a mosque with three or more towers. Churches with only one or two towers are not relevant. Buildings that are not churches, cathedrals or mosques are not relevant even if they have more than two towers.<narr>

- Method VCDT: find the “concept” in the <title>
- Method VCDTWN: find the “concept” in a list of synonyms (Wordnet) of the <title>

WordNet expansion

- TOPIC 5 : animal swimming
  - Animal: organism, plankton, mascot, fungus, …
  - Swimming: bathe, diving, floating, surfing, **water** sport, …
  - Use the VCDT-animal & VCDT-water



Rank 1

Rank 4

Rank 11

How we apply a filter on a list?

- After a text query: list of ranked images
  - “Good” images are highly ranked
  - Should we filter them out?
  - At what VCDT degree we decide that the concept is present?
- Since ImageClef focus on P20 and CR20
  - We filter the first the 50
  - And re-introduced them after rank 50.

“seals near water” – rank 1&2

Rank 1: water concept detected -> not filtered

Rank 2: water concept was not detected -> filtered to rank 50

“seals near water” – first error

Rank 3: water concept was not detected -> filtered to rank 51!

“seals near water” – rank 3 to …






"seals near water" – second error

Rank 20: water concept was detected -> not filtered!

“seals near water”

Rank 8: water concept was detected -> not filtered !!

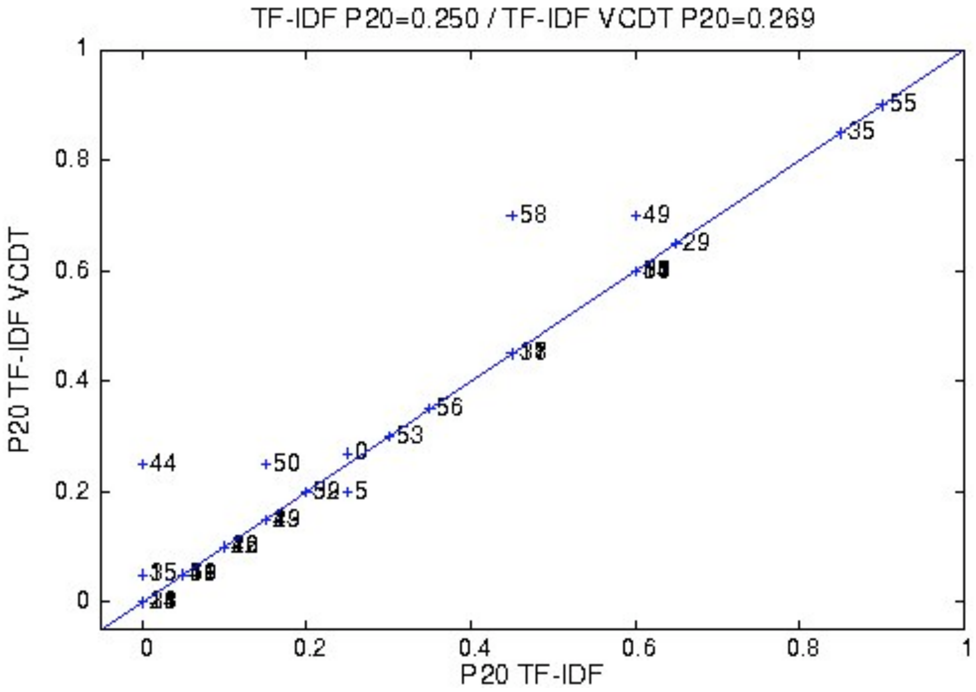
Was considered by ImageClef as WRONG !?

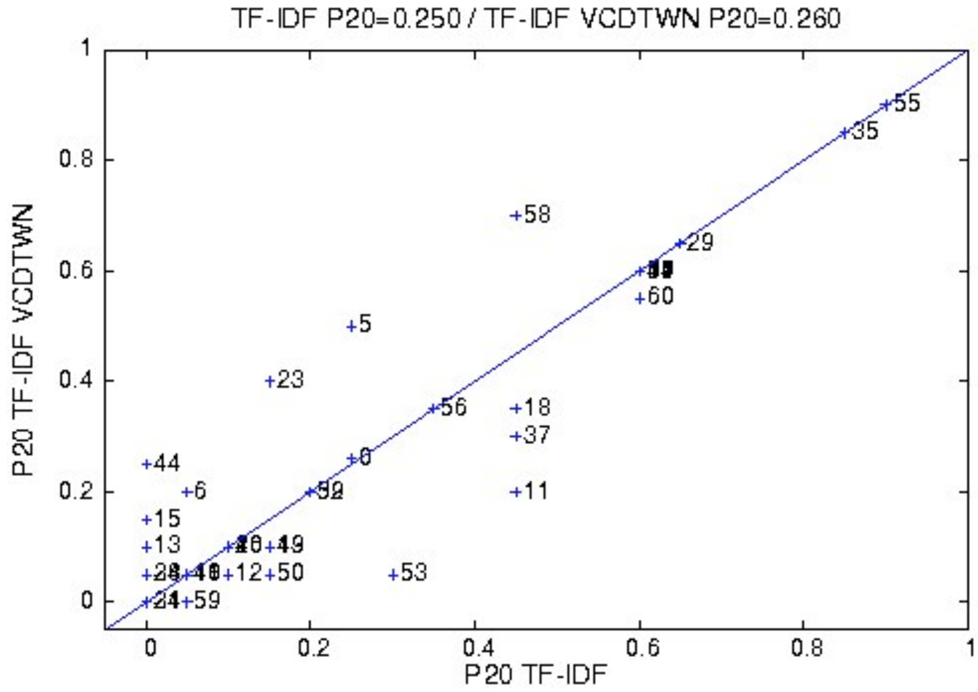
Results

Texte	Concept Filtering		All 39 topics		Only topics modified by filtering		
	VCDT	WN	P20 (gain %)	CR20 (gain %)	Nb of topics	P20 (gain %)	CR20 (gain %)
LM			0.185 (-)	0.247 (-)	11 25	0.041 (-) 0.148 (-)	0.090 (-) 0.254 (-)
	X		0.195(+6)	0.257(+4)	11	0.077 (+88)	0.126 (+40)
	X	X	0.176(-5)	0.248(+1)	25	0.134 (-9)	0.257 (+1)
TF -IDF			0.250 (-)	0.300 (-)	11 25	0.155 (-) 0.210 (-)	0.161 (-) 0.305 (-)
	X		0.269(+8)	0.313(+5)	11	**0.223 (+44)**	**0.209 (+30)**
	X	X	0.258(+4)	0.293(-2)	25	0.226 (+8)	0.294 (-4)

- VCDT - 11 topics modified and 7 VC were used
- VCDT+WN – 25 topics modified (some x Times) and 9 VC were used

VCDT filtering vs VCDT+WN





VCDT filtering always improved the Precision at 20

Conclusion

- Some **Visual Concepts** can be learned
  - Exploiting relations between concepts only slightly improve the results
  - Small simple “image2concept translator” is possible
- Text-based **image retrieval** can benefit from such a translator
  - For some queries strong improvement in terms of P20
  - Difficulty: how to detect in the query (text) the right visual concept?

Other results… diversity

- Why diversity is interesting?
- UPMC/LIP6
  - Visual diversity based on pre-segmenting the colors space
  - Slight improvement of CR20
- AVEIR
  - Diversity by the fusion of different runs (teams)
  - Several fusion strategies were compared
  - P20: AVEIR better than best individual - CR20: AVEIR close.
  - 3rd best team at ImageClefPhoto

http://aveir.lip6.fr

Thank you for your attention

Marcin.Detyniecki@lip6.fr