SZTAKI @ ImageCLEF 2008

Bálint Daróczy

joint work with

András Benczúr, Mátyás Brendel, Zsolt Fekete,
Attila Pereszlényi, Simon Rácz, Dávid Siklósi

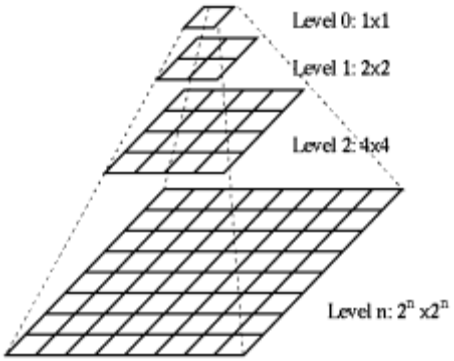
Data Mining and Web Search Group
Computer and Automation Research Institute
Hungarian Academy of Sciences

Overview

- Visual features, similarity measures used
  - Segmentation procedure
- Visual Concept Detection task
  - Comparison of global features vs. Segmentation
- WikimediaMM and ImageCLEF-photo
  - Similarities and differences
  - Is ImageCLEF really about images?

Image retrieval 1: segmentation

Pre-segmentation

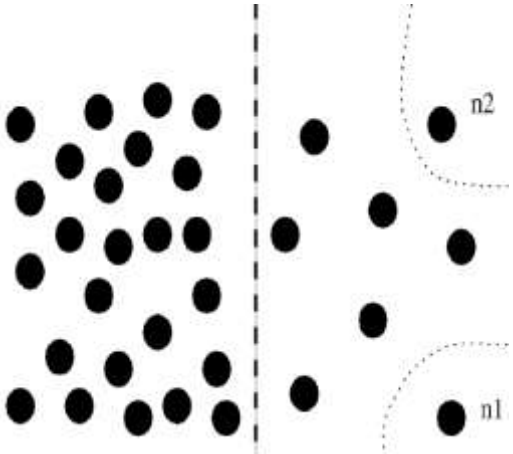


- Resize to 500x500 (OpenCV)
- Smooth to eliminate noise (OpenCV)
- Downsizing with Gaussian kernel

(Three-level Gaussian-Laplacian pyramid)

- Intra and inter level threshold for joining pixels

Graph-based method



- Undirected weighted graph over neighboring pixels
- Bottom-up clustering with dynamic thresholds
- Efficient heuristic solution, better than min-cut, close to normalized cut

[Felzenszwalb, Huttenlocher]

- Sobel gradient image to select important edges

Image retrieval 2: Feature vector

common for all:

- average RGB, RGB histogram, contrast

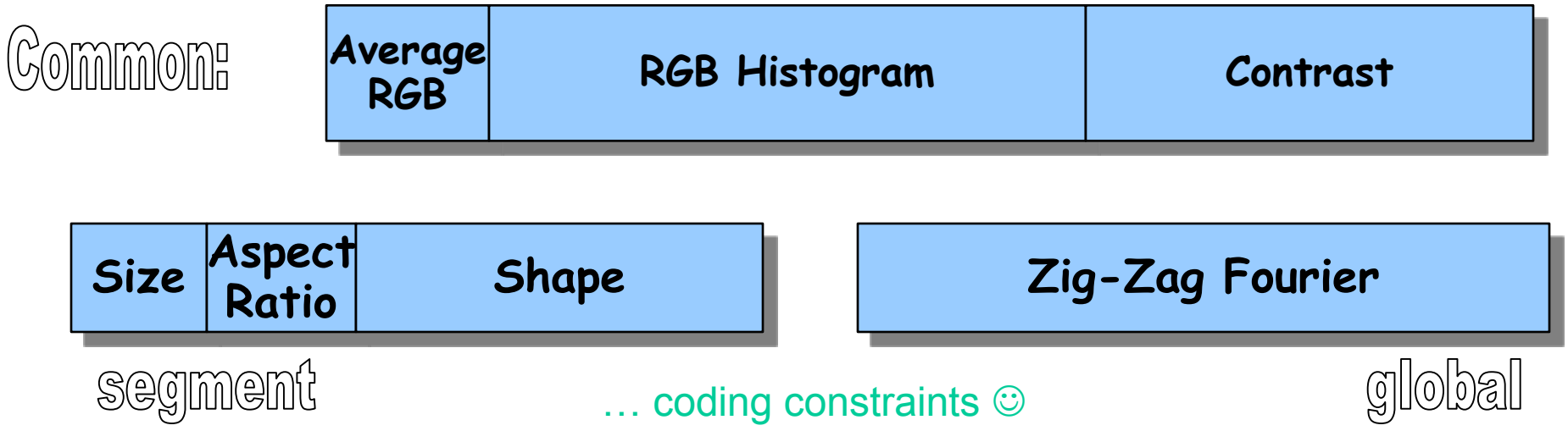
global:

- DFT values (zig-zag)

segments:

- shape, size, aspect ratio

Overall 91 (segment) and 176 (global) values.
Weighting of features is important.



Visual Concept Detection

- Two segmentation granularity (100< small, 100> large)
  - Minimal segment size threshold
  - Felzenbach-Huttenlocher cut parameter
- global AND segment based features
- Filtering segments over training data
  - ***relabeling*** false negative samples
  - Ignoring ***all false*** classified samples
  - Ignoring ***false positive*** samples
- classification: logisitic regression
- Classifier combination

Results

Visual Concept	EER	AUC
Large, no filter w/o DFT	32.10	74.18
Large, relabel w/o DFT	32.47	73.57
Large, all false w/o DFT	37.01	59.30
Large, false pos w/o DFT	32.47	73.61
Small, no filter w/o DFT	32.44	73.32
Small, relabel w/o DFT	32.48	73.03
Small, all false w/o DFT	36.07	67.15
Small, false pos w/o DFT	32.46	73.05
Global, w/o DFT	45.72	52.78
Global, with DFT	31.14	**74.90**
Combination	**29.92**	72.77

Conclusions for the visual concept task

- "Global" classes and "local" classes = object

EER/AUC	Night	Animal	Vegetation
Best global	10.85 / 94.86	39.07 / 63.58	36.23 / 68.52
Best local	30.30 / 80.01	32.80 / 74.87	31.26 / 75.22

- More positive cases this year in the test set, better -> We liked this task!
- But suggestion
  - Objects located on the image in the training set (like last year) would be better
  - One object per image in training set would be better

Wikimedia and Photo

Text Based Image Retrieval

- Okapi BM25
- query expansion (under progress…)
- KL dissimilarity to increase cluster recall

Two types of Content Based Image Retrieval

- CBIR: average distance between segments
- SCBIR: image similarity by novel Segment-Conformation Based method

Linear combination of textual and visual results

Image retrieval 3: measures

Similarity between segmented images and the query image

This slide is not ready yet

Measure of similarity of objects is defined in a certain feature-space.

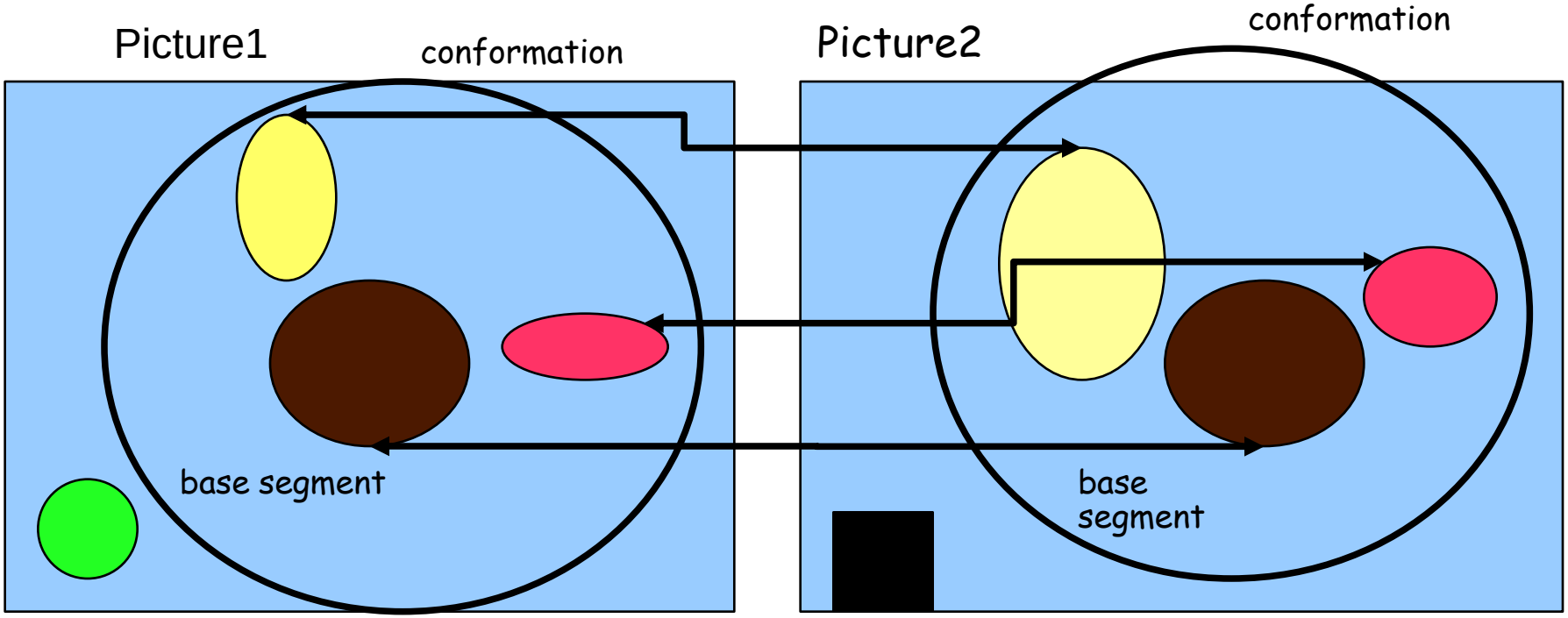
$$dist(S_i, O_j) = d(F(S_i), F(O_j)) : S_i \in S, O_j \in O$$

Euclidan distance to measure similarity between images.

$$dist(Q) = min_{i,j} dist(S_i, O_j) : S_i \in S, O_j \in O$$

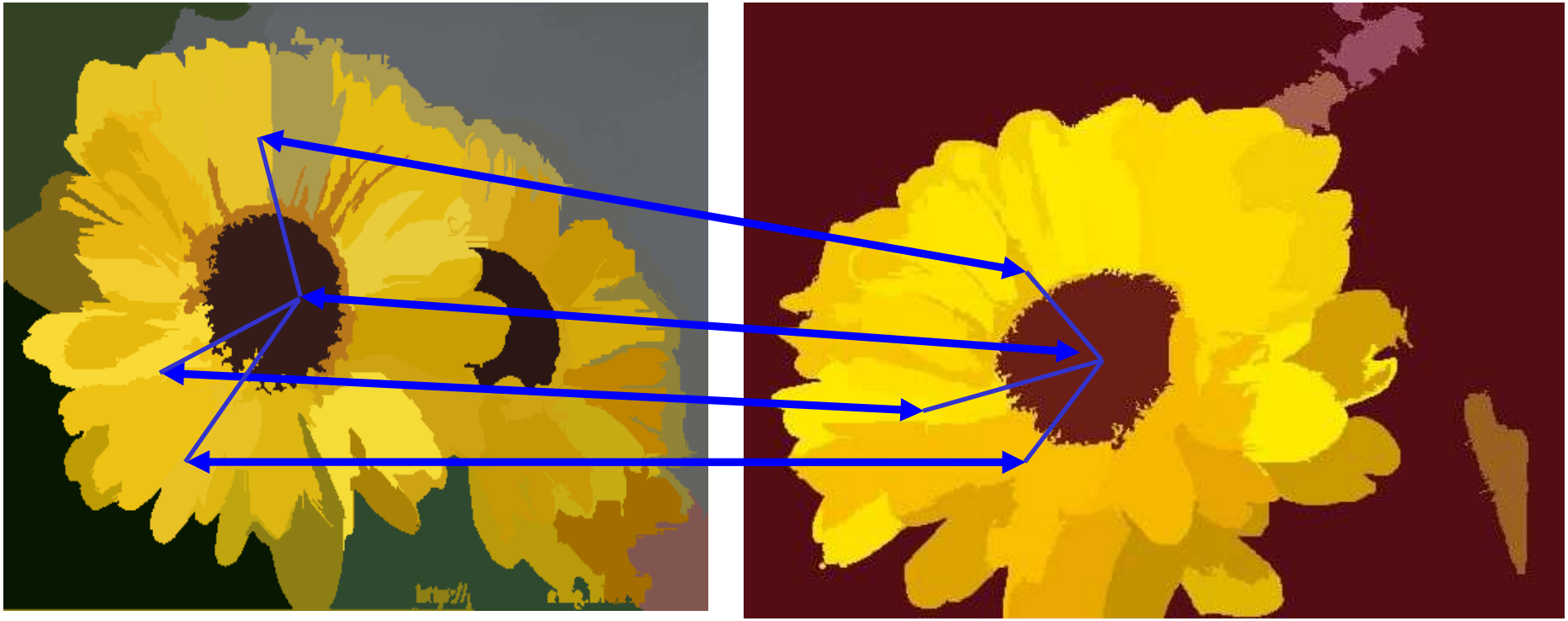
Segment-Conformation Based image retrieval

Find similar conformation of segments



Segment-Conformation Based image retrieval

Find similar conformation of segments



Results

Results (MAP)	Photo	WikiMedia
TBIR	0.2956	0.2551
CBIR	0.0381	0.0000
TBIR+CBIR	0.2981	0.2526
TBIR+QE	0.2988	0.2546
TBIR+CBIR+QE	0.3014	0.2514
SCBIR	0.0404	0.0000
SCBIR+TBIR	0.2984	0.2491
SCBIR+TBIR+QE	**0.3030**	0.2465

Some results of other teams

cea		
TXTCON	QE	0.2763
TXT	QE	0.2632
Chemnitz		
TXTIMGCON	QE	0.2195
TXT	NOFB	0.2166
TXTIMG	NOFB	0.2138
TXTIMGCON	NOFB	0.2048
curien		
TXT	NOFB	0.2453
TXTIMG	FB	0.1161
imperial		
TXT	NOFB	0.1918
TXTIMG	NOFB	0.1225
upmc		
TXT	NOFB	0.1193
TXTIMG	NOFB	0.1050
upeking		
TXT	QE	0.3444
TXTIMG	QE	0.1912
utoulon		
TXT	NOFB	0.0399
TXTIMG	NOFB	0.0296
ualicante		
TXT	NOFB	0.2700
TXTCON	NOFB	0.2587

Conlcusions for the Wikimedia task

- sample images given only for 40 topics out of 75
- sample images not representative visually (eg. Bush
- Only CEA used concept with moderate success
- CBIR could not improve TBIR significantly

- images were diverse in size, quality and content, this is makes things hard, but realistic
- database already large for brute force pairwise similarities, but still small for testing scalability -> larger databases ?

Conclusions for photo task

- Topics and clusters were mainly based on location, not really suitable for CBIR
- Cluster recall could be improved by textual content but not by visual features
- We did not make use of the result of visual concept (did anybody?)

Suggestions:

- Less location centric queries (and clusters)
- No topics as e.g. „stadium outside Australia"
- More than 3 sample images
- Image similarity centric topics