UvA-ISIS@ImageCLEF
photo annotation & concept-based retrieval

Koen van de Sande
Cees Snoek

University of Amsterdam
ImageCLEF 2011 - September 21st

Photo Annotation

- Annotation task
  - For 99 concepts, rank 10,000 photos
  - Concept detection system
  - Modalities: Visual & tags
- Concept-based retrieval task
  - For 40 topics, rank 200,000 photos
  - Can use existing concept detectors … or train new ones
  - Modality: Visual only

2

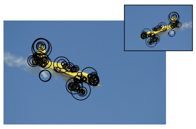
Concept detection system

3

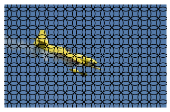
Image Feature Extraction: Point Sampling Strategies

Lowe IJCV 2004
Mikolajczyk IJCV 2005
Zhang IJCV 2007
Marszalek VOC 2007

- Orientation and scale of object changes
- Salient point methods robustly detect regions



Harris-Laplace



Dense sampling

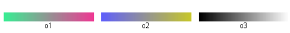
- Preferred for visual categorisation accuracy are
  - Harris-Laplace salient points
  - Dense sampling

4

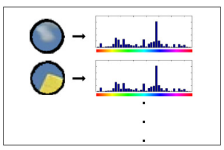
Invariant Visual Descriptors

Lowe IJCV 2004
van de Sande PAMI 2010

Color SIFT:
- Intensity-based SIFT
- OpponentSIFT
- C-SIFT
- *rg*SIFT
- RGB-SIFT



Add color, but also keep intensity information



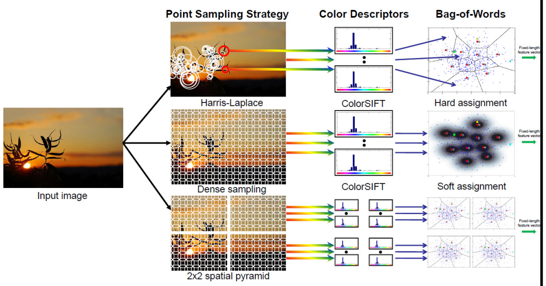
5

Example: light color change

RGB-SIFT descriptor is invariant

6

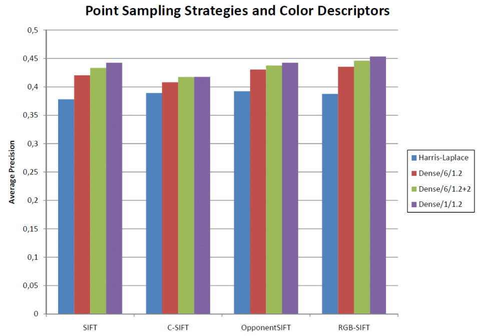
Visual Feature Overview



7

ImageCLEF book chapter 18

Experimental results



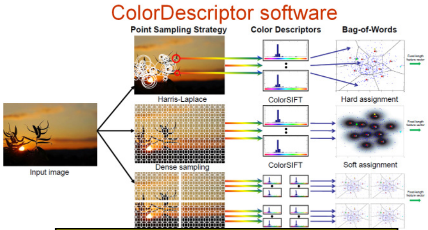
8

van de Sande PAMI 2010
van de Sande TMM 2011

ColorDescriptor Software

- Today at v3.0
- GPU-accelerated version (CUDA) for 64-bit Windows: end of next month

ColorDescriptor software



Visit http://www.colordescriptors.com

9

Sonnenburg JMLR 2010
Maji CVPR 2008
Perronnin ECCV 2010
Zhou ECCV 2010

Train concept model

- SVM with nonlinear kernel
  - χ^2 kernel
  - Histogram intersection kernel
- Fast to apply with intersection kernel
- Other approaches work with linear kernels (Fisher vector, supervector), but need much longer feature vectors

10

Annotation task

in 1 slide

Team	MiAP	Config.	Team	F-Ex	Config.	Team	SR-Prec.	Config.
TUBFI	0.443	M	ISIS	0.622	M	ISIS	0.742	M
LIRIS	0.437	M	CAEN	0.600	V	BPACAD	0.729	V
BPACAD	0.436	M	BPACAD	0.593	M	CAEN	0.727	V
ISIS	0.433	M	HHI	0.588	V	LIRIS	0.725	V
MLKD	0.402	M	LIRIS	0.576	M	HHI	0.718	V
CEALIST	0.384	M	TUBFI	0.566	M	IDMT	0.713	M

Selected 1 likelihood threshold for all concepts

11

Concept-based retrieval

12

Concept-based retrieval

- 40 topics
- 5 examples per topic
- 200,000 images in test set

What to do?

- More discriminative features? Keep features fixed
- More powerful classifiers? Keep classifier fixed
- Better training examples? Experiment with training & retrieval method

visual-only

13

Approach 1: Fully automatic

- Take 5 positive example images
- Take 10, 33 or 100 random negatives
- Train new model for topic

Topic: Airplane in the Sky

We'd like to find photos of real airplanes flying in the sky. Airplanes on the ground are not relevant, neither are airplanes pictured from the inside. Small models of airplanes are also not relevant.

Examples:

14

Results 1: Fully automatic

Topic	Fully automatic auto10	auto33	auto100
1 Graffiti on buildings/walls	0,017	0,074	0,062
2 Toy vehicle	0,001	0,000	0,002
3 1 person doing sports at sea	**0,123**	0,057	0,044
4 Airplane in the sky	0,045	0,000	0,024
28 Fireworks	0,000	0,384	0,415
29 Close-up of flowers with raindrops	0,000	0,000	0,000
30 Cute toys arranged as still-life	0,000	0,002	0,002
31 Ship/boat on a river	0,004	0,003	0,003
32 Underexposed photos of animals	0,001	0,000	0,000
33 Cars and motion blur	0,000	0,128	0,108
40 Close-up of bodypart	0,011	0,017	0,022
MAP	**0,018**	**0,037**	**0,043**
#pos	5	5	5
#neg	10	33	100

•Should have used >100 random negatives?
•Our overall best for 3/40 topics

15

Approach 2: Human topic mapping

- Topics are often Boolean combinations of existing concepts
  ➔ Use 99 concept detectors from the annotation task
- Parsing the relations between the concepts from the topic text is challenging
  ➔ Let a human pick 1 or 2 concepts per topic, fusion of classification scores

16

Results 2: Human topic mapping

Topic	Human topic mapping Concept #1	Concept #2	1concept	2concept **x**	2concept **+**
1 Graffiti on buildings/walls	Graffiti	Building_Sight	0,082	0,022	0,000
2 Toy vehicle	car	Toy	0,000	**0,023**	0,003
3 1 person doing sports at sea	Sea	Single_Person	0,002	0,002	0,002
4 Airplane in the sky	airplane	Sky	0,125	**0,185**	0,166
28 Fireworks	Night	Outdoor	0,006	0,005	0,005
29 Close-up of flowers with raindrops	Flowers	Rain	0,002	**0,011**	0,002
30 Cute toys arranged as still-life	Toy	Still_Life	**0,155**	0,108	0,101
31 Ship/boat on a river	ship	River	0,002	**0,025**	0,017
32 Underexposed photos of animals	Animals	Underexposed	0,025	0,035	0,036
33 Cars and motion blur	car	Motion_Blur	0,009	**0,522**	0,453
40 Close-up of bodypart	bodypart		0,207	0,207	0,207
MAP			**0,053**	**0,089**	**0,080**

•Didn't use negation, only 'AND'
•Works only if there are related concepts
•Our overall best for 21/40 topics

17

Approach 3: Human topic inspection

- Start with 5 given positive images
- Apply automatic model on the annotation task train set (8,000 images)
- Human adds more examples & retrains
- At most 7,5 minutes per topic
- Increases #pos from 5 to ~42
- We never look at the test set

18

Results 3: Human topic inspection

Topic	Human topic inspection normal	+100neg
1 Graffiti on buildings/walls	0,184	**0,253**
2 Toy vehicle	0,000	0,000
3 1 person doing sports at sea	0,007	0,008
4 Airplane in the sky	0,051	0,091
28 Fireworks	0,404	**0,423**
29 Close-up of flowers with raindrops	0,005	0,006
30 Cute toys arranged as still-life	0,044	0,060
31 Ship/boat on a river	0,012	0,024
32 Underexposed photos of animals	**0,066**	0,065
33 Cars and motion blur	0,340	0,345
40 Close-up of bodypart	0,230	**0,261**
MAP	**0,092**	**0,100**
#pos (on average)	42	42
#neg (on average)	228	328

- **Topic example images give starting point**
- **Build a new ‘concept’ on the fly**
- **Our overall best for 17/40 topics**

19

Demo

#28. Fireworks. **We like to find photos of fireworks at night in outdoor images**

#15. Sea sunset or sunrise. **We like to find pictures of a sunrise or sunset over the sea. So sunrises over cities or over lakes are not relevant. If the picture shows persons or vehicles it is also not relevant**

20

Conclusions

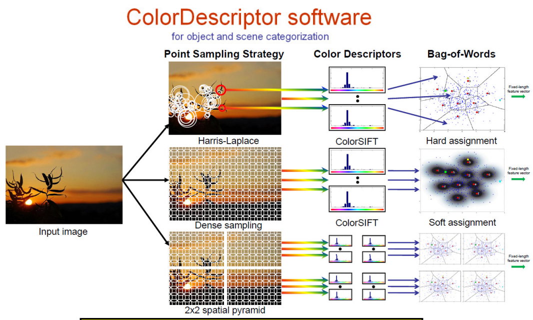
- Annotation task
  - Focus on feature representations, machine learning algorithms
- Concept-based retrieval is a new and different task
  - Focus on training approaches, using proven visual features

- Lessons learned
  - Should have used the text tags
  - For ~50% of topics, selecting 1-2 ‘relevant’ concepts works
  - Should support complex relations: ‘A but not B’
  - If approach 1 & 2 don’t work, difficult to bootstrap 3… unless you would exploit tags

21

van de Sande PAMI 2010
van de Sande TMM 2011

ColorDescriptor Software



Visit http://www.colordescriptors.com

22