Hierarchical Annotation of Large Image Collections

Mark Huiskes and Michael Lew

LIACS, Leiden University

Theseus/ImageCLEF workshop on visual information retrieval evaluation

Corfu, Greece, 29 Sep 2009

Outline

(Manual) annotation of the MIRFLICKR-25k collection

Goal

building a benchmark for evaluation of image retrieval systems based on relevance feedback

- Requirements
- Annotation method
- Initial results

by Mariah Michelle

by Wei Kin

by Chris Tarnawski

by Petar Kurschner

MIRFLICKR-25000

http://press.liacs.nl/mirflickr

open collection of 25000 *high-quality* images for image retrieval benchmarking

- downloaded from Flickr using their public API
- Creative Commons license
- high interestingness score
- Flickr tags (1386 tags with more than 20 images)
- EXIF metadata

Tag	# Images	Tag	# Images
sky	845	people	330
water	641	city/urban	308/247
portrait	623	sea	301
night	621	sun	290
nature	596	girl	262
sunset	585	snow	256
clouds	558	food	225
flower/flowers	510/351	bird	218
beach	407	sign	214
landscape	385	car	212
street	383	lake	199
dog	372	building	188
architecture	354	river	175
graffiti/streetart	335/184	baby	167
tree/trees	331/245	animal	164

Benchmarking Relevance Feedback

- Small sample size

- Large influence of image representation quality on performance

- Dual role for ground truth
  - Measuring retrieval accuracy
  - Simulating user behavior

Goals

1. Full topic annotations

   – partial topic annotation: topic annotation by a single annotator for a subset of the image collection
   – full topic annotation: topic annotation by a single annotator for entire image collection

2. Factor out representation quality before averaging performance results over different topics

Hierarchical Annotation

- Goal is to achieve full annotations at reasonable cost

- annotation set: set of images that has to be considered to annotate a certain topic

- Main idea: make the annotation set as small as possible without losing annotation quality

- Reduction of annotation set by stepwise refinement along two dimensions

  1. Abstraction level: from general to specific categories

  2. Relevance level: from (at least) weakly relevant to strongly relevant to the topic

General topic	Subtopics
sky	clouds
water	sea/ocean, river, lake
people	portrait, boy/man, girl/woman, baby
night	
plant life*	tree, flower
animals	dog, bird
man-built structures*	architecture, building, house, city/urban, bridge, road/street
sunset	
indoor	
transport*	car

Hierarchical Annotation

- Goal is to achieve full annotations at reasonable cost

- annotation set: set of images that has to be considered to annotate a certain topic

- Main idea: make the annotation set as small as possible without losing annotation quality

- Reduction of annotation set by stepwise refinement along two dimensions

  1. Abstraction level: from general to specific categories

  2. Relevance level: from (at least) weakly relevant to strongly relevant to the topic

'Pre-annotations': at least weakly relevant images
include any image that may potentially be relevant to the topic

by Gabriel de Andrade Fernandes

by David Sawyer

by Nebelkerze

'real' annotations: strongly relevant to the topic
include only images relevant for subjective interpretation of topic

by Edward Leger

by Pedro Pinheiro

by Philipp Klinger

Procedure

1. Pre-annotation stage

   Going down the semantic hierarchy, annotate topics for weak relevance

   Result: pre-annotations (or: potential labels)

   Small number of annotators, 'objective', hard work- but only once

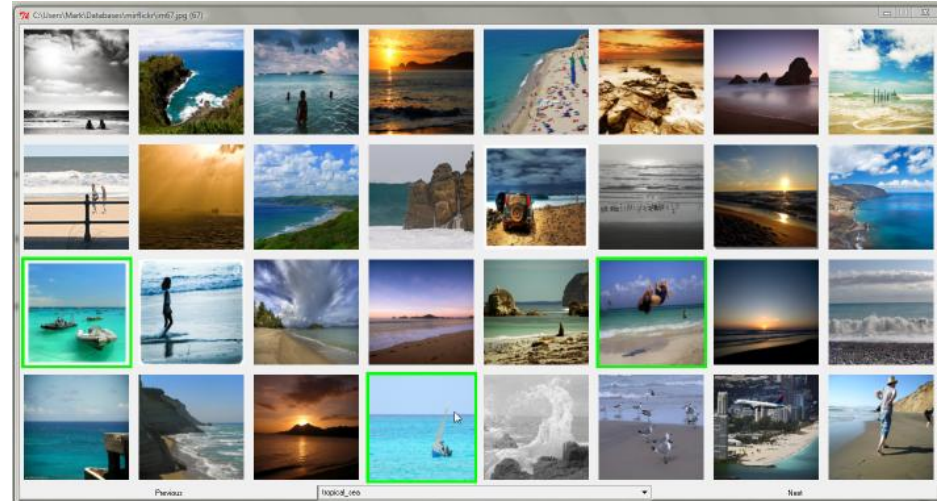2. Annotation stage

   Annotate personal interpretation of topics

   Many annotators, light work, one annotator per interpretation, subjective

Annotation/Tag	# Flickr Tags	# Pre-annotations	# Relevant Labels
clouds	558	3700	1350
sea/ocean	301	1322	214
river	175	894	149
flower	510	1823	1077
dog	372	684	590
bird	218	743	484
car	212	1177	380
baby	167	259	116

tag.pl

http://github.org/huiskes/tag.pl

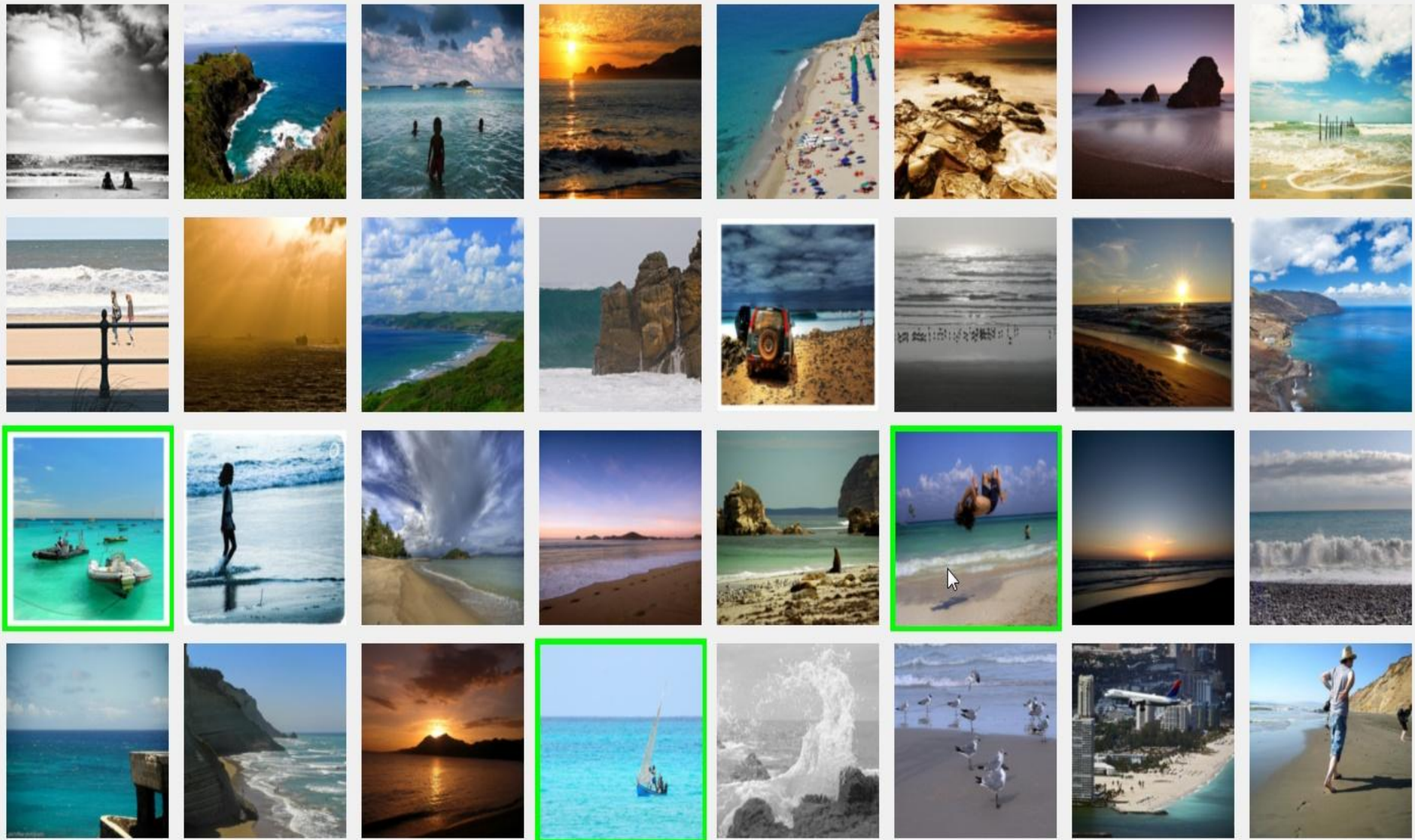**C:\Users\Mark\Databases\mirflickr\im242.jpg (242)**

Tag

```
A                : toggle autoforward
+                : increase autoforward speed
-                : decrease autoforward speed
q                : quit
- tagging
1-9 + extra keys : fast tagging
T                : tag (for tag id's >9)
d                : delete tag
s                : show image tags
- extra keys
r: street         j: house          u: urban
b: building       f: building house   i: bridge
>u
TAG ====> Updated tagfile for tag 5 (urban), image: 242
>
```

Features

1. HMMD Color Histogram descriptor
2. Spatial Color Mode descriptor
3. MPEG-7 Edge Histogram descriptor
4. MPEG-7 Homogeneous Texture descriptor
5. Flickr tags. 293 binary features for Flickr visual concept tags (>50 images per tag)

Results

		(a)						
	sky	water	night	people	plant life	animals	structures	sunset
MAP	80.8	56.7	63.0	73.4	69.8	52.3	71.5	52.3
prec@50	98	90	92	78	92	90	86	81
	indoor	transport	clouds	sea	river	lake	portrait	male
MAP	66.8	40.8	65.2	46.6	33.4	26.1	54.7	43.0
prec@50	66	72	90	84	80	60	80	61
	female	baby	tree	flower	dog	bird	car	
MAP	50.3	32.1	52.0	55.8	63.0	39.2	27.8	
prec@50	68	70	84	94	98	92	68	

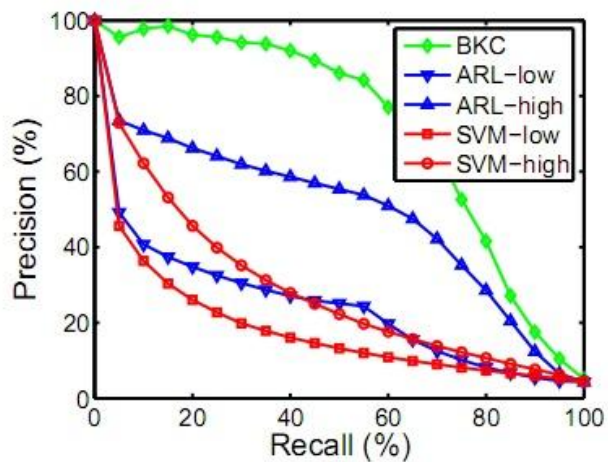|		(b)						
	sky	water	night	people	plant life	animals	structures	sunset
MAP	82.6	58.5	63.4	71.9	76.2	35.2	69.0	63.8
prec@50	100	100	98	100	100	80	90	96
	indoor	transport	clouds	sea	river	lake	portrait	male
MAP	61.8	40.0	67.3	37.9	17.0	17.9	53.3	40.2
prec@50	84	90	100	68	32	24	90	54
	female	baby	tree	flower	dog	bird	car	
MAP	52.2	21.9	55.1	55.9	57.5	16.6	28.2	
prec@50	96	42	92	96	94	40	62	

Results
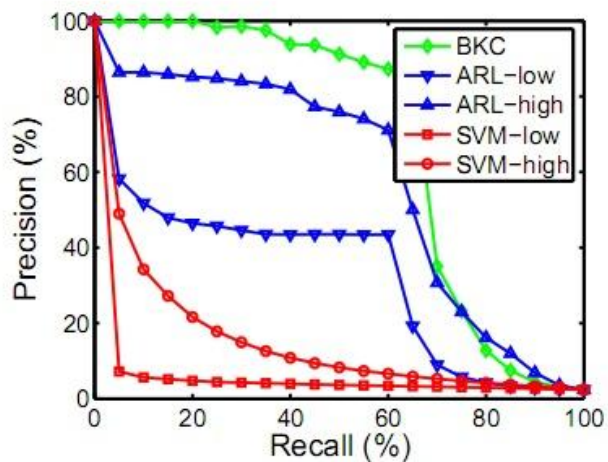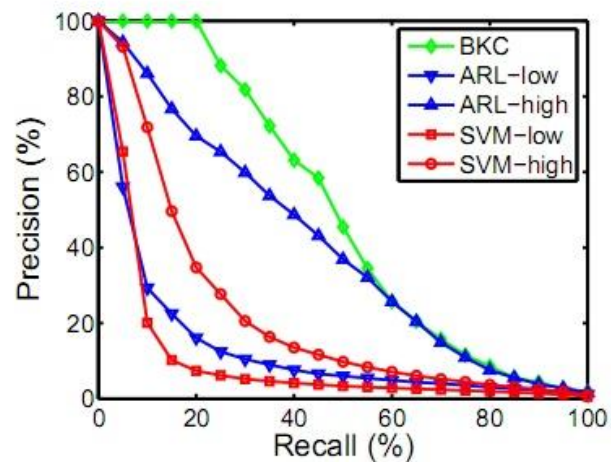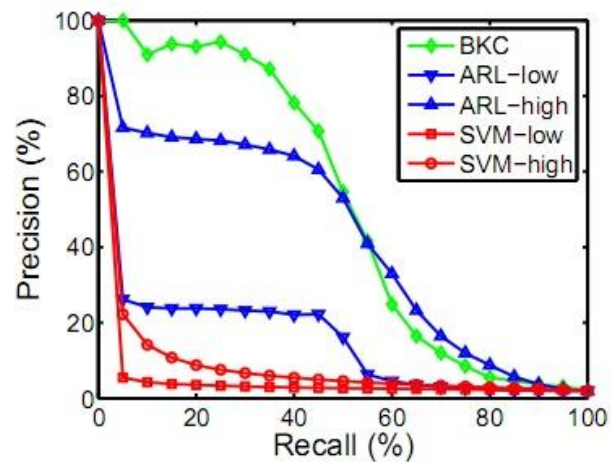

RF performance for 4 topics (a) dog, (b) dog on grass, (c) bird, (d) flower

Future work

- Extend number of topics
- Comparison of RF methods
- Comparison of MIRFLICKR and ImageCLEF annotations