LABORATORIO DE TECNOLOGÍAS DEL LENGUAJE

INSTITUTO NACIONAL DE ASTROFÍSICA, ÓPTICA Y ELECTRÓNICA

The SAIAPR TC12 Benchmark

http://imageclef.org/SIAPRdata

Hugo Jair Escalante, Manuel Montes &
Enrique Sucar*

September 29, 2009, Theseus / ImageCLEF workshop, Corfu, Greece

* In collaboration with Michael Grubinger

Contents

- The IAPR TC12 benchmark
- Extending the IAPR TC12
- Segmentation & annotation of IAPR TC12
- Statistics of the SAIAPR TC12 collection
- Baselines for region labeling
- Baselines for multimedia image retrieval

The IAPR TC12 Benchmark

- An established image retrieval benchmark:
  - 20,000 images with free-text descriptions in English, German and Spanish
  - Ground-truth data for the evaluation of image retrieval
  - Used in ImageCLEFPhoto 2006, 2007 and 2008

Image ID: annotations/00/25.eng
Title: Plaza de Armas

Description: a yellow building with white columns in the background; two palm trees in front of the house; cars are parking in front of the house; a woman and a child are walking over the square;

Notes: The Plaza de Armas is one of the most visited places in Cochabamba. The locals are very proud of the colourful buildings.

Location: Cochabamba, Bolivia
Date: 1 February 2002
Originator: Michael Grubinger

The SAIAPR TC12 Benchmark

- We extended the IAPR TC12 collection by manually **S**egmenting every image and **A**nnotating each resultant region

- **The new stuff:**
  - 20,000 segmentation masks
  - 99,535 labeled regions (single-label)
  - 99, 535 labeled regions according to a conceptual hierarchy (multi-label)
  - Basic spatial relationships for each region in every image
  - Basic visual features were extracted from each region

The SAIAPR TC12 Benchmark

- **What for?**
  - The evaluation of automatic image annotation methods at both image-level and region-level
  - The evaluation of the impact of image annotation methods into the multimedia image retrieval task
  - The evaluation of methods for image segmentation and multiclass classification (both single-label and multi-label)

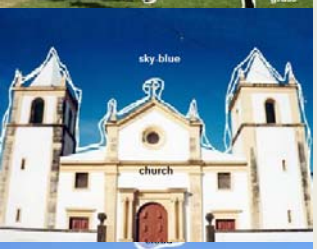
The SAIAPR TC12 Benchmark

- **Segmentation:** annotators[1] marked points around objects, which were joined by using splines

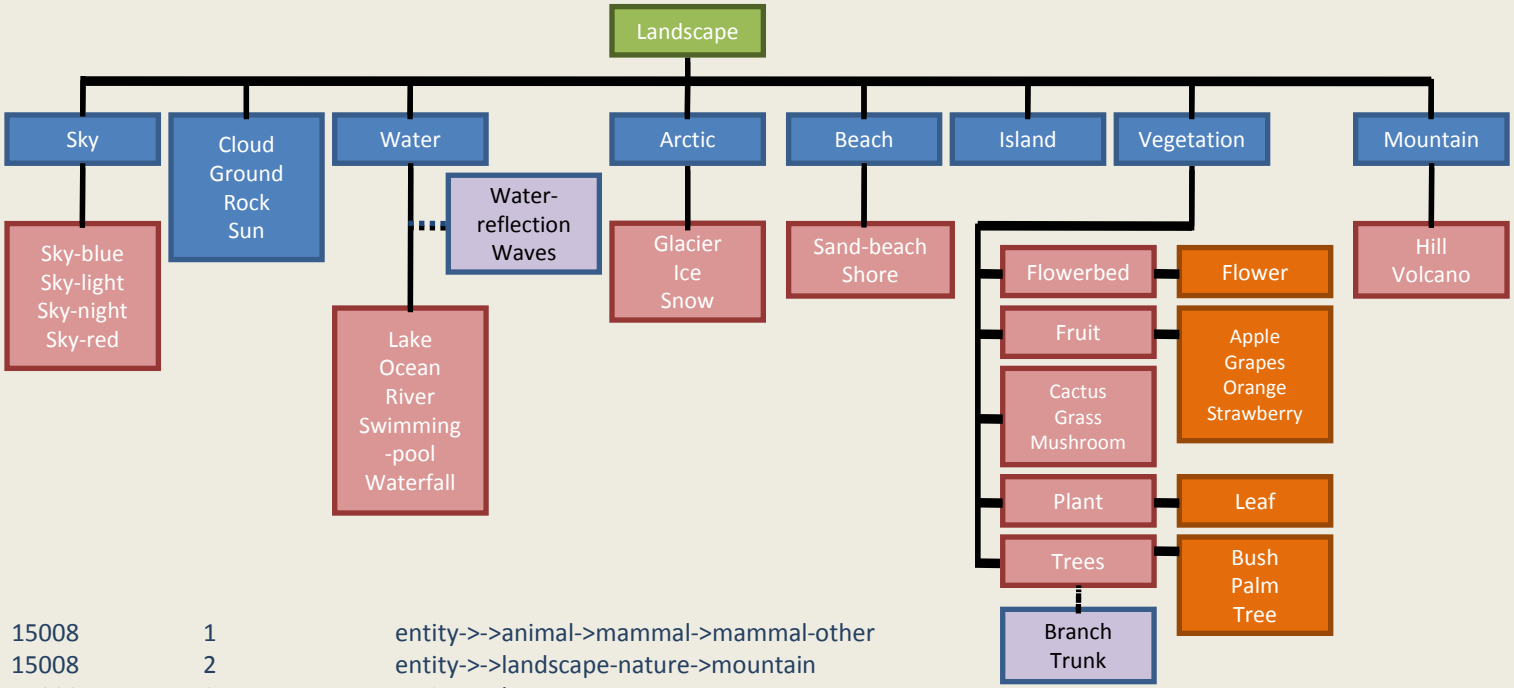
- **Annotation:** the vocabulary was organized into a conceptual hierarchy for improving the annotation process and for the soft evaluation of image annotation methods.

Branch	Animal	Humans	Food	Man-made	Landscape	Other
Frequency	1,991	16,950	705	34,024	45,308	622
Descendants	70	14	6	110	45	6
Leaves	56	12	5	88	33	6

[1] *Four annotators were considered.*



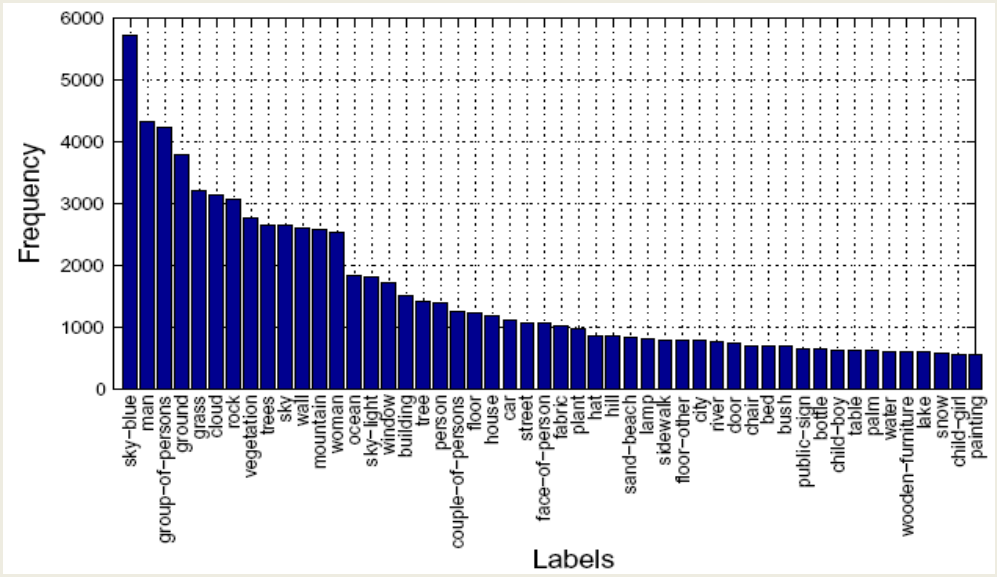
Annotation hierarchy



15008	1	entity->->animal->mammal->mammal-other
15008	2	entity->->landscape-nature->mountain
15008	3	entity->->humans->person->woman
15008	4	entity->->landscape-nature->dessert->sand-dessert
15009	1	entity->->landscape-nature->mountain
15009	2	entity->->landscape-nature->_sky->sky-light
15009	3	entity->->landscape-nature->_water
15009	4	entity->->landscape-nature->beach->sand-beach
15009	5	entity->->humans->_group-of-persons
15009	6	entity->->animal->mammal->mammal-other
15012	1	entity->->landscape-nature->_sky->sky-light

....

Statistics

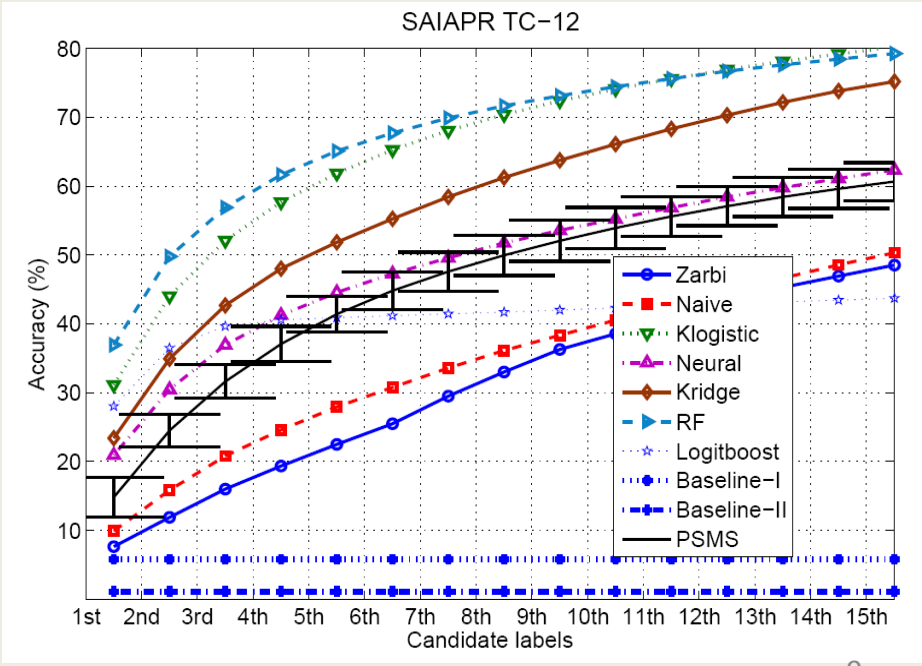


Feature	Statistics
Segmentation masks	20,000
Labels and features	99,535
Labels used	255 / 275 (92.8%)
Avg. area of regions	15.61 %
Avg. regions per image	4.97
Leaves used	200

Automatic Image Annotation

- Labels with at least 200 examples were considered
  - 90 labels
  - 91,139 regions (91.7%)
- 10-fold cross validation
- 7-basic classifiers[2] were tested (single-label / one-vs-all multiclass classification)

Classifier	Accuracy
zarbi	7.79 %
Naïve Bayes	8.05 %
Klogistic	31.12 %
Neural	20.98 %
Kridge	23.54 %
Random forest	**37.04 %**
Logitboost	28.20 %
Baseline-I	5.89 %
Baseline -II	1.11%

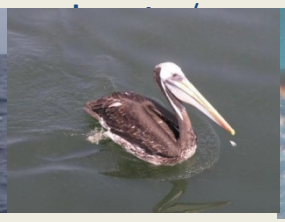


[2] http://clopinet.com/CLOP

Annotation Based Image Retrieval

- Experiments combining text + labels, using topics used for ImageCLEF 2007-2008:
  - **Late fusion (LF):** Combination of the outputs of independent retrieval models
  - **Early fusion (EF):** Concatenation of the representations for text and labels, then use a simple retrieval method
  - **Inter-media pseudo-relevance feedback:** 1) Retrieval using X-modality; 2) the top documents from 1) are used to build a query for modality Y; 3) a second retrieval stage is performed.

<title> **animal swimming** </title>
<narr> **Relevant images will show one or more animals (fish, birds, reptiles, etc.) swimming in a body of water. Images of people swimming in water are not relevant. Images of animals that are not swimming are not not**


sea, whale, sky **water, bird** **ocean, fish**

Annotation Based Image Retrieval

- Retrieval results

α weights		ImageCLEF 2007				ImageCLEF 2008			
α^l	α^t	MAP	P20	R20	RR	MAP	P20	R20	RR
Individual performance									
1	0	0.0587	0.1417	0.1066	1201	0.053	0.141	0.1133	727
0	1	0.1241	0.1767	0.1694	1424	0.1033	0.1795	0.1534	1014
Late fusion image retrieval									
1	1	0.1273	0.185	0.1817	1731	0.106	0.191	0.1683	1138
1	2	0.1348	0.1858	0.1879	1703	0.1126	0.1936	0.1759	1139
Early fusion image retrieval									
1	1	0.1276	0.2475	0.2498	1722	0.1167	0.2551	0.2342	1044
1	**8**	**0.189**	**0.2508**	**0.2996**	**2226**	**0.1565**	**0.2372**	**0.2695**	**1416**

Order		ImageCLEF 2007				ImageCLEF 2008			
Initial	Final	MAP	P20	R20	RR	MAP	P20	R20	RR
L	L	0.0841	0.165	0.136	1273	0.0714	0.1577	0.123	706
L	**T**	**0.1659**	**0.2142**	**0.262**	**1952**	**0.1326**	**0.1987**	**0.2205**	**1302**
T	T	0.1435	0.1867	0.2029	1717	0.1253	0.1974	0.2009	1255
T	L	0.055	0.1008	0.0884	1227	0.0484	0.1077	0.0847	774

Annotation Based Image Retrieval

- Retrieval results

α weights		ImageCLEF 2007				ImageCLEF 2008			
α^l	α^t	MAP	P20	R20	RR	MAP	P20	R20	RR
Individual performance									
1	0	0.0587	0.1417	0.1066	1201	0.053	0.141	0.1133	727
0	1	0.1241	0.1767	0.1694	1424	0.1033	0.1795	0.1534	1014
Late fusion image retrieval									
1	1	0.1273	0.185	0.1817	1731	0.106	0.191	0.1683	1138
1	2	0.1348	0.1858	0.1879	1703	0.1126	0.1936	0.1759	1139
Early fusion image retrieval									
1	1	0.1276	0.2475	0.2498	1722	0.1167	0.2551	0.2342	1044
1	**8**	**0.189**	**0.2508**	**0.2996**	**2226**	**0.1565**	**0.2372**	**0.2695**	**1416**

Order		ImageCLEF 2007				ImageCLEF 2008			
Initial	Final	MAP	P20	R20	RR	MAP	P20	R20	RR
L	L	0.0841	0.165	0.136	1273	0.0714	0.1577	0.123	706
L	**T**	**0.1659**	**0.2142**	**0.262**	**1952**	**0.1326**	**0.1987**	**0.2205**	**1302**
T	T	0.1435	0.1867	0.2029	1717	0.1253	0.1974	0.2009	1255
T	L	0.055	0.1008	0.0884	1227	0.0484	0.1077	0.0847	774

Annotation Based Image Retrieval

- Retrieval results

α weights		ImageCLEF 2007				ImageCLEF 2008			
α^l	α^t	MAP	P20	R20	RR	MAP	P20	R20	RR
Individual performance									
1	0	0.0587	0.1417	0.1066	1201	0.053	0.141	0.1133	727
0	1	0.1241	0.1767	0.1694	1424	0.1033	0.1795	0.1534	1014
Late fusion image retrieval									
1	1	0.1273	0.185	0.1817	1731	0.106	0.191	0.1683	1138
1	2	0.1348	0.1858	0.1879	1703	0.1126	0.1936	0.1759	1139
Early fusion image retrieval									
1	1	0.1276	0.2475	0.2498	1722	0.1167	0.2551	0.2342	1044
1	**8**	**0.189**	**0.2508**	**0.2996**	**2226**	**0.1565**	**0.2372**	**0.2695**	**1416**

Order		ImageCLEF 2007				ImageCLEF 2008			
Initial	Final	MAP	P20	R20	RR	MAP	P20	R20	RR
L	L	0.0841	0.165	0.136	1273	0.0714	0.1577	0.123	706
L	**T**	**0.1659**	**0.2142**	**0.262**	**1952**	**0.1326**	**0.1987**	**0.2205**	**1302**
T	T	0.1435	0.1867	0.2029	1717	0.1253	0.1974	0.2009	1255
T	L	0.055	0.1008	0.0884	1227	0.0484	0.1077	0.0847	774

Annotation Based Image Retrieval

- Retrieval results

α weights		ImageCLEF 2007				ImageCLEF 2008			
α^l	α^t	MAP	P20	R20	RR	MAP	P20	R20	RR
Individual performance									
1	0	0.0587	0.1417	0.1066	1201	0.053	0.141	0.1133	727
0	1	0.1241	0.1767	0.1694	1424	0.1033	0.1795	0.1534	1014
Late fusion image retrieval									
1	1	0.1273	0.185	0.1817	1731	0.106	0.191	0.1683	1138
1	2	0.1348	0.1858	0.1879	1703	0.1126	0.1936	0.1759	1139
Early fusion image retrieval									
1	1	0.1276	0.2475	0.2498	1722	0.1167	0.2551	0.2342	1044
1	**8**	**0.189**	**0.2508**	**0.2996**	**2226**	**0.1565**	**0.2372**	**0.2695**	**1416**

Order		ImageCLEF 2007				ImageCLEF 2008			
Initial	Final	MAP	P20	R20	RR	MAP	P20	R20	RR
L	L	0.0841	0.165	0.136	1273	0.0714	0.1577	0.123	706
L	**T**	**0.1659**	**0.2142**	**0.262**	**1952**	**0.1326**	**0.1987**	**0.2205**	**1302**
T	T	0.1435	0.1867	0.2029	1717	0.1253	0.1974	0.2009	1255
T	L	0.055	0.1008	0.0884	1227	0.0484	0.1077	0.0847	774

Current research . . .

- A complete evaluation on the impact of labels in multimedia image retrieval
  - How complimentary/redundant are labels and text?
  - Adaptive adjustment of weights for labels and textual information, according to the topic
- How can we improve the annotation performance?
  - Using the conceptual hierarchy
    - Hierarchical classification
    - Applying preference learning techniques
- How to combine labels and text so that retrieval performance is maximized?
- How can we take advantage of spatial information for annotation and retrieval

Final remarks

- The SAIAPR TC12 collection is out there
  - http://imageclef.org/SIAPRdata
- Results in image annotation confirm it is a very challenging problem
- Retrieval results show that labels can be helpful for image retrieval; thus, motivating the development of better methods for information fusion

Use the SAIAPR TC12 !

Acknowledgements: We are grateful with Thomas Deselaers for its valuable support.

Relevant publications:

- [1] H. J. Escalante, M. Grubinger, C. Hernandez, J. A. Gonzalez, A. Lopez, M. Montes, E. Morales, E. Sucar, and L. Villaseñor. **The Segmented and Annotated IAPR- TC12 Benchmark.** In Computer Vision and Image Understanding, in press, doi:10.1016/j.cviu.2009.03.008, 2009.

- [2] M. Grubinger. **Analysis and Evaluation of Visual Information Systems Performance.** PhD Thesis. School of Computer Science and Mathematics, Faculty of Health, Engineering and Science, Victoria University, Melbourne, Australia, 2007.

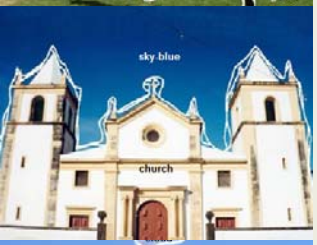

mountain


vegetation
mammal
ground


church-interior
group-of-persons
bench
floor-wood


wall
paper
wall
paper
child-girl
child-girl
child-girl
child-girl


sky-blue
trees
trees
fabric
horse
horse
fabric
grass


sky-blue
church


sky-blue
ocean
diver
diver
shore
sand-beach