SZTAKI @ ImageCLEF 2008 Visual Concept Detection*

Bálint Daróczy Zsolt Fekete Mátyás Brendel

Data Mining and Web search Research Group, Informatics Laboratory
Computer and Automation Research Institute of the Hungarian Academy of Sciences
{daroczyb, zsfekete, mbrendel}@ilab.sztaki.hu

Abstract

We describe our approach to the ImageCLEF-VisualConcept 2008 task. Our method is based on image segmentation, using a feature vector describing the visual content of image segments or the entire image. Logistic regression was used for classification. Images were segmented by a home developed segmenter. While in this preliminary report classification by global image features performed best, preliminary results suggest the importance of segmentation for certain classes. We are planning to provide improved analysis in the near future.

Categories and Subject Descriptors

H.3 [**Information Storage and Retrieval**]: H.3.1 Content Analysis and Indexing; H.3.3 Information Search and Retrieval; H.3.4 Systems and Software; H.3.7 Digital Libraries

General Terms

Measurement, Performance, Experimentation

Keywords

1 Introduction

In this paper we describe our approach to the ImageCLEF Visual Concept 2008 evaluation campaign over the IAPR TC-12 Benchmark [6]. The Visual Concept Detection Task has the objective to identify visual concepts. Both the training and test-set was a part of the IAPR TC-12 database. 1,800 images were published, which were classified according to a small concept hierarchy with 17 concepts. The test database consisted of 1,000 images. For each of these images it was required to determine the presence or absence of the concepts.

Our method is based on our approach to the object classification track of the previous year ([4]). As a main difference, we used global features in addition to segment based ones, since concepts such as night and day are characterized by the entire image. For this reason our CBIR method is based on segmentation of the image and on the comparison of features globally as well as segmentwise. While may of the existing CBIR systems rely on so called blobs, regions, or segments [3, 8, 2, 7], the specialty of our method is our special segmentation method and the combining of global and segment based approach.

*This work was supported by the EU FP7 project JUMAS – Judicial Management by Digital Libraries Semantics and by grants OTKA NK 72845 and NKFP-07-A2 *TEXTREND*.

Due to processing and classification costs we show preliminary results only that we plan to revise in the near future. As a main issue, we were not able to perform method selection and blending on a separate heldout set that, as expected, resulted in overfitting both for our classificator combination and for our segment filtering methods.

2 Visual feature generation

Our CBIR system [4, 1] relies on so called blobs, regions or segments. Classes such as building or people are classified by extracting specific features from the segments. For segmentation we use the code of the Felzenszwalb and Huttenlocher [5] graph-based method. Global classes such as outdoor are classified by using the entire image as a single segment.

By the distinction of classes that characterize global and local features of the image, respectively, we experimented with the number and size of the segments starting from a single segment per image for global classes down to a very large number of segments. After resizing images to a size of maximum 500x500 by keeping the aspect ratio, we tuned the minimum segment size and the cut parameters of the Felzenszwalb–Huttenlocher algorithm to select a *small* and a *medium* granularity segmentation. The *small* version resulted typically in more than 100 while *medium* in less than 100 segments per image. The minimum segment size is 50 pixels for *small* and 1500 for *medium*.

The runs submitted also differ in the features used to characterize the segments. We use mean color, RGB histogram and the 2D Fourier transform of the image in addition to shape values formed by converting segments to binary pattern, then resizing to 10x10 so that binary values are converted to grayscale values proportionally.

glob1: 33 values per image for mean color (RGB) and a 10-bin histogram for all the 3 channels (RGB). No segmentation is performed.

glob2: 173 values per image for mean color (RGB), a 20-bin histogram, 2x5 contrast (5 maximal and 5 minimal values of L-channel in HSL color-space) and 100 values of a 2D Fourier transform (sampled along zig-zag). No segmentation is performed.

medium: 135 values per segments for mean color, 3x10 histogram and 10x10 shape. Segments are of medium size, i.e. less than 100 in number per image.

small: Same as medium with small size segments, i.e. more than 100 segments per image.

3 Classification

We use logistic regression for classification with the global or segment features as input. The output real value is interpreted as the probability of the image or segment belonging to the specific class. For a single image we averaged the segment based predictions, which turned out more accurate than either the minimum or the maximum. Finally, a threshold of 0,5 was applied to get binary values. We did not use the logical information included in the class hierarchy, which could improve our method.

In our *mixed* run for each class we used the classifier that performed best on the training data. Due to time constraints we did not use a heldout set, which resulted in overtraining for this run. By closer analysis the *glob1* run was overtrained the most. By replacing *glob1* by *glob2* the combined performance improved over the best single run even in this overtrained scenario. The explanation for the overtraining for *glob1* may lie in the low number of features used.

4 Results

Table 1 summarizes our runs. The results were evaluated by the ImageCLEF organizers using the measures of equal error rate (EER) and area under ROC curve (AUC).

	Glob1	Glob2	Small			
			no filter	relabel	ppnpnn	ppnn
EER	45.72	31.14	32.44	32.48	32.46	36.07
AUC	52.78	74.90	73.32	73.03	73.05	67.15

	Medium				Logreg	Mixed	Mixed2
	no filter	relabel	ppnpnn	ppnn			
EER	32.10	32.47	32.47	37.01	37.12	38.34	29.92
AUC	74.18	73.57	73.61	59.30	66.53	63.80	72.77

Table 1: Performance of the three basic methods and their combination, evaluated by different measures

The three main variants are based on the granularity of the segmentation. We distinguish the single, 100- and 100+ segments per image labeled *Glob*, *Small* and *Medium*, respectively.

In the case of the segment based classification we introduced further variants for filtering out irrelevant segments from the training data. After filtering a new training was applied. The lack of a heldout data resulted in overfitting in this case as well. The four variants are

no filter: all segments are used for training the class;

ppnn: stands for discarding all segments from the training set except for those labeled correctly (positive for positive, negative for negative);

ppnpnn: is a more admissive filter that discards only segments with positive true label classified as negative.

relabel: stands for changing the true label of negatively classified segments to negative before the second training step.

Finally we submitted three combinations, all of them suffering from overtraining due to the lack of a heldout set.

Logreg: the output of the classifiers are combined by logistic regression again.

Mixed: For each class the method performing best on the training data was selected.

Mixed2: Partially resolving the overfitting of *Mixed*, *glob1* is always replaced by *glob2*. This run is included only in the post submission error analysis.

Conclusion and future work

In summary we may observe best overall performance for the high dimensional global feature space, closely followed by the medium resolution segmentation. We also reached improvement (although not among the submitted runs) by combination. Results in this report are preliminary and we are planning to rerun all our classificators by using separate heldout sets for segment filtering and combination.

References

[1] András Benczúr, István Bíró, Mátyás Brendel, Csalogány Károly, Bálint Daróczy, and Dávid Siklósi. Cross-modal retrieval by text and image feature biclustering. In *Working Notes for the CLEF 2007 Workshop*, Budapest,Hungary, 2007.

[2] Chad Carson, Serge Belongie, Hayit Greenspan, and Jitendra Malik. Blobworld: Image segmentation using expectation-maximization and its application to image querying. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(8):1026–1038, 2002.

[3] Yixin Chen and James Z. Wang. Image categorization by learning and reasoning with regions. *J. Mach. Learn. Res.*, 5:913–939, 2004.

[4] Thomas Deselaers, Allan Hanbury, Ville Viitaniemi, András Benczúr, Mátyás Brendel, Bálint Daróczy, Hugo Jair Escalante Balderas, Theo Gever, Carlos Arturo Hernández Gracidas, Steven C. H. Hoi, Jorma Laaksonen, Mingjing Li, Heidy Marisol Marin Castro, Hermann Ney, Xiaoguang Rui, Nicu Sebe, Julian Stöttinger, and Lei Wu. Overview of the imageclef 2007 object retrieval task. In *Working Notes for the CLEF 2007 Workshop*, Budapest,Hungary, 2007.

[5] Pedro F. Felzenszwalb and Daniel P. Huttenlocher. Efficient graph-based image segmentation. *International Journal of Computer Vision*, 59, 2004.

[6] Michael Grubinger, Paul Clough, Henning Müller, and Thomas Deselears. The IAPR TC-12 benchmark - a new evaluation resource for visual information systems. In *OntoImage*, pages 13–23, 2006.

[7] Qin Lv, Moses Charikar, and Kai Li. Image similarity search with compact data structures. In *CIKM '04: Proceedings of the Thirteenth ACM International Conference on Information and Knowledge Management*, pages 208–217, New York, NY, USA, 2004. ACM Press.

[8] B. G. Prasad, K. K. Biswas, and S. K. Gupta. Region-based image retrieval using integrated color, shape, and location index. *Comput. Vis. Image Underst.*, 94(1-3):193–233, 2004.