SZTAKI @ ImageCLEF 2010

Bálint Daróczy István Petrás András A. Benczúr
Dávid Nemeskey Róbert Pethes
Data Mining and Web search Research Group, Informatics Laboratory
Computer and Automation Research Institute of the Hungarian Academy of Sciences
{benczur, daroczyb, ndavid, petras, rpethes}@ilab.sztaki.hu

Abstract

Our approach to the ImageCLEF 2010 tasks is based on Histogram of Oriented Gradients descriptors (HOG) and Okapi BM25 based text retrieval. We extracted feature vectors to describe the visual content of an image region or the entire image. We trained a Gaussian Mixture Model (GMM) to cluster the feature vectors extracted from the image regions. To represent each image with only one vector we computed a Bag-of-Words (BOW) model from GMM probabilities of HOG descriptors. We trained linear regression classificators for the Photo Annotation task. To improve our textual ranking in the Wikipedia MM task we successfully expanded the textual query based on the visual ranking.

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; H.2.3 [**Database Managment**]: Languages—*Query Languages*

General Terms

Measurement, Performance, Experimentation

Keywords

HOG, Gaussian mixtures, BOW, Okapi BM25, rank aggregation

1 Introduction

In this paper we describe our approach to the ImageCLEF 2010 Wikipedia Retrieval and Photo Annotation evaluation campaigns [4, 1]. Our CBIR method is based on a Bag-of-Visual Words model based on HOG descriptors extracted from the images. We used the Hungarian Academy of Sciences search engine [2] as our information retrieval system that is based on Okapi BM25 [6] and query expansion by thesaurus.

2 Visual feature extraction and textual search engine

We extracted Histogram of Oriented Gradients descriptors over four color channel of each image (brightness and RGB channels). This resulted four times 81 dimensional descriptor for near 5 thousand regions per image. We trained a Gaussian Mixture Model on a randomly sampled subset of the descriptors. Finally we represented each image with a 128-entry vector (BOG -

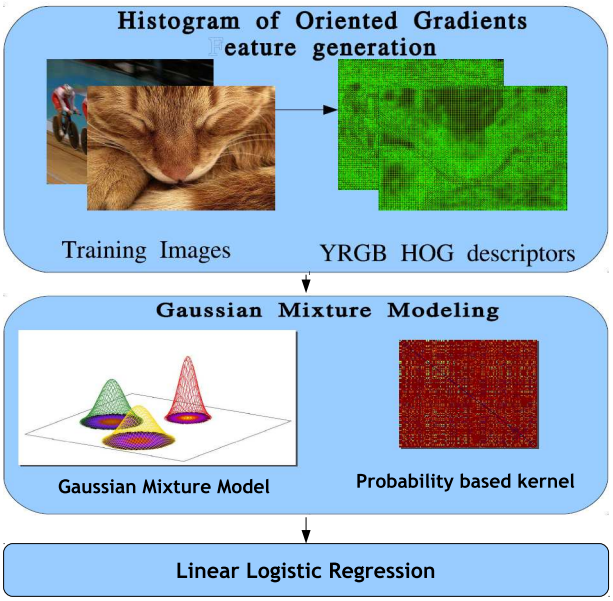

Figure 1: Feature extraction and classification procedure (Photo Annotation)

Bag-of-Gaussians), where each entry is refers to the sum of corresponding cluster probabilities (Fig. 1).

We use the Hungarian Academy of Sciences search engine [2] as our information retrieval system based on Okapi BM25 ranking [6] with the proximity of query terms taken into account [5, 3]. We deployed stopword removal and stemming by the Porter stemmer. We extended of stop word list with terms such as "photo" or "image" that are frequently used in annotations but does not have a distinctive meaning in this task.

We applied query term weighting to distinguish definite and rough query terms, the latter may be obtained from the topic description or a thesaurus. We multiplied the BM25 score of each query term by its weight; the sum of the scores gave the final rank.

3 The WikipediaMM Task

Nearest neighbor search was performed over the BOG model. We trained a Gaussian Mixture Model with 128 Gaussians over a sampled subset of HOG descriptors extracted from the 243 thousand images of the WikipediaMM data collection. Each image was represented by a smoothed aggregate of the GMM probabilities of the HOG descriptors extracted from the image. We preprocessed the annotation text by regular expressions to remove author and copyright information. We made no differentiation between the title and the body of the annotation.

Since file names often contain relevant keywords and also often as substring, we gave score proportional to the length of the matching substring. Since the indexing of all substrings is infeasible, we only performed this step for those documents that already matched at least one query term in their body.

For the WikipediaMM task we also deployed query expansion by an online thesaurus[1]. We added groups of synonyms with reduces weight so that only the score of the first few best performing synonym was added to the final score to avoid overscoring long lists of synonyms.

We performed a query expansion to combine the visual and the textual based systems. We found the text based score to be more accurate. Therefore we expanded the query with words

[1] `http://thesaurus.com/`

Table 1: WikiMediaMM ad hoc search evaluation.

	Modality	FB/QE	MAP	P10	P20
Text + Image (weight of expansion=0.3)	Mixed	QE	0.1794	0.4857	0.4329
Text + Image (weight of expansion=0.1)	Mixed	QE	0.1779	0.4729	0.4257
Text + Thesaurus	Textual	NOFB	0.1768	0.4714	0.4229
Text only	Textual	NOFB	0.1737	0.4671	0.4171
Image only	Visual	NOFB	0.0015	0.0071	0.0079

from the textual description of the top K ranked images. Ranking algorithm weighting the words from the query differently.

As seen in Table 1, our our combination of the visual and textual based ranking improved performance in terms of MAP. The results showed the weightning of the expansion words with 0.3 instead of 0.1 increased the performance of the system. Expansion by thesaurus improved the performance in a similar sense.

4 Photo Annotation

We produced a global visual vocabulary that approximate the per-image distribution of the low level features by clustering with a 128 dimensional GMM. First we obtained a variable number of visual words per image that we processed by Bag-of-Words model (BOG). We trained linear logistic regression model on χ^2 based kernel for each category. Our system performed an average MAP 0.283178 over the 94 categories of the Photo Annotation 2010 Task.

5 Conclusions

- For image retrieval, we successfully combined visual and textual based systems.
- The use of the thesaurus and other query expansion techniques needs further analysis and refinement.
- We also plan to strengthen our results by using more sophisticated methods for text and image retrieval fusion.

References

[1] Theodora Tsikrika Adrian Popescu and Jana Kludas. Overview of the Wikipedia Retrieval task at ImageCLEF 2010. In *Working Notes of CLEF 2010*, 2010.

[2] András A. Benczúr, Károly Csalogány, Eszter Friedman, Dániel Fogaras, Tamás Sarlós, Máté Uher, and Eszter Windhager. Searching a small national domain—preliminary report. In *Proceedings of the 12th World Wide Web Conference (WWW)*, Budapest, Hungary, 2003.

[3] Stefan Büttcher, Charles L. A. Clarke, and Brad Lushman. Term proximity scoring for ad-hoc retrieval on very large text collections. In *SIGIR '06*, pages 621–622, New York, NY, USA, 2006. ACM Press.

[4] Stefanie Nowak and Mark Huiskes. New Strategies for Image Annotation: Overview of the Photo Annotation Task at ImageCLEF 2010. In *Working Notes of CLEF 2010*, 2010.

[5] Yves Rasolofo and Jacques Savoy. Term proximity scoring for keyword-based retrieval systems. In *Advances in Information Retrieval*, LNCS, pages 207–218. Springer, 2003.

[6] Stephen E. Robertson and Karen Sparck Jones. Relevance weighting of search terms. In *Document retrieval systems*, pages 143–160. Taylor Graham Publishing, London, UK, UK, 1988.