CLaC at ImageCLEF 2009

Osama El Demerdash, Sabine Bergler and Leila Kosseim
Concordia University
osama_el,bergler,kosseim@cse.concordia.ca

Abstract

This paper describes our participation at ImageCLEF 2009. We participated in the photographic retrieval task (ImageCLEFPhoto). Our method is based on intermedia pseudo-relevance feedback. We have enhanced the pseudo-relevance feedback mechanism by using semantic selectional restrictions. We use Terrier for text retrieval and our own simple block-based visual retrieval engine. The results obtained at imageCLEF 2009 show that our method is robust and promising. However, there is room for improvement on the visual retrieval as well as the topics without cluster descriptions.

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

Pseudo-relevance feedback, Semantic retrieval

1 Introduction

This paper describes our participation at ImageCLEF 2009. We participated in the photographic retrieval task (ImageCLEFPhoto). This year's task targeted promoting diversity in image search. It involved an annotated image collection of approximately half a million images, and fifty queries divided into two sets: one with a subject and provided specific subtopics (clusters), while the other with only a topic. A full description of the task can be found in [4].

We submitted four runs, aiming at evaluating our method as well as the resources used. Similar to our participation last year, our method is based on intermedia pseudo-relevance feedback. However, to in order to account for the much larger data set, we have introduced some modifications to our visual component. We have also enhanced the textual retrieval component, as well as to the pseudo-relevance feedback mechanism by using semantic selectional restrictions.

The results obtained at imageCLEF 2009 show that our method is robust and promising. However there is room for improvement on the visual retrieval as well as the topics without cluster descriptions.

The rest of this paper is organized as follows: Section 2 describes the visual retrieval component, Section 3 the text processing of the query, Section 4 the enhanced pseudo-relevance feedback and Section 5 the results we obtained at ImageCLEF 2009, then we conclude the paper.

2 Visual Retrieval

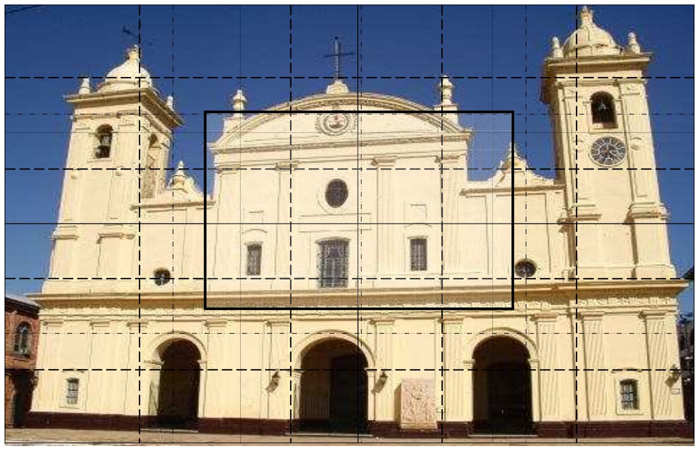

Figure 1: Block-based Visual Features

Figure 1 shows the different regional divisions used to analyze the image. In order to capture different levels of detail, we divide the image into 2X2, 3X3, 4X4 blocks yielding 4, 9, 16 equal partitions respectively. Due to the the much larger size of the data set compared to the IAPR TC-12 collection (60,000 images) used in previous years, we resorted to reducing the index by eleminating some of the descriptors we used previously, such as the grey-level and gradient-magnitude descriptors. The image is first converted to the Intensity/Hue/Saturation (IHS) color space, a perceptual color space which is more intuitive and reflective of human color perception than the RGB color space. This also allows for assigning more weight to the hue component which is a better discriminating feature as shown in [6].

As has been illustrated before in [2], the moments of histograms are efficient approximations of the entire histogram. Therefore, for each of the three-band color histograms of the divisions, the first two moments (the mean and the average energy) as well as the standard deviation are stored in the index.

For retrieval, the different partitions are compared to their counter parts in the query images. We selected the Manhattan distance (L^1 Norm) after investigating several other measures including the Euclidean and the Mahalanobis distances, combined with a measure for the number of blocks within a minimum threshold for the distance. Since all features were represented as histograms with the same number of bins (256), no normalization was necessary. The images in the database were ranked according to their highest proximity to any of the three query images. This choice presumes that our simple features do not perform equally well on all example images.

3 Text Retrieval

The text is tokenized and preprocessed by removing stop words (grammatical words which do not contribute to the meaning) and punctuation. The rest of the terms are converted to lower case and stemmed using the Snowball stemmer [5].

Queries are tokenized and preprocessed similarly; stop words and punctuation are removed and the rest of the terms are stemmed. The queries consist of a topic and a cluster description when available, in addition to the expansion terms from the top visual results. Named-entities are given more weight and multiple-token named-entities are chunked into one term by adding quotes around them.

For text retrieval, we use the Terrier Information Retrieval platform, a Java-based Information Retrieval platform available from the University of Glasgow [3]. Terrier includes boolean, vector-space and probabilistic model capabilities. We use the vector-space model, which slightly surpasses the probabilistic models in our experiments. In the vector-space model, documents and queries are represented as vectors of terms weighed by Term Frequencies multiplied by the Inverse Document Frequency (TF-IDF). Terrier also has the option of block-indexing for phrase querying which we employ. Query terms are considered unioned by Terrier in order to promote recall.

4 Pseudo-Relevance Feedback with Semantic Selectional Restrictions

In this phase, the query is expanded with terms potentially related to the query. Common ways for text query expansion include adding synonyms and other related terms to the query. However, according to our experiments, this approach leads to the introduction of many noisy terms. Instead, we opted for the extraction of related terms from the five highest-ranked results retrieved by the content-based system described in Section 2. For the data set in our experiments, all the terms associated with the image are extracted except for stop words. In order to expand the query without introducing noise, the candidate text is compared to the query topic. If the image is found to be potentially related to the topic, the text query is expanded with the relevant terms. To compute the relatedness of the image annotaion to the topic, we use the minimum threshold of one common non-grammatical word, due to data sparseness.

The purpose of the query expansion module is not only to augment the query by adding new candidate related terms to it, but also to enhance it by adding weights to its key terms and filtering out potentially noisy terms from expansion. We also avoid expanding the query with named entities that do not have a semantic relationship with the query. This is crucial in photographic collections, since by their nature, photographs and image queries are often bound by geographical constraints. In order to ensure that potential expansion images do not introduce conflicting geographical terms in the query, we first build a filter from the location specified in the query. We make use of WordNet [1], a lexical database, by traversing its *PartMeronym* hierarchy. A *PartMeronym* is a relationship between two nouns where the child noun constitutes a part of the parent noun. For geographical locations, this translates by the divisions of the parent noun. For example for the USA, a traversal of the hierarchy produces the names of the states, then major cities and towns followed by specific locations. While similar filters are possible for common nouns and using other relations such as Hyponymy (sub-classes of a term), we limit the expansion to named-entities, so as to avoid the problem of disambiguation of the specific sense of the term.

5 Results

Table 1 shows the results our runs obtained at ImageCLEF 2009. The first two runs are purely visual and textual respectively. The *PRF* run combines visual and text retrieval using Pseudo-relevance feedback and separate queries for each cluster. The *Combined* run uses the same method as the *PRF*, while combining all clusters into one query. Tables 2 and 3 show the break down of these runs by query set. The text run on the queries without given clusters shows a significant degradation which might be attributed to a glitch in our system.

Our results show a significant improvement of the pseudo-relevance feedback method over the use of a single modality. We also note a significant difference between the precision and cluster recall at ten (P10 & CR10) and at 20 (P20 & CR20) retrieved results. Contrary to ImageCLEFPhoto 2008 the F-measure was computed this year using a cut-off of the first ten results, which was a disadvantage to our method. The MAP, GMAP and the Relevant Retrieved figures are promising and show consistency over the different topics.

Table 1: Results at ImageCLEFPhoto 2009

Description	**P10**	**P20**	**CR10**	**CR20**	**MAP**	**GMAP**	**Rel_Ret**	**F-measure**
Visual	0.096	0.099	0.298	0.434	0.006	0.0009	657	0.1452
Text	0.434	0.435	0.4187	0.4437	0.226	0.0069	9603	0.4262
With PRF	0.55	0.643	0.7027	0.8212	0.3939	0.3164	16600	0.6171
Combined	0.586	0.672	0.6605	0.7569	0.4106	0.325	16677	0.621

Table 2: Queries with Given Clusters

Description	**P10**	**P20**	**CR10**	**CR20**	**MAP**	**GMAP**	**Rel_Ret**	**F-measure**
Visual	0.072	0.082	0.2603	0.3934	0.0026	0.0008	241	0.1128
Text	0.732	0.748	0.7416	0.7726	0.4274	0.3178	8438	0.7368
With PRF	0.548	0.688	0.7482	0.8772	0.4155	0.3602	8638	0.6327
Combined	0.604	0.724	0.6741	0.7702	0.4459	0.3846	8785	0.6371

Table 3: Queries without Given Clusters

Description	**P10**	**P20**	**CR10**	**CR20**	**MAP**	**GMAP**	**Rel_Ret**	**F-measure**
Visual	0.12	0.116	0.3357	0.4757	0.0095	0.001	416	0.1768
Text	0.136	0.122	0.0957	0.1148	0.0247	0.0001	1165	0.1124
With PRF	0.552	0.598	0.6572	0.7652	0.3939	0.3164	7962	0.6171
Combined	0.568	0.62	0.6469	0.7435	0.3753	0.2747	7892	0.6049

6 Conclusion

We experimented at ImageCLEF 2009 with applying semantic selectional restrictions to enhance intermedia pseudo-relevance feedback and different methods of query formulation for clustered queries. We will further analyze the results in order to understand the significance of the chosen measures, given that the precision varies significantly at different levels of recall.

References

[1] Christiane Fellbaum, editor. *WordNet An Electronic Lexical Database*. The MIT Press, Cambridge, MA ; London, May 1998.

[2] M.K. Mandal, T. Aboulnasr, and S. Panchanathan. Image indexing using moments and wavelets. *IEEE Transactions on Consumer Electronics*, 42:557–565, 1996.

[3] I. Ounis, G. Amati, V. Plachouras, B. He, C. Macdonald, and C. Lioma. Terrier: A High Performance and Scalable Information Retrieval Platform. In *Proceedings of ACM SIGIR'06 Workshop on Open Source Information Retrieval (OSIR 2006)*, 2006.

[4] M. Paramita, M. Sanderson, and P. Clough. Diversity in Photo Retrieval: Overview of the ImageCLEFPhoto Task 2009. In *CLEF working notes 2009, Corfu, Greece*, 2009.

[5] Martin F. Porter. Snowball: A language for stemming algorithms. Published online, October 2001. Accessed 16.04.2009, 18.00h.

[6] Markus A. Stricker and Markus Orengo. Similarity of color images. In *Storage and Retrieval for Image and Video Databases (SPIE)*, pages 381–392, 1995.