CLaC at ImageCLEFPhoto 2008

Osama El Demerdash, Leila Kosseim and Sabine Bergler

Concordia University

{osama_el,kosseim,bergler}@cse.concordia.ca

Abstract

This paper presents our participation at the ImageCLEFPhoto 2008 task. We submitted six runs, experimenting with our own block-based visual retrieval as well as with query expansion. The results we obtained show that despite the poor performance of the visual and text retrieval components, good results can be obtained through Pseudo-relevance feedback and the fusion of the results.

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 Management**]: Languages—*Query Languages*

General Terms

Measurement, Performance, Experimentation

Keywords

Image Retrieval, Pseudo-relevance feedback

1 Introduction

This paper presents our participation at the ImageCLEFPhoto 2008 task. We submitted six runs experimenting with our own block-based visual retrieval as well as with query expansion. While the 2008 task introduced a focused clustering theme for the first time, we did not attempt to use the cluster target information. Our resources comprise a text search engine and a content-based search system that we developed. The results we obtained show that despite the poor performance of the visual and text retrieval components, good results can be obtained through Pseudo-relevance feedback and the inter-media fusion of the results.

2 Related Work

Block-based methods have been used extensively in image retrieval. Examples can be found in [2] and [4]. [3] investigated pseudo-relevance feedback and fusion methods on the same dataset used in our experiments, the IAPR TC-12 collection. They reported a precision gain with feedback but not with fusion.
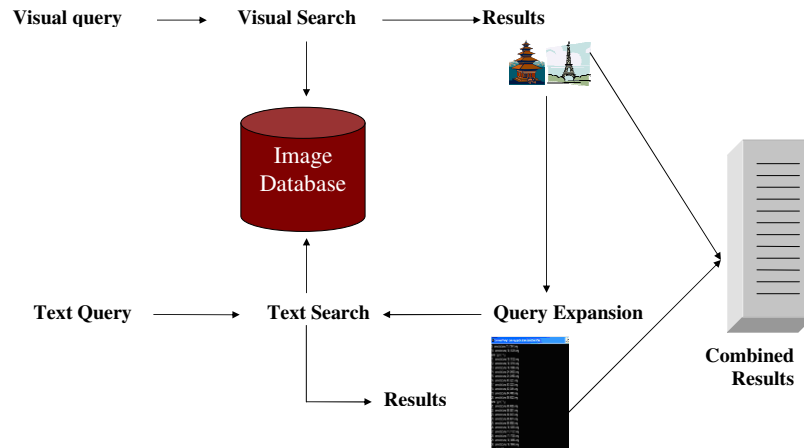
Figure 1: Architecture of the System

3   Resources

For text retrieval, we used the Apache Lucene engine [1] , which implements a TF-IDF paradigm. Stemming was done using the Snow-ball stemmer based on the porter algorithm and available separately under the BSD license at http://snowball.tartarus.org/. For image analysis, we utilized the Java Advanced Imaging (JAI) API v. 1.1.3. JAI is available from http://www.java.net.

4   Text Retrieval

As mentioned earlier, for text retrieval, we used the Apache Lucene engine, which implements a TF-IDF paradigm. Stop-words were removed, the rest of the terms were stemmed using the porter stemming algorithm. The documents were then indexed as *field data* retaining only the *title*, *notes* and *location* fields, all of which were concatenated into one field. All text query terms were joined using the *OR* operator.

When searching the text, the query is also stemmed and stop words are removed. we found that using the first sentence of the narrative field in addition to the title field improves the result. By contrast, the rest of the narrative needs semantic processing to avoid introducing noise.

Figure 2: Partitioning the Image for Visual Retrieval

5 Visual Retrieval

For visual retrieval, we implemented a system based on unsupervised analysis of the image. We sought to capture basic global and local color, texture and shape information. In order to achieve this goal, we divided the image into 2X2, 3X3, 4X4 and 5X5 blocks yielding 4, 9, 16 and 25 equal partitions respectively. We also used the image as a whole as well as a center block occupying half the image dimensions. Figure 2 shows the different regional divisions used to analyze the image. The image was first converted to the Intensity/Hue/Saturation (IHS) color space, a perceptual color space which more intuitive and reflective of human color perception than the RGB color Space, then the following features were extracted:

- A three-band color histogram for each of the image divisions

- A histogram of the grey level image

- A histogram of the gradient magnitude image for each of the divisions of the grey-level image

- A three-band color histogram of the thumbnail of the image

The first feature captures the color characteristics of the image, while the grey-level histogram conveys some texture information. The gradient magnitude adds the outline of the shapes in the image.

For retrieval, the different partitions are compared to their counter parts in the query image. We did not experiment with assigning different weights to features. After experimenting with several measures including the Euclidean distance, the Manhattan distance was chosen as the distance measure. The images in the database were ranked according to their highest proximity to any of the three query images.

6   Query Expansion and Fusion of the Results

The next step in processing the query is the text query expansion involving a pseudo-relevance feedback mechanism and the fusion of the text and visual search results.

6.1   Query Expansion

We attempted several ways of query expansion. The highest ranked results from each of the text and visual search engines were passed to the other engine. We also attempted adding noun synonyms from WordNet to the query.

Our best run uses pseudo-relevance feedback for query expansion. While the MAP of the visual only run is only 0.055, its precision at five retrieved documents (0.328) is significantly higher than that of the text only run (0.236). For this reason, we use the highest ranked document for expansion of the text query. This is only done if the document meets a confidence level that we determined empirically. The confidence score is assigned based on the proximity score to the query image.

6.2   Fusion of Image and Text Search Results

To combine the results from the different media searches, we again took into consideration the confidence level in the visual results (i.e. the level of proximity from the query images). A maximum of three highest ranked images is taken from the visual query result depending on the confidence score, followed by a majority of the text results after query expansion.

7   Results

We submitted six runs at ImageCLEFPhoto 2008. Following is their description:

- clacTX: Uses text search with *title* field only

- clacTXNR: uses Text Search with Title field and the first sentence of Narrative field with query expansion

- clacIR: uses visual search only

- clacIRTX: combines the results from clacTXNR and clacIR

- clacNoQE: uses Text Search on Title and Narrative Field without query expansion

- clacNoQEMX: same as clacNoQE combined with clacIR

Table 1 shows the results we obtained in comparison to the mean, median and best runs of the track, taken from the best 4 runs from each participating group (25 groups and 100 runs in total) . As we expected we obtained our highest Mean Average Precision (MAP) for the run that utilized the maximum of our resources and methods.

References

[1] Otis Gospodnetic and Erik Hatcher. *Lucene in Action*. 2005.

[2] Jun-Hua Han and De-Shuang Huang. A novel BP-Based Image Retrieval System. In *International Symposium on Circuits and Systems (ISCAS 2005), 23-26 May 2005, Kobe, Japan*, pages 1557–1560. IEEE, 2005.

Table 1: Results at ImageCLEFPhoto 2008.

Run ID	Modality	MAP	P10	P20	P30	GMAP	Rel
clacTX	Text	0.1201	0.1872	0.1487	0.1462	0.019	1155
clacTXNR	Text	0.2577	0.4103	0.3449	0.3085	0.1081	1859
clacIR	Visual	0.0552	0.2282	0.1615	0.1214	0.0268	629
clacIRTX	Mixed	0.2622	0.4359	0.3744	0.3308	0.1551	1630
clacNoQE	Mixed	0.2034	0.3205	0.2705	0.2487	0.078	1701
clacNoQEMX	Mixed	0.218	0.4026	0.3269	0.2855	0.129	1546
Average run	N/A	0.2187		0.3203			
Median run	N/A	0.2096		0.3203			
Best run	N/A	0.4288		0.6962			

[3] Nicolas Maillot, Jean-Pierre Chevallet, and Joo-Hwee Lim. Inter-media pseudo-relevance feedback application to imageclef 2006 photo retrieval. In Carol Peters, Paul Clough, Fredric C. Gey, Jussi Karlgren, Bernardo Magnini, Douglas W. Oard, Maarten de Rijke, and Maximilian Stempfhuber, editors, *CLEF*, volume 4730 of *Lecture Notes in Computer Science*, pages 735–738. Springer, 2006.

[4] Takala V., Ahonen T., and Pietikäinen M. Block-based methods for image retrieval using local binary patterns. 2005. In: Image Analysis, SCIA 2005 Proceedings, Lecture Notes in Computer Science 3540, Springer, 882-891.