Improving the Image Retrieval System by Ranking

Petra Budikova
budikova@ics.muni.cz

Michal Batko
batko@fi.muni.cz

Pavel Zezula
zezula@fi.muni.cz

Faculty of Informatics, Masaryk University, Brno, Czech Republic

Keywords

image retrieval system, ranking, relevance-feedback

1. INTRODUCTION

With the rapid growth of multimedia data, a lot of attention has been recently devoted to the development of multimedia retrieval systems. The research has followed two main directions: The first one applies existing text-search mechanisms to retrieve multimedia data based on its descriptive annotations, the second approach retrieves data by content. In case of text-based searching, the quality of results depends on the quality of text metadata, which is often not very high (especially in large general-purpose collections such as web image galleries). In the content-based approach, data objects are indexed and searched using features extracted from the data that describe their important characteristics. However, this solution suffers from the well-known *semantic gap* problem, i.e. the discrepancy between the similarity as computed using the descriptors and human understanding of similarity.

In our approach, we propose to bridge the semantic gap by combining both the orthogonal views. This method has already been proved to be very successful in the text-based searching – some of the major search engines (Google[1], Bing[2]) recently launched a new type of searching based on visual similarity of images. Both solutions exploit visual ranking of search results acquired by text retrieval [1, 5]. Result post-processing has also been employed in some content-based strategies to filter out less interesting objects from the result, usually by means of result clustering [4]. However, the existing content-based approaches do not employ additional measures of similarity. In our system, we provide a novel solution for large-scale content-based retrieval which enables to obtain high quality results by incorporating several measures of similarity into the search process.

[1]http://images.google.com/
[2]http://www.bing.com/images

SISAP'10 September 18–19, 2010, Istanbul, Turkey

2. RANKING

Ranking is often considered an integral part of the search process – search engines deliver ranked results. However, we model searching as a two-phase process, as depicted in Figure 1. During the initial search, suitable candidates are selected from the dataset and submitted to the ranking phase, where the more relevant objects from the candidate set are pushed to the top of the list. The ranking can be done either automatically, using the properties of candidate objects and statistics, or in cooperation with users that can actively participate in the process of defining the ranking function.

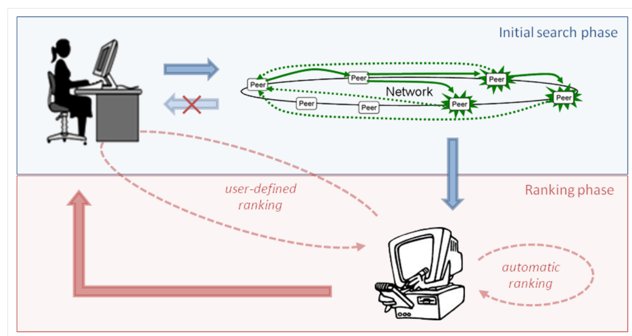

Figure 1: The two-phase search schema

In our scenario, we use image descriptors that form a metric space $\mathcal{M} = (\mathcal{D}, d)$. The initial search $F_{initial}$ is performed by any standard metric search query operation, e.g. the k-nearest neighbor search. In the ranking phase, a function $F_{rank} : \mathcal{D} \mapsto \mathbb{N}$ is applied on the result of $F_{initial}$ to establish a new rank of each object. The ranking function depends on the context in which it is evaluated and its computation may contain additional context-derived parameters.

Even though a user is interested in the first k objects, with k typically ranging from 10 to 100, the initial search should provide significantly more objects in order to allow the ranking to show interesting new data. The choice of the initial result size k' needs to balance the following three factors: the costs of the initial search for k' best objects, the cost of ranking the k' objects, and the probability that there are at least k relevant objects in the initial result of size k'.

We define several different types of ranking functions that are orthogonal to the content-based similarity. As our target data are images from a commercial microstock site with rich annotations, the ranking mainly exploits the text information. However, any other metadata such as time, location or popularity of a given object could be used as well.

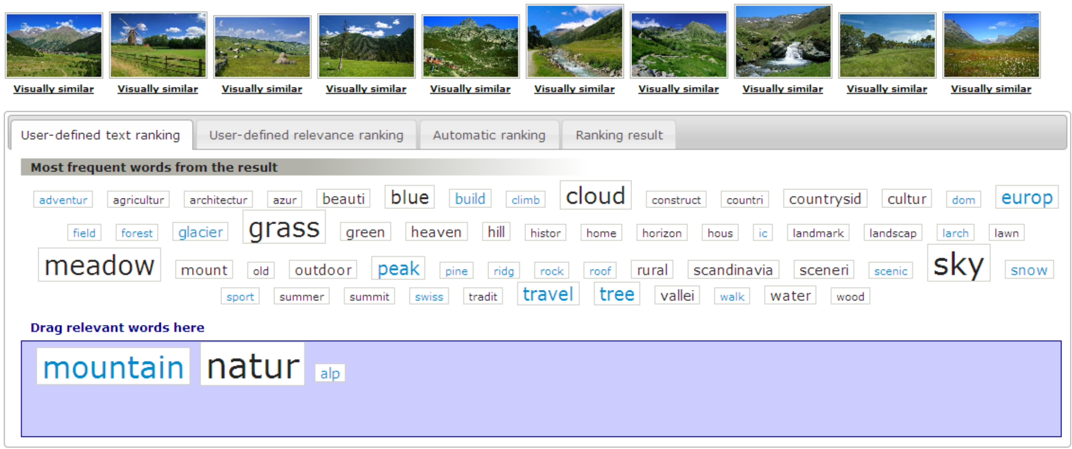

Figure 2: Interface of the demonstrated application

Keyword ranking Inversely to the search model applied by the common web search engines that combine text-based retrieval and visual ranking, we propose to rank the content-based search result with respect to keywords of the query image. The similarity between two sets of keywords is measured by the Jaccard coefficient.This ranking method is intended for data with rich and reliable annotations and can be enhanced using WordNet semantic relationships.

Word cloud ranking For data with sparse and erroneous text metadata, the keyword ranking is not applicable. In this case, we exploit the keywords of all objects in the initial result and compute their frequencies. We call the resulting set of keywords and frequencies the *word cloud*. Finally, the ranking employs the most frequent words from the cloud.

Combined visual and text ranking In the previous methods, we have only used the textual (keyword) information for the ranking, ignoring the initial ranking of the visual (content-based) search. However, it may also be useful to include it into the final ranking – we provide both the keyword and word cloud variants.

Selected descriptors ranking In the initial search, similarity of images is evaluated using a combination of several visual descriptors which refer to image's color, shape and texture. In the ranking phase, it is possible to choose the most important descriptor.

Relevance feedback ranking We propose to implement the relevance feedback as a ranking function. Users choose relevant objects from the initial result and the ranking function defines the final rank as a function on the content-based similarity to each of the objects marked as relevant.

User-defined keyword ranking Keywords may provide a strong ranking tool but automatic approaches may not always guess the optimal set of words. This method allows users to define the relevant keywords themselves.

3. APPLICATION

We demonstrate our solution over a dataset of commercial microstock photos with systematic annotations, which contains 8.3 million images. The content-based similarity of images is defined as a combination of several MPEG-7 descriptors [2] and each image is annotated by about 25 keywords on average.

We use an instance of the Multi-Feature Indexing Network (MUFIN) [3] for the initial search. The capabilities of MUFIN allow us to retrieve the results very fast even for large collections of data and its interfaces made it possible to plug-in the proposed ranking algorithms seamlessly into the web user interfaces. For each query image, 200 nearest images are submitted to the ranking phase and processed by a selected ranking function. The 10 most relevant objects are then shown to the user.

The web interface of the application (see Figure 2) enables easy selection of the preferred ranking function and a simple drag-and-drop marking of the relevant images or keywords in case of user-defined ranking. Users are then shown both the initial result and the ranked one. The application is available at **http://mufin.fi.muni.cz/ranking/**.

Acknowledgments

This work has been partially supported by the national research projects GACR 201/08/P507, GACR 201/09/0683 and by Brno PHD Talent Scholarship. The hardware infrastructure was provided by the METACentrum under the research intent MSM6383917201.

4. REFERENCES

[1] Y. Jing and S. Baluja. VisualRank: Applying PageRank to large-scale image search. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(11):1877–1890, 2008.

[2] MPEG-7. Multimedia content description interfaces. Part 3: Visual. ISO/IEC 15938-3:2002, 2002.

[3] D. Novak, M. Batko, and P. Zezula. Generic similarity search engine demonstrated by an image retrieval application. In *SIGIR '09*, page 840, 2009.

[4] R. H. van Leuken, L. Garcia, X. Olivares, and R. van Zwol. Visual diversification of image search results. In *WWW '09*, pages 341–350, New York, NY, USA, 2009. ACM.

[5] L. Wang, L. Yang, and X. Tian. Query aware visual similarity propagation for image search reranking. In *ACM Multimedia '09*, pages 725–728, New York, NY, USA, 2009. ACM.