Online Image Annotation

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, search by content, annotation

1. INTRODUCTION

With the rapid growth of the amount of multimedia data, the need for efficient ways of data organization and searching is indisputable. Some tasks can be accomplished using content-based retrieval which uses only data-specific features, but there are also applications where text annotations of the multimedia objects are necessary, e.g. to evaluate queries formulated in natural language or to categorize data objects. Since a manual creation of text metadata is a tedious work, automatic annotation techniques are much needed. The image annotation is a challenging research topic with many subproblems, ranging from tag recommendations to a full recognition of objects in the image.

In this paper, we present our system for online annotation of common web images. There are many situations where such a system can be used, e.g. to pre-select relevant keywords for users who want to tag their images, or to clean or expand existing user-provided annotations. In these applications, the speed of the annotation is one of the most important qualities. Our solution, which is based on a large scale content-based search system, is able to provide an annotation of the submitted image in real-time.

2. RELATED WORK

In recent years, many approaches to image annotation have been proposed, differing in the targeted applications and, subsequently, in the techniques used. Basically, these approaches can be divided into two categories: text-based methods and image-based methods. The text-based methods are used when the image under examination is surrounded by some text, typically a web page, from which the relevant textual information can be mined. Various semantic tools such as the WordNet database [2] are utilized to extract the most informative keywords from the text [4].

SISAP '11, June 30-July 1, 2011, Lipari, Italy

On the other hand, the image-based approaches [3, 6, 7] require just an image as input and need to exploit its visual features in order to select the relevant keywords. For these techniques, there are two main annotation strategies. Either some form of machine learning is employed to connect the visual features with keywords, or a content-based search is evaluated over a collection of annotated images and the relevant keywords are selected from among the keywords of the nearest neighbors. The machine learning techniques, described for instance in [3], comprise classifiers and probabilistic approaches based on text and image co-occurrences in a training dataset. Such approaches typically work with a limited set of class labels. On the contrary, the techniques which exploit content-based retrieval have a potentially unlimited vocabulary provided by an underlying tagged image collection, usually downloaded from some photo-sharing site. Current research in this area focuses on heuristics for selecting the relevant keywords and statistical or semantic associations between the tags [6, 7]. Both the image-based strategies can be further combined with text-based techniques to clear or expand the annotation.

3. MUFIN ANNOTATION STRATEGY

The MUFIN annotation system is built over the MUFIN similarity search engine [5] which is capable of on-line searching in very large data collections. The annotation system exploits this capability to search in several large collections in parallel. As depicted in Figure 1, the whole annotation process is composed of four phases: multiple *Content-based searches*, *Results processing*, optional *Additional search*, and final *Annotation cleaning*.

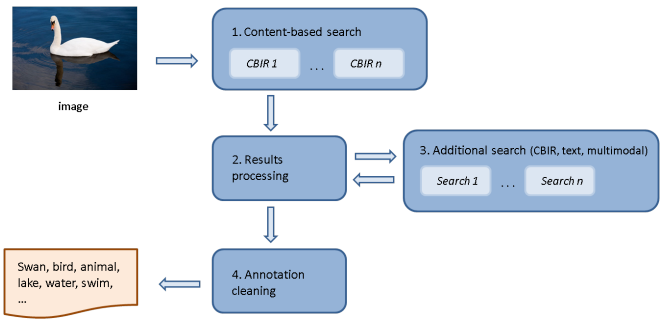

Figure 1: MUFIN annotation system architecture

In the first phase, the query image is sent to several CBIR subsystems, each of which is built over a different dataset.

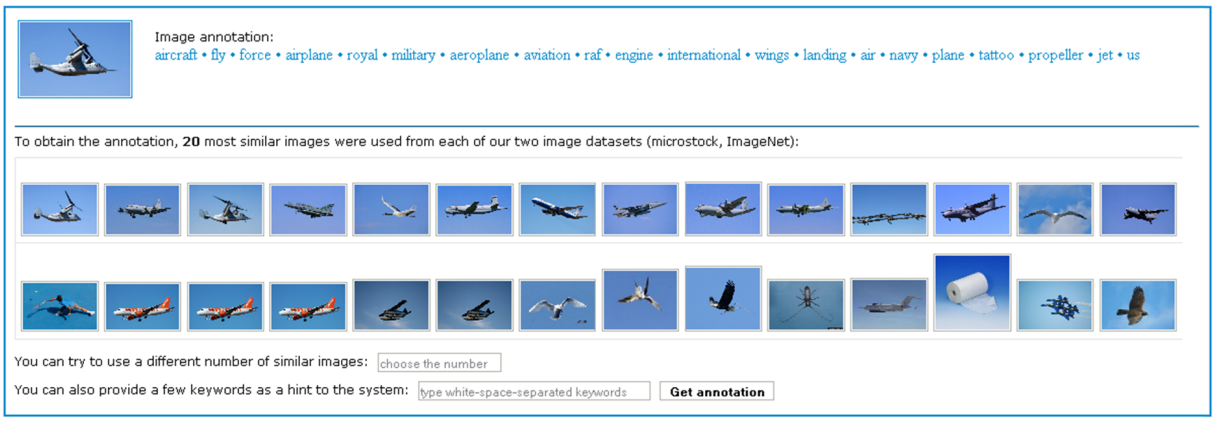

Figure 2: Interface of the MUFIN annotation system

The combination of several heterogeneous data sources brings the following advantages into the annotation process:

- it eliminates the risk of using a closed vocabulary of a specific community that created the data source;
- it allows to verify the results obtained from one data source by the others.

Each of the search subsystems produces a list of the most similar objects, which are in the second phase merged together and a set of candidate keywords is formed. The keywords and the query image can then be repeatedly used as an input for a new search. However, each subsequent search phase increases the time costs of the annotation, therefore result postprocessing methods are often more convenient than new queries. Finally, the keyword set is cleaned from irrelevant word types, names, etc. and displayed to the user.

4. APPLICATION

The demonstrated annotation system uses two different data collections. The first one contains 20 million images from a commercial microstock photo site, each of which is accompanied by a systematic annotation with about 20 keywords in average. As the second dataset we use the ImageNet [1], which is an image database that is being created to illustrate the concepts of the WordNet [2]. The ImageNet subset we use contains about 12 million images describing more than 15,000 WordNet concepts.

Both the datasets are indexed and searched using five MPEG-7 global visual descriptors (see [5] for more details). In the result-processing phase, the frequencies of the keywords from the microstock collection result are computed using the *tf-idf* weighting scheme and the most important ones are selected as the candidate keywords. From the ImageNet collection, we similarly select the most important concepts. Next, the WordNet database is engaged to expand the candidate concepts. The final list of keywords is composed of (1) the candidate keywords that belonged to the candidate concepts, and (2) the synonyms of the qualifying keywords provided by the WordNet.

The web interface of the application (see Figure 2) enables to annotate any web image by providing its URL. If a duplicate image is found in our dataset, e.g. by using an image from the demonstration itself, we do not consider its keywords during the processing to provide a fair annotation that is not influenced by the one that is already given for that image.

Apart from the keywords, the system also shows the k nearest visual neighbors of the query image which were used to obtain the annotation. To test the dependence of the result on the number of nearest neighbors, it is possible to set a different value of k. In addition, users may provide a hint to the annotation system. The hinted keywords are used together with the visual features in the initial search over the microstock collection. The application is available at **http://mufin.fi.muni.cz/annotation/**.

Acknowledgments

This work has been partially supported by the national research projects GAP 103/10/0886, VF 20102014004 and by Brno PhD Talent Financial Aid. The hardware infrastructure was provided by the METACentrum under the programme LM 2010005.

5. REFERENCES

[1] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. In *CVPR09*, 2009.

[2] C. Fellbaum, editor. *WordNet: An Electronic Lexical Database*. The MIT Press, May 1998.

[3] H. Kwasnicka and M. Paradowski. Machine learning methods in automatic image annotation. In *Advances in Machine Learning II*, pages 387–411. 2010.

[4] S. A. Noah, D. A. Ali, A. C. Alhadi, and J. M. Kassim. Going beyond the surrounding text to semantically annotate and search digital images. In *ACIIDS (1)*, volume 5990 of *Lecture Notes in Computer Science*, pages 169–179. Springer, 2010.

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

[6] J. J. Verbeek, M. Guillaumin, T. Mensink, and C. Schmid. Image annotation with tagprop on the MIRFLICKR set. In *Multimedia Information Retrieval*, pages 537–546, 2010.

[7] Y. Yang, Z. Huang, H. Shen, and X. Zhou. Mining multi-tag association for image tagging. *World Wide Web*, 14:133–156, 2011. 10.1007/s11280-010-0099-8.