

Automatic Photo Annotation Based on Visual Content Analysis

Eduard KURIC*

*Slovak University of Technology
Faculty of Informatics and Information Technologies
Ilkovičova 3, 842 16 Bratislava, Slovakia
xkuric@stuba.sk*

Automatic photo annotation is the process by which a computer system automatically assigns metadata to a target photo. With increasing popularity of digital and mobile phone cameras, the need for quick and exact searching grows, for example by general category or focusing on a specific object. Manual creation of annotations is very time-consuming and produces often subjective results thus making automatic photo annotation a very desirable albeit challenging task.

Generally, approaches for automatic annotation are categorized into two scenarios [1]: learning-based methods primary focus on determining complex categories or groups of specific objects and web-based methods that use crawled web image data to obtain relevant annotations.

In learning-based methods, a statistical model is built to teach a classifier. Automatic face recognition in photos is good a example of automatic annotation of specific objects. One possibility is that a retrieval process uses a robust dictionary of visual terms to identify people. Similarity can be evaluated by comparing local descriptors which are computed over local features such as edges, small patches around points of interest. The local descriptors are much more precise and discriminating than global descriptors. For searching for specific objects, this feature is useful, but by searching complex categories it can be an obstacle. Another obstacle is the need to store the huge number of the extracted features.

While the Web provides unlimited vocabulary for web-based methods, its problem is the initial query and the lack of information about target photos. Without providing information such as key caption, searching for similar photos to a target photo on the Web is like finding a needle in a haystack. Additional important drawbacks include performance and noise in obtained annotations.

In our method (Figure 1), we combine local and global features to retrieve the best results. For detection and extraction of local features, we use Scale Invariant Feature Transform (SIFT). To computation of the global descriptors, we use Joint

* Supervisor: Mária Bielíková, Institute of Informatics and Software Engineering

Composite Descriptor (JCD). With their combination, we are able to ensure robustness and generalization needed for complex queries.

We place great emphasis on real-time performance. To cope with the huge number of extracted features, we implemented disk-based locality-sensitive hashing to index descriptors. By searching for similar candidates during extraction of keywords, we focus on photos analysis in terms of probability, that the retrieved photos contain the right keywords for the target photo. For example, we prefer photos where extracted objects of interest from the target photo are dominant in retrieved photos or their frequency of occurrence is greater.

Our method is designed for versatile use, e.g. identifying key objects in specific photo albums; complex automatic photo annotation in large web photo galleries; searching similar photos according to objects of interest (query by image content).

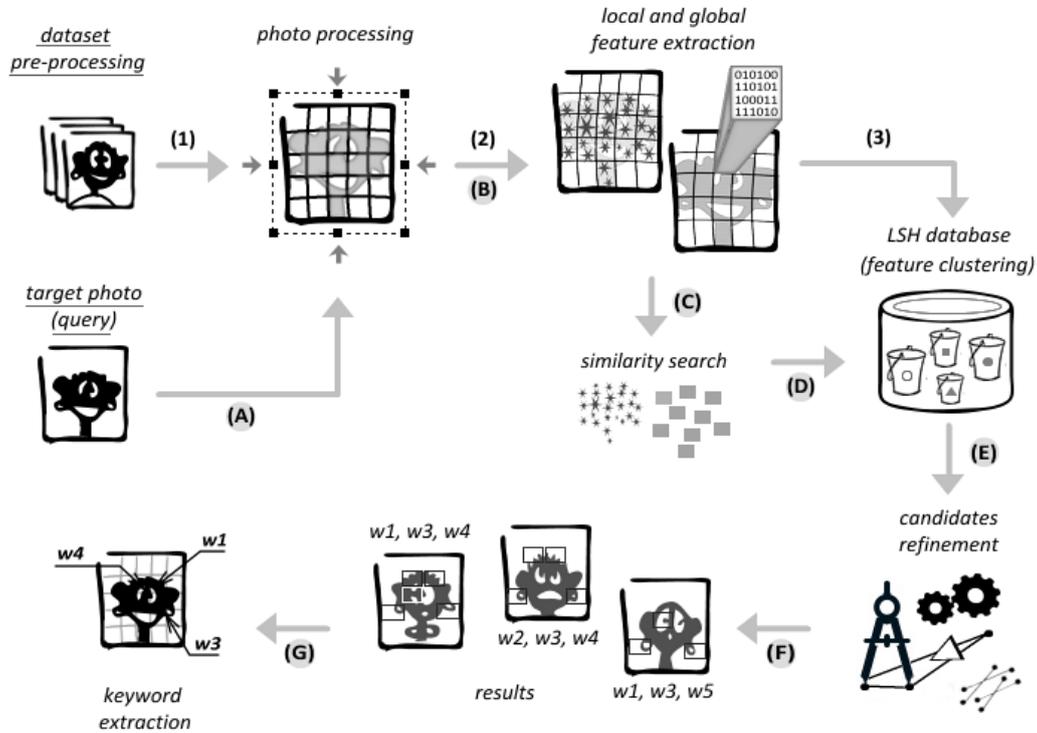


Figure 1. A scheme of our annotation method consists of two independent phases, namely dataset pre-processing (1) – (3) and processing of a target photo (A) – (G).

Acknowledgement. This work was partially supported by the Scientific Grant Agency of Slovak Republic, grant No. VG1/0508/09.

References

- [1] Wang, X., et al.: AnnoSearch: Image Auto-Annotation by Search. In *Proc. of the IEEE CS Conf. on Computer Vision and Pattern Recognition – Vol. 2, CVPR '06*, Vol. 2. IEEE CS, Washington, pp. 1483-1490, 2006.