Interactive Photo Retrieval based on Semi-Automatic Annotation using Visual Content and Folksonomies

Eduard KURIC*

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

Nowadays web photo management systems provide features for users, such as organizing, sharing and searching photos. With increasing popularization of digital and mobile phone cameras, there occurs a need of quick and exact searching. Content based indexing of photos is more difficult than text documents because the photos do not contain units like words. Searching is based on annotations and semantic keywords that are entered by a user and associated with photos. However, manual creating of annotations is very time-consuming and results are often subjective. Therefore, photo semi-automatic annotation is most challenging task.

Traditional approaches for semi-automatic annotation are based on combining keyword-based and content-based photo retrieval [3]. The user enters a query consisting of a target photo and keywords, typically only a caption. The aim is to find most similar photos and to extract related keywords. After a retrieval process, the user selects the best relevant keywords and associates them with the target photo. The process usually takes place in three steps. First, a keyword-based technique is used to obtain a list of candidate photos that are also associated with the input caption. Second, content-based photo retrieval technique is used to assemble a ranked list of visually related photos. Finally, a method is used to combine the ranked list into an annotation list which represents keyword proposals. These solutions employ global low-level features like color and texture for a content comparison of photos. However, the user query can include a full photo or a just part of the whole photo which we call object-of-interest.

In our work, we propose a novel method for annotating photos which extends existing solutions of searching similar photos primary according to objects-of-interest [2]. Often, those objects represent a foreground of a photo that is in comparison with a background less dominant. Therefore, in traditional approaches of content-based photo

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

Spring 2010 PeWe Workshop, April 18, 2010, pp. 7–8.

retrieval foregrounds can be ignored (low-rated) despite of the fact that can represent most important elements of the photo. We use an interactive photo segmentation to determine objects-of-interest. To capture local photo information that is in object retrieval essential, we use scale invariant feature transform (SIFT) in combination with a hash-based method known as locality sensitive hashing [1].

A SIFT detector transforms a photo into a collection of feature points that are invariant to photo scaling, translation, rotation and partially to illumination changes. Such photo can be viewed as a bag-of-feature-points and any object in the photo is represented as a subset of the points (local descriptors). Unfortunately, with such representation, there arises a many-to-many matching problem in a high-dimensional space because the photo typically contains from hundreds to thousands of feature points. Therefore, our proposed solution provides the interactive photo segmentation whereby a user can select a subset of feature points of a target photo instead of a whole set. The query subset represents objects-of-interest and remaining points of the target photo are used to a refinement. A group of points of the subset can be associated with keywords and consequently each of database photos can contain exactly named objects-of-interest.

Our annotation process takes place in four steps. First, a system creates a candidate list consisting of database photos in which each one contains the same objects as the query. Second, the candidate list is refined by comparison with remaining points. Third, from the list there are gathered and ranked all named objects of which is created an annotation list (keyword proposals). Finally, other keywords associated with the candidates are ranked and combined with the annotation list.

By reason of using the local descriptors, the solution is high resistant to cropping and other common transforms against approaches based on global features. Our proposed solution does not require input caption but in the case of insufficient results allows extending input query of the keywords. Thus, our solution allows identifying objects in the photo and using relevance feedback user can improve performance of our content-based photo retrieval.

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

References

[1] Ke, Y., Sukthankar, R., and Huston, L. 2004. An efficient parts-based near-duplicate and sub-image retrieval system. In *Proc. of the 12th Annual ACM Int. Conf. on Multimedia* (New York, NY, USA, 2004). New York, NY, pp. 869-876.

[2] Kuo, Y., Chen, K., Chiang, C., and Hsu, W. H. 2009. Query expansion for hash-based image object retrieval. In *Proc. of the 17th ACM Int. Conf. on Multimedia* (Beijing, China, 2009). ACM, New York, NY, pp. 65-74.

[3] Wang, X., Zhang, L., Jing, F., and Ma, W. 2006. AnnoSearch: Image Auto-Annotation by Search. In *Proc. of the 2006 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition – Vol. 2* (2006). IEEE Computer Society, Washington, DC, pp. 1483-1490.