

SBIA: Search-based Image Annotation by Leveraging Web-Scale Images*

Xirong Li¹, Xin-Jing Wang³, Changhu Wang², Lei Zhang³

¹Department of Computer Science & Technology, Tsinghua University, Beijing 100084, China

²Department of EEIS, University of Science and Technology of China, Hefei 230027, China

³Microsoft Research Asia, 49 Zhichun Road, Beijing 100080, China

lxr@mails.tsinghua.edu.cn, wch@ustc.edu

{i-xinjaw, leizhang}@microsoft.com

ABSTRACT

In this technical demonstration, we showcase the SBIA system – a search-based image annotation system. At the heart of the system lies a very large-scale image search engine which indexed three million Web images and supports both text and visual queries. Given an image (with initial annotations), SBIA first finds semantically/visually similar images via the search engine, and then mines representative keywords from the retrieved images. These keywords, after annotation rejection and relevance ranking, are finally used to annotate the query image.

Categories and Subject Descriptors

I.4.8 [Image Processing and Computer Vision]: Scene Analysis – Object Recognition, H.2.8 [Database Management]: Database Application – Image databases.

General Terms

Algorithms, Experimentation, Performance

Keywords: Image Annotation, Search, Clustering, Rejection

1. INTRODUCTION

Automatic image annotation (AIA) has been an active research topic in recent years due to its potentially large impact on both image understanding and Web image search.

Most previous AIA models, borrowed from either Machine Learning or Information Retrieval fields, focus on learning projections or correlations (e.g. translation, joint distribution, image classification as conditional probability estimation) between images and words given a number of training images [4].

However, compared with the potentially unlimited vocabulary existing in a Web-scale image database, only a very limited number of concepts can be modeled on a small-scale image database. Moreover, due to the data sparseness in the high-dimensional visual feature space, a large amount of training samples are required to ensure a reasonable accuracy, which is known as “the curse of dimensionality”. Therefore, collecting and leveraging large-scale training data in an effective and efficient way is becoming a key challenge to conquer the AIA problem.

By leveraging numerous Web pages consisting of images and their descriptions (e.g. titles, anchor texts, and surrounding texts) search based approaches seem to be a promising way to overcome the challenge. Motivated by Web search techniques in many commercial systems, we have proposed the AnnoSearch system [1]. Assuming that an accurate keyword of the image to annotate is available, the keyword is used to retrieve a set of semantically relevant images. The resulting images are further re-ranked based on visual features to ensure visual coherence. A search result clustering (SRC) algorithm [5] is then adopted to mine annotations from the top ranked images. Although the initial keyword might speed up the search process and improve the search result quality, it may not always be available in real environment, especially for the personal images.

We later proposed an upgraded version to relax the initial-word assumption and to efficiently exploit large-scale images [2]. In the search stage, we directly perform an approximate K -NN search in the visual feature space, which is facilitated by high-dimensional indexing techniques. Further, considering that SRC was originally designed for general Web search and might not be optimal for annotation mining, we utilize the $tf-idf$ ¹ principle to rank annotations for each retrieved image, and fuse these ranked lists into the final annotation list [3].

In this demo, we present the SBIA system – a search-based annotation system which integrates the work in [1,2,3]. We improve upon the previous systems in several ways: i.e., support for multi-modality retrieval, annotation relevance ranking based on SRC results and query-independent ranks, and a rejection scheme to control the annotation quality which is the main contribution in this work.

2. ANNOTATION SYSTEM

2.1 Framework

The intention of image annotation is to find a keyword set \mathbf{w}^* that maximizes the conditional probability $P(\mathbf{w}|I_q)$, where \mathbf{w} are keywords in the vocabulary and I_q the image to annotate. In an ideal case where there exists a well-annotated and unlimited-scale image database, then for any query image, we can find its duplicates in this database and thus solve the above optimization problem simply by annotation propagation: $\mathbf{w}^* \leftarrow \mathbf{w}_i$, where

* This work was performed at Microsoft Research Asia.

¹ $tf-idf$ (term frequency – inverse document frequency) is the best known term weighting scheme in the Information Retrieval area.

image J is the duplicate of I_q and w_j the annotation of J . In a more realistic case where the database is of limited yet very large-scale, we can still find a group of very similar images in terms of semantic and/or visual measurement.

The annotation process is composed of three main steps:

1. **Searching similar images.**
2. **Mining representative keywords from the retrieved images as annotation candidates.**
3. **Annotation Rejection & Relevance Ranking.**

The system framework is shown in Figure 1. One merit of the system is its ability to exploit numerous rapid-growing Web resources effectively and efficiently.

2.2 Learning to Predict Annotation Quality

It is worth noting that global features are helpful for image-level concept annotation, but are ineffective in object-level annotation [2]. It is thus necessary to develop a scheme to estimate the annotation quality. Viewing each image in the database as a K -NN classifier, for one image, we obtain its k visually similar images and measure their semantic relevance based on their annotations. The higher the relevance is, the better the annotation quality might be. Besides, from the language modeling perspective, the search result quality can be characterized to some extent by the model divergence between the relevant document set and the whole collection. Bearing these in mind, we formalize annotation rejection as a regression problem,

$$f: (x_1, x_2, \dots, x_n) \rightarrow y, y \in \{acceptance, rejection\}$$

where $x_i, i=1, \dots, n$ are features extracted from the SRC results and f the regression model. We will detail the algorithm elsewhere.

2.3 Annotation Relevance Ranking

Analogous to ranking web pages, the ranking scores of candidate annotations can also be decomposed into query-dependent ones which are determined by the query image and query-independent ones which can be calculated beforehand. In this implementation, we combine the *tf-idf* weighting strategy in [3] and the prediction

results (see Sec. 2.2) to rank phrases generated by SRC and select the top ones as final annotations.

2.4 Implementation

The SBIA system indexed 3 million Web images. These images were crawled from several photo forum sites, because images in photo forums have rich and less noisy descriptions provided by photographers. A 64-dimensional global feature [6] is extracted, and the K-means clustering algorithm is used to index the 64-D feature for the sake of simplicity.

The system was implemented on a Web server with two 2.0GHz CPUs and 2GB memory. It can efficiently find similar images from the 3 million image database within 20 milliseconds, and annotate the query image within 50 milliseconds on average.

3. ACKNOWLEDGMENTS

We would like to thank Shuo Wang for designing the UI.

4. REFERENCES

- [1] Wang, X., Zhang, L., Jing, F., and Ma, W. AnnoSearch: Image Auto-Annotation by Search. In *Proc. of CVPR*, 2006.
- [2] Li, X., Chen, L., Zhang, L., Lin, F., and Ma, W. Image Annotation by Large-Scale Content-based Image Retrieval. In *Proc. of ACM Multimedia*, 2006.
- [3] Wang, C., Jing, F., Zhang, L., and Zhang, H. Scalable Search-Based Image Annotation of Personal Images. In *Proc. of ACM MIR*, 2006.
- [4] Barnard, K., Duygulu, P., Freitas, N., Forsyth D., Blei, D., and Jordan, M. Matching Words and Pictures. *JMLR*, 2003.
- [5] Zeng, H., He, Q., Chen, Z., Ma, W., and Ma, J. Learning to cluster web search results. In *Proc. of SIGIR*, 2004.
- [6] Zhang, L., Hu, Y., Li, M., Ma, W., and Zhang, H. Efficient Propagation for Face Annotation in Family Albums. In *Proc. of ACM Multimedia*, 2004.

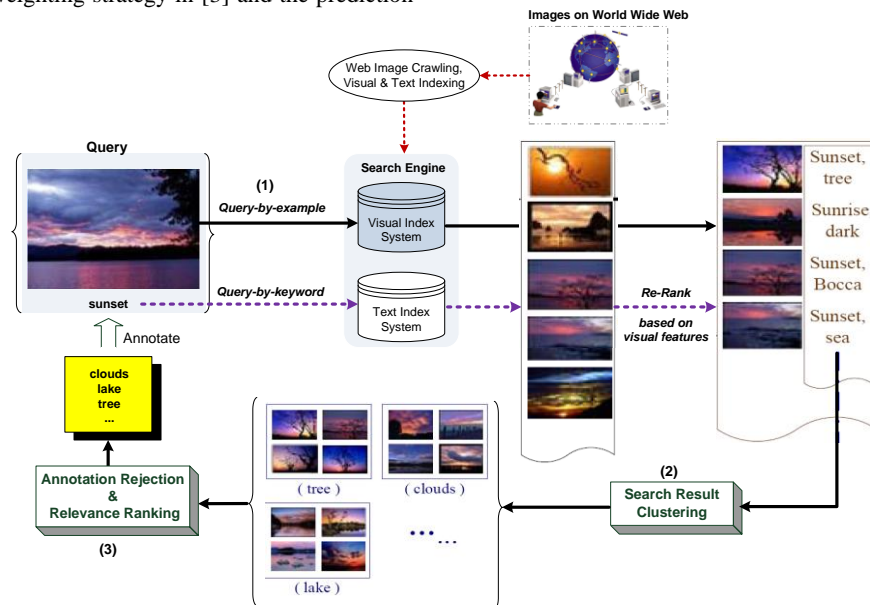


Figure 1. Framework of the SBIA system