Information Sciences 358–359 (2016) 151–163

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Combined retrieval: A convenient and precise approach for Internet image retrieval

Kehua Guo[a], Ruifang Zhang[a], Zhurong Zhou[b], Yayuan Tang[a], Li Kuang[c,*]

[a] School of Information Science and Engineering, Central South University, Changsha 410083, China
[b] School of Computer Science and Information, Southwest University, Chongqing 400715, China
[c] School of Software, Central South University, Changsha 410083, China

ARTICLE INFO

Article history:
Received 4 May 2015
Revised 25 March 2016
Accepted 3 April 2016
Available online 22 April 2016

Keywords:
Internet image retrieval
Retrieval intention
User experience
Perceptual hash
Image search engine

ABSTRACT

In Internet image retrieval, returned results may fail to satisfy the retrieval intentions of users because of noisy annotations. Solving the ambiguity in image retrieval by combining text features and visual information has been a challenging problem. In this paper, we propose a convenient and precise approach for Internet image retrieval called combined retrieval (CR), which costs minimized extra feedback to retrieve more results reflecting the query intentions of users. CR is used as a plug-in to commercial image search engines, such as Google and Bing, which are defined as host image search engines (HISE). First, in the returned result from HISE, document analysis is utilized to construct the image categories based on the Wikipedia categorical index. Returned images will be automatically categorized, and a convenient interface is provided for user feedback. Second, we describe the re-retrieval algorithm in which image data combined with particular text information will be sent to the HISE for re-retrieval. Finally, a perceptual hash based re-rank algorithm to optimize the returned images is proposed. Experimental results indicate that CR can significantly improve the retrieval performance with minimum effort and can provide a notably convenient user experience.

1. Introduction

With the explosive growth of Internet images, effective retrieval techniques are urgently required. Many image retrieval approaches have been extensively used to search for massive amounts of Internet images. For example, people can use search engines such as Google and Bing to easily find images from the Internet; however, it still has been difficult to retrieve images that satisfy the query intentions of users.

The keyword-based approach is commonly used in traditional image retrieval. After submitting the query, images that contain the keywords in the surrounding text will be returned by the retrieval engine. However, there are many notable disadvantages in the keyword-based approach, as follows: (1) keywords cannot describe the image content and the intentions of users accurately. For example, the word 'Apple' may mean a type of fruit, a famous company, or a brand of computers and phones. (2) Users may not have enough sufficient knowledge to accurately express the query intention, and the keywords tend to be too short to describe the image content. Hence, the ambiguity of keywords may lead to a result with considerable noise, and users have to spend more time finding the images they really want.

* Corresponding author. Tel.: +86 731 88830212.
E-mail address: corrine.li2009@gmail.com, kuangli@csu.edu.cn (L. Kuang).

http://dx.doi.org/10.1016/j.ins.2016.04.001
0020-0255/

To solve the problems above, many approaches have been taken to capture the intentions of users. The most commonly used approach is to employ linguistically related methods to find synonyms or relevant text with the query keywords and provide more detailed text keyword options. However, the query intention is highly diverse, and users may have difficulty finding the images that truly reflect their intentions because of keyword option limitations. For example, "Related Searches" in the Google image search engine only provides six options. Users may have to attempt several times to obtain the final result.

The image-based approach is another traditional image retrieval method. Visual features are extracted for image retrieval in this approach. After the user uploads an image, the search engine searches the images that are similar to it using content-based image retrieval (CBIR) algorithms [24,6]. Because feature extraction costs computation resources, it is important to find a balance between the accuracy of feature extraction and the computation time. Moreover, the personal understanding of users is ignored in CBIR algorithms, and similar results satisfying the intentions of users cannot be retrieved. Although we can add user feedback to CBIR approaches [23], users will suffer from being asked for too much substantial feedback.

The third approach is semantic-based image retrieval method [27,20]. In this approach, semantic annotations of Internet images are described by ontology and stored in a database, and the engine will execute the retrieval based on the annotations. However, annotating the images on the Internet is a heavy workload, and it cannot be guaranteed that all of the images have accurate annotations. In addition, if the images are copied or moved to another place, rebuilding the semantic information is complex work.

To solve the problems mentioned above, in this paper, we propose a novel Internet image retrieval approach called combined retrieval (CR). The characteristics of CR are as follows: (1) convenient interface. CR runs as a plug-in to HISE (such as Google and Bing) and provides retrieval patterns similar to them. Thus, the input data to CR is based on the result from HISE, and the interface gives the users an extra option for obtaining the preferred images; (2) full intention analysis. CR analyzes the images' surrounding text to automatically categorize the returned images and display the categories to roughly capture the retrieval intentions of users; (3) minimum effort. Users are required to participate in only one extra feedback session to retrieve results and obtain the intention-reflected result. The retrieval performance and convenient user experience of CR will be reported in the experimental result.

The rest of this paper is organized as follows: Section 2 presents the related work of Internet image retrieval. Section 3 describes the overview and key technologies of CR, including image categorization, document analysis, user feedback, and the re-retrieval process. Section 4 shows the prototype implementation, performance evaluation model, and experimental results. Finally, conclusions and future work are presented in Section 5.

2. Related work

Researchers have proposed many approaches to solve the problem that keywords cannot closely express the query intentions of users. The most representative method is text-based keyword expansion. In the expansion, keywords are generally expanded with their linguistically related words [17] to ensure the textual description of the query in more detail [26]. In addition, some typical algorithms [10,11] provide tag suggestions or annotations in the image retrieval procedure. However, there are several disadvantages in the keyword expansion approach for Internet image retrieval. First, this approach requires supervised training and costs considerable computation resources. Second, it cannot be guaranteed that the expanded list contains the desired keywords for users.

In addition to text-based image retrieval approaches, CBIR approaches are widely employed for Internet image retrieval. The fundamental idea of CBIR is to extract the visual features, which can be divided into two categories: global and local. A representative global feature is GIST [30], which is a low-dimensional global image representation that provides relevant information for place recognition and categorization. The other feature is HoG [9], which represents edge distributions over different parts of images. A typical local feature descriptor is SIFT [21], which provides the underlying image patch descriptor for matching scale invariant key points [26]. However, the intentions of users are not just based on low-level visual features, such as color, texture, and shape [32].

To solve the problem, several researchers combined CBIR approaches with user feedback to improve the retrieval performance. Pinjarkar et al. [23] presented comprehensive surveys for user feedback in image retrieval. Su et al. [25] collected the navigation data and visual features in the CBIR system and adjusted them to adapt to the query intentions of users. Zhang et al. [37] refined the discriminant features from a large feature set for relevance feedback based on the biased discriminant analysis (BDA) approach. Babenko et al. [3] investigated the use of neural codes within image retrieval and illustrated the performance of neural codes by quantitative experiments. Paulin et al. [22] proposed an unsupervised framework called Patch-CKN (Convolutional Kernel Network) as an alternative to SIFT for CBIR and achieved good results. However, the CBIR approaches increase the burden to users. To gain a better result, users have to spend considerable time on giving relevance feedback to the result, which significantly reduces the quality of user experience [38].

In the field of semantic-based retrieval, topic models, such as probabilistic latent semantic analysis (PLSA) [28] and latent Dirichlet allocation (LDA) [4], are widely used in text semantic extraction. The bag-of-words (BoW) model [5] is a typical model to express visual words. Semantic representations are generally based on the object ontology method [12,13,1]. Gutiérrez et al. [14] investigated the ontology expression to virtual humans, covering the features, functions, and skills of a human. Wang et al. [31] used semantic web rule language to define semantic matching. Yang et al. [36] proposed a hierarchical ontology-based knowledge representation model. These approaches can support user intention-reflected retrieval,