A Study on Improving Annotation Based Image Retrieval

陳, 華

https://doi.org/10.15017/1931930

出版情報：九州大学, 2017, 博士（工学）, 課程博士
バージョン：
権利関係：

（別紙様式 2）

氏　　名　：陳　華　(Chen Hua)

論 文 名　：A Study on Improving Annotation Based Image Retrieval
（アノテーションに基づく画像検索の改善に関する研究）

区　　分　：甲

論　文　内　容　の　要　旨

With the growing use of computers, mobile devices and other digital products, a huge amount of digital images is being generated every day. Consequently, how to quickly and accurately find relevant images has become a hot research topic.

There are many image retrieval solutions have been introduced to browse and retrieve images from large image databases. These solutions are typically classified into two categories: content based image retrieval (CBIR) solutions and annotation based image retrieval (ABIR) solutions. Basically, CBIR solutions use visual features (such as color, texture, or shape) to retrieve images. With these solutions, users can submit a query image to find similar images from a collection of images. However, because the visual features are often low-level and cannot accurately represent the complicated semantics of images, the CBIR solutions suffer from the semantic gap problem. On the other hand, ABIR solutions use textual descriptions as image metadata and find images with text retrieval techniques. As can be seen in many of today's image retrieval systems, the ABIR solutions are considered more practical (e.g., Flickr, Google and Bing). However, the results of previous image retrieval systems are often inaccurate for semantically complex queries. For example, when searching images with a query "two animals run on grass", retrieval systems cannot understand the exact meaning of this query and cannot return accurate results. Therefore, it is necessary to find better solutions to solve this problem and then improve the results of image retrieval.

By analyzing the previous image retrieval systems, the author finds that natural language processing (NLP) technologies are not well leveraged for image retrieval. For example, the object counts are often represented as general attributes of objects, but not the actual numbers of objects. Therefore, previous solutions cannot support image retrieval by counting objects. However, recent approaches in NLP field can easily deal with counting problems by using concepts and instances to represent information.

This thesis aims at improving ABIR for complex natural language queries by using NLP technologies, Semantic Web technologies and other related technologies. This thesis first proposes an image annotation model and introduces how to use it to address the counting problem in image retrieval. Then, the thesis presents an ABIR system based on the proposed image annotation model, which can support image retrieval for semantically complex queries. Finally, two techniques are proposed to improve the accuracy of image retrieval for complex queries.

The main contributions of this thesis are as follows.

1. This thesis proposed an image annotation model based on the notions of concept and instance. With this model, the properties of objects, the relations of objects, and the numbers of objects can be expressed in

RDF (Resource Description Framework) to describe the detail information of images. Because this model contains the numbers of objects in images, it can support image retrieval by counting objects. Besides, SPARQL (SPARQL Protocol and RDF Query Language) provides the ability to query RDF triples with counting expression; with this feature, it can easily and effectively find images by counting objects.

2. This thesis designed an ABIR system based on the proposed image annotation model for complex queries by using a semantic parser (Knowledge Parser or K-Parser). From text written in natural language, the K-Parser extracts a graphical semantic representation of the objects involved, their properties as well as their relations. Both the image textual descriptions and the natural language queries are analyzed by the K-Parser. As a technical solution, RDF is leveraged in two ways: (1) the parsed image captions are stored as RDF annotations; (2) image queries are translated into SPARQL queries to query RDF annotations for image retrieval. When applied to the Flickr8k dataset with a set of custom queries, the K-Parser exhibits some biases that negatively affect the accuracy of the queries.

3. Two optimization techniques were proposed to address the weaknesses mentioned above: (1) a set of rules were introduced to transform the output of K-Parser and fix some basic, frequent mistakes that occurred on the captions of Flickr8k dataset; and (2) commonsense knowledge databases (e.g., ConceptNet) were leveraged to raise the accuracy of queries on broad concepts.

Two experiments were conducted with the proposed ABIR system: (1) searching images with a set of custom complex queries; and (2) searching images by counting objects. The experimental results showed that the proposed approach achieved highly satisfactory accuracy with complex queries. Especially, it excelled in advanced domains such as counting objects in images (e.g., “two animals run on grass”), finding images with fuzzy description queries (e.g., “animal wearing something runs”).