

Chapter 4.17

Annotating Images by Mining Image Search

Xin-Jing Wang

Microsoft Research Asia, China

Lei Zhang

Microsoft Research Asia, China

Xirong Li

Microsoft Research Asia, China

Wei-Ying Ma

Microsoft Research Asia, China

ABSTRACT

Although it has been studied for years by computer vision and machine learning communities, image annotation is still far from practical. In this chapter, the authors propose a novel attempt of modelless image annotation, which investigates how effective a data-driven approach can be, and suggest annotating an uncaptioned image by mining its search results. The authors collected 2.4 million images with their surrounding texts from a few photo forum Web sites as our database to support this data-driven approach. The entire process contains three steps: (1) the search process to discover visually and semantically similar search results;

(2) the mining process to discover salient terms from textual descriptions of the search results; and (3) the annotation rejection process to filter noisy terms yielded by step 2. To ensure real time annotation, two key techniques are leveraged – one is to map the high dimensional image visual features into hash codes, the other is to implement it as a distributed system, of which the search and mining processes are provided as Web services. As a typical result, the entire process finishes in less than 1 second. Since no training dataset is required, our proposed approach enables annotating with unlimited vocabulary, and is highly scalable and robust to outliers. Experimental results on real Web images show the effectiveness and efficiency of the proposed algorithm.

DOI: 10.4018/978-1-60960-818-7.ch4.17

INTRODUCTION

The number of digital images has exploded with the advent of digital cameras which requires effective image search techniques. Although it is an intuitive way of image search since “a picture worths a thousand word”, currently the Query-By-Example (QBE) scheme (i.e. using images as queries) is seldom adopted by commercial image search engines. The reasons are at least twofold: 1) the semantic gap problem. It is also the fundamental problem in Content-based Image Retrieval (CBIR) field. The techniques to extract low-level features, such as color, texture, and shape etc., are far from powerful to represent the semantics contained in an image, and hence given an image, its search results may be quite different conceptually to the query, although they possess same chromatic or textural appearances. 2) Computational expensiveness. It is well known that the inverted indexing technique ensures the practical and successful usage of current text search engines. This technique uses keywords as entries to index documents containing them, which also aligns seamlessly with the Query-by-Keyword (QBK) scheme adopted by text search engines. Thus given a number of query keywords, the search result can just be the intersection of documents indexed by these keywords separately (if no ranking functions applied). However, in the image search case, since images are 2D media, and the spatial relationship between pixels is critical in conveying the semantics of an image, how to define image “keyword” is still an open question. This prevents the inverted indexing technique from being directly used in image search, which creates a critical efficiency problem of using image visual features to search.

Due to the above reasons, there is a surge of interests on image auto-annotation and object recognition in recent years. Researchers try to define ways to automatically assign keywords onto images or image regions. While all the previous work build learning models to annotation images,

in this chapter, we attempt to investigate how effective a modeless and data-driven approach, which leverages the huge amount of commented image data in the Web, could be.

This is reasonable because (1) to manually label images is a very expensive task (Naphade, Smith, Tesic, Chang, Hsu, Kennedy, Hauptmann, & Curtis, 2006), and in many cases, different people tend to have different explanations on a single image. Such inconsistency even makes image labeling a research topic that many questions should be addressed: What kind of images can be labeled consistently? What is the strategy of labeling images that can ensure the consistency, etc? All these obstacles lead to the lack of training data, which discourages researchers from learning a practically effective annotation model even if we have such a huge image data set in the Web. (2) Some researchers proposed very efficient and effective image encoding approaches (Fan, Xie, Li, Li, & Ma, 2005) which convert an image into an N-bit hash code so that image visual distance can be measured simply by the Hamming distance between the hash codes. This enables large-scale content-based image search in real-time and set a light on combining image visual appearance into the commercial search engines that are purely based on text search since both image indexing (e.g. matching the first n bits of two hash codes) and retrieval are of $O(1)$ complexity.

Imagine that an image set that is large enough is available, so that for each query image at least one duplication can be found in the image set, and then what we need to do is just to annotate the query by the comments or surrounding texts of the duplication. As analyzed above, this is applicable in real-time since each image is converted into a hash code so that large-scale image indexing is enabled.

This motivated us to leverage Web data and crawl Web images for a data-driven image annotation approach since the Web not only deposits huge amount of images but also they generally have human assigned comments or descriptions as

22 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:
www.igi-global.com/chapter/annotating-images-mining-image-search/56187

Related Content

A Novel Cross Folding Algorithm for Multimodal Cancelable Biometrics

Padma P. Pauland Marina L. Gavrilova (2012). *International Journal of Software Science and Computational Intelligence* (pp. 20-37).

www.irma-international.org/article/novel-cross-folding-algorithm-multimodal/76268

Distributed Intelligence for Constructing Economic Models

Ting Yu (2012). *Machine Learning: Concepts, Methodologies, Tools and Applications* (pp. 1538-1550).

www.irma-international.org/chapter/distributed-intelligence-constructing-economic-models/56212

LVQ Neural Networks in Color Segmentation

Erik Cuevas, Daniel Zaldivar, Marco Perez-Cisnerosand Marco Block (2010). *Soft Computing Methods for Practical Environment Solutions: Techniques and Studies* (pp. 45-63).

www.irma-international.org/chapter/lvq-neural-networks-color-segmentation/43144

Cooperative Encoding Strategy for Gate Array Placement

Hong-Bo Wang, Qing-Dong Suand Ruolei Zeng (2018). *International Journal of Software Science and Computational Intelligence* (pp. 29-43).

www.irma-international.org/article/cooperative-encoding-strategy-for-gate-array-placement/223493

Leveraging AI for Human Rights in Digital Security

Sudhir Kumar Dwivedi, Prashant Pandey, Mudit Rohilla, Pankaj Dwivediand Saquib Ahmed (2026). *Cross-Sector Cyber Insurance for the Intelligent Society* (pp. 223-250).

www.irma-international.org/chapter/leveraging-ai-for-human-rights-in-digital-security/387641