Chapter 7
Collaborative Bayesian Image Annotation and Retrieval

Rui Zhang
Ryerson University, Canada

Ling Guan
Ryerson University, Canada

ABSTRACT
With nearly twenty years of intensive study on the content-based image retrieval and annotation, the topic still remains difficult. By and large, the essential challenge lies in the limitation of using low-level visual features to characterize the semantic information of images, commonly known as the semantic gap. To bridge this gap, various approaches have been proposed based on the incorporation of human knowledge and textual information as well as the learning techniques utilizing the information of different modalities. At the same time, contextual information which represents the relationship between different real world/conceptual entities has shown its significance with respect to recognition tasks not only through real life experience but also scientific studies. In this chapter, the authors first review the state of the art of the existing works on image annotation and retrieval. Moreover, a general Bayesian framework which integrates content and contextual information and its application to both image annotation and retrieval are elaborated. The contextual information is considered as the statistical relationship between different images and different semantic concepts for image retrieval and annotation, respectively. The framework has efficient learning and classification procedures and the effectiveness is evaluated based on experimental studies, which demonstrate its advantage over both content-based and context-based approaches.

INTRODUCTION
Ever-lasting growth of multimedia information has been witnessed and experienced by human beings since the beginning of the information era. An immediate challenge resulting from the information explosion

DOI: 10.4018/978-1-61692-859-9.ch007
Collaborative Bayesian Image Annotation and Retrieval

is how to intelligently manage and enjoy the multimedia databases. In the course of the technological development of multimedia information retrieval, various approaches have been proposed with the ultimate goal of enabling semantic-based search and browsing. Among those intensively explored topics, content-based image retrieval (CBIR), born at the crossroad of computer vision, machine learning and database technologies, has been studied for more than a decade, yet still remaining difficult (Smeulders, Worring, Santini, Gupta, Jain, 2001), (Datta, Joshi, Li, Wang, 2008). In a nutshell, the content-based approaches to image retrieval primarily rely on the pictorial information, a.k.a. low level visual features such as color, texture, shape and layout, which can be automatically extracted from images for similarity measure. The essential challenge is that the low level visual features accurately characterizing the semantic meaning of images are difficult to discover. Therefore, semantically relevant images may be located far away from each other in the space of the pictorial information. To reduce the gap between the high level semantics and low level features, human knowledge was expected to help refine the representation of the semantic meaning in a user’s query. To this end, the relevance feedback (RF), a technique originally proposed for traditional document retrieval, was adapted to solve the problem of image retrieval (Crucianu, Ferecatu, Boujemaa, 2004), (Zhou, Huang, 2003). To enable more efficient search within a large scale database, content-based image classification has been proposed for structured indexing. Existing along with the advantages of content-based approaches is the inherent difficulty in terms of the query formulation based on representation completely different from the human language. To human beings, identifying discriminative visual features for expressing high level semantic meaning, such as someone’s first day at a university or the most exciting scene in a movie, is a fairly difficult task. Therefore, automatic image annotation aiming at constructing the correspondence between visual features and textual words has also been intensively studied. After so many years of research on the above-mentioned topics, it can be identified that none of the individual modalities, e.g. visual content, text, and metadata such as time, location, etc, is sufficient for effectively accomplishing the goal. Therefore, it has become a leading trend (Guan, Muneesawang, Wang, Zhang, Tie, Bulzacki, & Ibrahim, 2009) within the research community to collect and integrate the sources of information with respect to the modalities that are distinct from and complementary to each other, which is known as the information fusion. In fact, as long as we believe that human beings are incomparable in terms of recognizing high level semantics, information fusion is indeed a promising direction not only for the topics covered herein but also for the general domain of pattern classification, as human beings are the most proficient user of the synergy across distinct modalities.

While research on image annotation and retrieval involves the effort from many different aspects, such as computer vision, machine learning, and even psychology, each of which has its own active research frontier, we intend to devote this chapter to a review of the representative works on image annotation and retrieval as well as the elaboration of our recent trend-aligned endeavor regarding this topic. The goal is to provide a general picture on the state of the art in order to draw the attention from the prospective readers and motivate more interesting ideas and development in the research community. The rest of this chapter is organized as follows. In section II, landmark achievements after the year of 2000 in the related research fields are reviewed. For those developed in 1990s, readers are referred to the literature mentioned in the section of introduction. In section III, we focus on a general Bayesian framework and its application to both image annotation and retrieval. Following the elaboration of the Bayesian framework, we continue with further discussion on the experimental results. Finally, the chapter is concluded with a summary including some future research directions.
22 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the product's webpage: www.igi-global.com/chapter/collaborative-bayesian-image-annotation-retrieval/49107?camid=4v1

This title is available in InfoSci-Books, InfoSci-Intelligent Technologies, Business-Technology-Solution, Science, Engineering, and Information Technology, InfoSci-Computer Science and Information Technology, InfoSci-Select. Recommend this product to your librarian: www.igi-global.com/e-resources/library-recommendation/?id=1

Related Content

Mass Media Strategies: Hybrid Approach using a Bioinspired Algorithm and Social Data Mining
www.igi-global.com/chapter/mass-media-strategies/64928?camid=4v1a

Image Retrieval based on HSV Feature and Regional Shannon Entropy
Liang Lei, Jun Peng and Bo Yang (2012). International Journal of Software Science and Computational Intelligence (pp. 64-80).
www.igi-global.com/article/image-retrieval-based-hsv-feature/72880?camid=4v1a

Simulation and Visualization of Concept Algebra in MATLAB
www.igi-global.com/article/simulation-and-visualization-of-concept-algebra-in-matlab/114095?camid=4v1a

The Formal Design Models of a Universal Array (UA) and its Implementation
Yingxu Wang, Jason Huang and Jingsheng Lei (2011). International Journal of Software Science and Computational Intelligence (pp. 69-89).
www.igi-global.com/article/formal-design-models-universal-array/60750?camid=4v1a