Image Retrieval Practice and Research

M

JungWon Yoon
University of South Florida, USA

INTRODUCTION

With emerging digital technologies, images become a significant communication venue, and image retrieval gains attention from both research and practitioner environments. It is a common view to distinguish image retrieval approaches into two categories, text-based and content-based (Chu, 2001; Enser, 2008). The text-based approach, which uses verbal descriptions for image indexing and retrieval, is a dominant method adopted in the practitioner environment and Library and Information Science field; whereas the content-based approach relies on low-level features (color, shape, texture, etc.), and has been explored primarily by the field of Computer Science (Enser, 2008). The content-based approach is practically adopted for special types of images, such as medical images, face and finger print recognition, trademark or logo searching, and so on, but for general images the content-based approach is still limited to the research environment.

As the image retrieval research has progressed, research which combines these two approaches are vigorously conducted, and recently, user-generated tags are considered as a promising source of image indexing. The purpose of this article is to overview state-of-the-art image indexing and retrieval approaches overarching research and practitioner environments.

BACKGROUND

Image Type and Features

Visual materials have been a form of written communication for a long time. Images can communicate information which cannot be delivered through phonetic alphabet; however, compared to text documents, there were difficulties in generating, distributing, and accessing images (Jörgensen, 2003). Recently, the development of digital technologies enhanced the usability and accessibility of images. The use of images is pervasive; it is used for face and fingerprint identification, medical purposes, trademark or logo searching, education, art and historical research, journalistic work, entertainment, online shopping, and so on. As much as the use of images is diverse, types of images also vary. Enser (2008) suggested image taxonomy as follows: direct picture which can be viewed with normal human visible spectrum, indirect picture which requires equipment for viewing the picture (e.g., images used in medical field, molecular biology, archeology, etc.), hybrid picture which integrates texts (e.g., posters or cartoons), and visual surrogate including drawing, diagram, map/chart/plan, and device. Smeulders, Worring, Santini, Gupta, and Jain (2000) categorized image domain into broad and narrow domains. In broad domain, images are polysemic and the interpretation of an image is not unique, whereas in narrow domain, images can be interpreted in a limited and predictable way. Among various types of images, this article focuses on image indexing and retrieval of general photographic images, which belong to the direct picture in Enser's taxonomy and the broad domain in Smeulders et al.'s category. Therefore, special types of images, such as medical images, face, trademarks, drawings, maps, symbols, cartoons, are beyond the scope of this article.

Images, especially those included in Smeulders et al.' broad domain, have unique features that make image indexing and retrieval difficult. First, an image has multi-layered meanings, and image interpretation is subjective and personal. An image may convey different meanings to different people, depending on their socio-cultural background, image need and usage purpose, disciplines, and other contextual background. Therefore, any set of index terms given by (an) indexer(s) or a creator may be different from viewers' interpretations on that image, and this has been the main problem of text-based image retrieval. Second, visual similarity does not always match with conceptualization. In other words, one conceptual object (e.g., glass) may have very different visual ap-

DOI: 10.4018/978-1-4666-5888-2.ch587

pearances, and two different conceptual objects (e.g., a starfish and the statue of liberty) may have similar visual appearance (Johansson, 2000). Third, there are concepts which do not have specific visual features, such as places (e.g. Florida), events (e.g., party), and abstract, symbolic, conceptual, and emotional concepts (e.g., poverty, celebrity, stylish), and so on. The gap between visual features of an image and semantic meanings that people recognize from the image are named "semantic gap" in the Content-Based Image Retrieval (CBIR) research field. The CBIR community recognizes that semantic meanings cannot be extracted solely from visual features, because human image retrieval process associates contextual background (Johansson, 2000; Jörgensen, 2007).

Image Attributes

Although it is a unique feature of images that they have multi-layered meanings, there is a lack of general agreement on what image attributes should be indexed and to what level attributes should be indexed (Matusiak, 2006). Therefore, researchers have attempted to identify and characterize image attributes, so that they can provide a rationale on what should be accomplished in the indexing of images (Jamies & Chang, 2000; Layne, 1994). Various frameworks have been developed for demonstrating image attributes, and Table 1 shows the comparison among selected frameworks. Jamies and Chang (2000), Hollink, Schreiber, Wielinga, and Worring (2004), Eakins, Briggs, and Burford (2004), Shatford, (1986), and Panofsky (1955) proposed frameworks based on their conceptual analysis on images. Jörgensen (2003) extracted image attributes by analyzing image-related tasks (description, sorting, and searching). Hare, Lewis, Enser, and Sandom (2007) analyzed index terms from a museum collection. Finally, Yoon and Chung (2011) analyzed narrative descriptions of image queries from a social Q&A site.

Although there are differences and commonalities among those frameworks, overall, image attributes can be classified into four broad categories. First, the descriptive metadata category includes attributes which are not part of the visual content of an image but describe the image, and this category includes attributes which are related to bibliographical information (title, creator, right, etc.) and physical information (format, size, medium, etc.). Second, attributes in the syntactic category depend on image viewers' visual perception. Since most attributes in the syntactic category are based on low-level features, world knowledge or contextual background does not influence image indexing. Third, the semantic category involves generic, specific, and abstract attributes. Most of users' image needs are related to attributes in the semantic category, but it is not easy to index semantic attributes which incorporate personal interpretation and world knowledge. Therefore, indexing semantic attributes has been a main focus of image indexing and retrieval research. Finally, the associated information category includes attributes which cannot be derived from an image itself, but are related to the image in some sense, such as related story, similar or comparative images, and related concepts. The attributes in this category are influenced by the context of the image and the background of users/indexers, and may provide information which cannot be obtained from the image itself.

APPROACHES TO IMAGE INDEXING AND RETRIEVAL

Metadata and Standards for Image Indexing

In the practitioner environments, text-based approach is a dominant way of providing access to images. The text-based approach indexes non-linguistic features through linguistic representation, so image indexing tends to be more dependent on knowledge organization systems, such as metadata and controlled vocabulary systems (Stvilia, Jörgensen, & Wu, 2012). Dublin Core, MAchine Readable Catalog (MARC), Anglo-American Cataloging Rules, 2nd ed. (AACR2), and Resource Description and Access (RDA) are often adopted for cataloging image collections, although they are not specialized for image collections. Visual Resources Association (VRA) Core (currently, version 4.0) and Categories for the Description of Works of Art (CDWA) are metadata standards specialized for image collections. VRA Core is designed for describing images and works and the relationship between images and works. Also, collection level cataloging (groups of works or groups of images) is available. VRA Core 4.0 includes 19 elements and their sub-elements, and provides an XML Schema for representing the metadata. CDWA is designed for describing and accessing information