Chapter I

Toward High-Level Visual Information Retrieval

Yu-Jin Zhang, Tsinghua University, Beijing, China

Abstract

Content-based visual information retrieval (CBVIR) as a new generation (with new concepts, techniques, mechanisms, etc.) of visual information retrieval has attracted many interests from the database community. The research starts by using a low-level feature from more than a dozen years ago. The current focus has shifted to capture high-level semantics of visual information. This chapter will convey the research from the feature level to the semantic level by treating the problem of semantic gap under the general framework of CBVIR. This high-level research is the so-called semantic-based visual information retrieval (SBVIR). This chapter first shows some statistics about the research publications on semantic-based retrieval in recent years; it then presents some existing approaches based on multi-level image retrieval and multi-level video retrieval. It also gives an overview of several current centers of attention by summarizing certain results on subjects such as image and video...
annotation, human-computer interaction, models and tools for semantic retrieval, and miscellaneous techniques in application. Before finishing, some future research directions, such as domain knowledge and learning, relevance feedback and association feedback, as well as research at even a high level such as cognitive level, are pointed out.

Introduction

It is said that “a picture is worth a thousand words.” Human beings obtain the majority of information from the real world by visual sense. This could include all entities that can be visualized, such as image and video (a chain/sequence of images) in a narrow sense, as well as animation, charts, drawings, graphs, multi-dimensional signals, text (in fact, many documents are used in image form, as indicated by Doermann, 1998), and so forth in a more general sense.

With the fast technique progress of computer science, electronics, medium capturing, and so forth, and the rapidly rising use of the Internet and the growing capability of data storage, the quantity of visual information expands dramatically and results in many huge visual information databases. In addition, many data are created and collected by amateurs, which is quite different than by professional people (Luo, Boutell, & Brown, 2006). In addition, visual media become a widespread information format in the World Wide Web (WWW) in which data are dispersed in various locations. All these make the search of required visual information more complex and time-consuming (Zhang, 2006). Along with the quickly increasing demands to create and store visual information comes the need for a richer set of search facilities. Providing tools for effective access, retrieval, and management of huge visual information data, especially images and videos, has attracted significant research efforts. Several generations of techniques and systems have been developed.

Traditionally, textual features such as captions, file names, and especially keywords have been used in searching required visual information. However, the use of keywords in the search is not only cumbersome but also inadequate to represent the riche content of visual information. Images are snapshots of the real world. Due to the complexity of scene content, there are many images for which no words can exactly express their implications. Image is beyond words, so it has to be seen and must be searched as image by content (i.e., object, purpose, scene, style, subject, etc.).

Content-based visual information retrieval has attracted many interests, from image engineering, computer vision, and database community. A large number of researches, especially on feature-based techniques, have been developed and have achieved plentiful and substantial results (Bimbo, 1999; Rui, Huang, & Chang, 1999; Smeulders et al., 2000; Zhang, 2003). However, in light of the complexity of the real world, low-level perceptive cues/indexes are not enough to provide suitable interpretation. To probe further, some higher-level researches and techniques for content understanding are mandatory. Among three broad categories of high-level techniques—synthetic, semantic, and semiotic—the semantic approach is quite natural from the understanding point of view. Nevertheless, from feature to semantic, there is a semantic gap. Solving this problem has been a focal point in
Semantic Annotation of Objects
www.igi-global.com/chapter/semantic-annotation-objects/35730?camid=4v1a