IDEA GROUP PUBLISHING
701 E. Chocolate Avenue, Hershey PA 17033-1117, USA
Tel: 717/533-8845; Fax 717/533-8661; URL-http://www.idea-group.com
ITB7266

Chapter II

Bridging the Semantic Gap in Image Retrieval

Rong Zhao and William I. Grosky
Wayne State University, USA

INTRODUCTION

The emergence of multimedia technology and the rapidly expanding image and video collections on the Internet have attracted significant research efforts in providing tools for effective retrieval and management of visual data. Image retrieval is based on the availability of a representation scheme of image content. Image content descriptors may be visual features such as color, texture, shape, and spatial relationships, or semantic primitives.

Conventional information retrieval was based solely on text, and those approaches to textual information retrieval have been transplanted into image retrieval in a variety of ways. However, "a picture is worth a thousand words." Image content is much more versatile compared with text, and the amount of visual data is already enormous and still expanding very rapidly. Hoping to cope with these special characteristics of visual data, content-based image retrieval methods have been introduced. It has been widely recognized that the family of image retrieval techniques should become an integration of both low-level visual features addressing the more detailed perceptual aspects and high-level semantic features underlying the more general conceptual aspects of visual data. Neither of these two types of features is sufficient to retrieve or manage visual data in an effective or efficient way (Smeulders, et al., 2000). Although efforts have been devoted to combining these two aspects of visual data, the gap between them is still a huge barrier in front of researchers. Intuitive and heuristic approaches do not provide us with satisfactory performance. Therefore, there is an urgent need of finding the latent correlation between low-level features and high-level concepts and merging them from a different perspective. How to find this new perspective and bridge the gap between visual features and semantic features has been a major challenge in this research field. Our chapter addresses these issues.

This chapter appears in the book, Distributed Multimedia Databases: Techniques and Applications by Timothy K. Shih.

Image Retrieval

Image retrieval is an extension to traditional information retrieval. Approaches to image retrieval are somehow derived from conventional information retrieval and are designed to manage the more versatile and enormous amount of visual data that exist.

Low-level visual features such as color, texture, shape and spatial relationships are directly related to perceptual aspects of image content. Since it is usually easy to extract and represent these features and fairly convenient to design similarity measures by using the statistical properties of these features, a variety of content-based image retrieval techniques have been proposed in the past few years. High-level concepts, however, are not extracted directly from visual contents, but they represent the relatively more important meanings of objects and scenes in the images that are perceived by human beings. These conceptual aspects are more closely related to users' preferences and subjectivity. Concepts may vary significantly in different circumstances. Subtle changes in the semantics may lead to dramatic conceptual differences. Needless to say, it is a very challenging task to extract and manage meaningful semantics and to make use of them to achieve more intelligent and user-friendly retrieval.

Challenges

High-level conceptual information is normally represented by using text descriptors. Traditional indexing for image retrieval is text-based (Jorgensen, 1998). In certain content-based retrieval techniques, text descriptors are also used to model perceptual aspects (Kelly & Cannon, 1995; Gimel'farb & Jain, 1996). Unfortunately, the inadequacy of text description is an obvious and very problematic issue. Meanwhile, image contents are much more complicated than textual data stored in traditional databases, yet there is an even greater demand for retrieval and management tools for visual data, since visual information is a medium more capable of conveying ideas and is more closely related to human perception of the real world. Image retrieval techniques should provide support for user queries in an effective and efficient way, just as conventional information retrieval does for textual retrieval.

However, the dynamic and versatile characteristics of image content require expensive computations and sophisticated methodologies in the areas of computer vision, image processing, data visualization, indexing, and similarity measurement. In order to manage image data effectively and efficiently, several schemes for data modeling and image representation have been proposed (Ahmad & Grosky, 1997; Aslandogan, et al., 1995; Chang & Liu, 1984; Chang & Wu, 1992; Chang, Shi, & Yan, 1987; Gudivada, 1995; Gudivada, 1997; Huang & Jean, 1994; Pentland, Picard, & Sclaroff, 1994; Smith & Chang, 1998; Tao & Grosky, 1999). Due to the lack of any unified framework for image representation and retrieval, certain methods may perform better than others under certain query situations. Therefore, these schemes and retrieval techniques have to be somehow integrated and adjusted on the fly to facilitate effective and efficient image data management.

Research Goals

The motivation of our research is to improve several aspects of content-based image retrieval by finding the latent correlation between low-level visual features and high-level semantics and integrating them into a unified vector space model. To be more specific, the significance of this approach is to design and implement an effective and efficient framework of image retrieval techniques, using a variety of visual features such as color, texture, shape