Advanced Techniques for Object-Based Image Retrieval

Y. J. Zhang
Tsinghua University, Beijing, China

INTRODUCTION

Along with the progress of imaging modality and the wide utility of digital images (including video) in various fields, many potential content producers have emerged, and many image databases have been built. Because images require large amounts of storage space and processing time, how to quickly and efficiently access and manage these large, both in the sense of information contents and data volume, databases has become an urgent problem. The research solution for this problem, using content-based image retrieval (CBIR) techniques, was initiated in the last decade (Kato, 1992). An international standard for multimedia content descriptions, MPEG-7, was formed in 2001 (MPEG). With the advantages of comprehensive descriptions of image contents and consistence to human visual perception, research in this direction is considered as one of the hottest research points in the new century (Castelli, 2002; Zhang, 2003; Deb, 2004).

Many practical retrieval systems have been developed; a survey of near 40 systems can be found in Veltkamp (2000). Most of them mainly use low-level image features, such as color, texture, and shape, etc., to represent image contents. However, there is a considerable difference between the users' interest in reality and the image contents described by only using the above low-level image features. In other words, there is a wide gap between the image content description based on low-level features and that of human beings' understanding. As a result, these low-level feature-based systems often lead to unsatisfying querying results in practical applications.

To cope with this challenging task, many approaches have been proposed to represent and describe the content of images at a higher level, which should be more related to human beings' understanding. Three broad categories could be classified: synthetic, semantic, and semiotic (Bimbo, 1999; Djeraba, 2002). From the understanding point of view, the semantic approach is natural. Human beings often describe image content in terms of objects, which can be defined at different abstraction levels. In this article, objects are considered not only as carrying semantic information in images, but also as suitable building blocks for further image understanding.

The rest of the article is organized as follows: in "Background," early object-based techniques will be briefly reviewed, and the current research on object-based techniques will be surveyed. In "Main Techniques," a general paradigm for object-based image retrieval will be described; and different object-based techniques, such as techniques for extracting meaningful regions, for identifying objects, for matching semantics, and for conducting feedback are discussed. In "Future Trends," some potential directions for further research are pointed out. In "Conclusion," several final remarks are presented.

BACKGROUND

Early Object-Based Techniques in Content-Based Image Retrieval

CBIR techniques are distinguished from traditional retrieval techniques by many aspects. Two of the most pertinent are that CBIR is a somehow subjective process, as for a given image, its means may have different interpretations for different users; and image retrieval is often a computationally expensive process, as the image database is often large in size and contains heterogeneous information. Due to these particular aspects, the results of CBIR could not be judged objectively—human perception should be considered. In addition, performing an exhaustive search for finding optimal solutions in CBIR is not feasible, and therefore, some suboptimal solutions will be chosen.

Because of the unique aspects of CBIR, object-based representation and description must be used even in so-called low-level feature-based image retrieval, though in these works, object recognition is not evidently performed and semantic information is not explicitly searched.

One typical example is in shape-based retrieval, as the shape features are generally extracted from individual objects (Latecki, 2002). In contrast, color features and textural features are often obtained by taking the whole image as a unit. From this point of view, shape-based retrieval is already at some higher level than color-based retrieval and texture-based retrieval (Zhang, 2003).

Structural query model is another instance in which partial matches are allowed and outputs related to the score of similarity can be provided. This type of retrieval is based on the relations between the individual objects and components in images (Zhou, 2001). In query by visual sketch, users sketch a scene by drawing a collection of objects. (It is assumed that these objects could fully define a scene.) For example, the objects are first identified and then used in a search (Chang, 1998).

Current Object-Based Techniques in Content-Based Image Retrieval

Currently, researchers seek explicit semantics and use the high-level descriptions that are common to humans, such as articles, people, places, and things. It is generally accepted that high-level features are crucial to improve the performance of CBIR up to so-called semantic-based querying. For this purpose, object-based content analysis, especially segmentation that segments the semantically meaningful objects from images, is an essential step (Zhang, 2001).

Complete image understanding should start at interpreting image objects and their relationships. Objects can be further identified in line with appropriate knowledge. For example, some object grammars based on rules for concept inference have been proposed (Petkovic, 2003). When domain knowledge is available, objects can be classified even without the explicit determination of object regions (Li, 2002b).

To extract high-level descriptions from images and to fill the gap between the low-level features and human beings' understanding of image contents, techniques to describe the whole image with a hierarchical structure to reach progressive image analysis are proposed (Castelli, 1998; Jaimes, 1999; Hong, 1999). The contents of images can be represented in different levels (Amir, 1998), such as the three-level content representation, including feature level content, object level content, and scene level content (Hong, 1999); and the five-level representation, including region level, perceptual region level, object part level, object level, and scene level (Jaimes, 1999). The problem here is how to implement these levels efficiently and effectively.

Another direction for extracting semantics information from an image is to map low-level visual features to high-level semantics. In other words, to fill the semantic gap, one makes the retrieval system work with low-level features, while the user puts in more high-level knowledge (Zhou, 2002). Two typical methods are to optimize query requests by using relevance feedback and semantic visual templates (Chang, 1998) and to interpret progressively the content of images by using interactive interfaces (Castelli, 1998). In both approaches, relevance feedback plays an important role, as humans are much better than computers at extracting semantic information from images (Rui, 1998; Ciocca, 1999).

MAIN TECHNIQUES FOR OBJECT-BASED IMAGE RETRIEVAL

A General Paradigm

In general, people distinguish three levels of abstraction when talking about image databases: raw data level, feature level, and semantic level. The raw data are original images in the form of a pixel matrix. The feature level shows some significant characteristics of the pixel patterns of the image. The semantic level describes the meanings of identified objects in images. Note that the semantic level should also describe the meaning of an image as a whole. Such a meaning could be obtained by the analysis of objects and the understanding of images.

According to the above discussions, a multilayer approach should be used for efficiently treating image data. Though the number of layers and the definitions and functions of these layers could have some variations in different approaches, some principle steps are common for object-based image retrieval. A general paradigm is shown in Figure 1.

First, objects should be determined. Two important tasks are as follows:

Figure 1. A general paradigm for object-based image retrieval