Content-Based Multimedia Retrieval

Chia-Hung Wei
University of Warwick, UK

Chang-Tsun Li
University of Warwick, UK

INTRODUCTION

In the past decade, there has been rapid growth in the use of digital media such as images, video, and audio. As the use of digital media increases, effective retrieval and management techniques become more important. Such techniques are required to facilitate the effective searching and browsing of large multimedia databases.

Before the emergence of content-based retrieval, media was annotated with text, allowing the media to be accessed by text-based searching (Feng et al., 2003). Through textual description, media can be managed, based on the classification of subject or semantics. This hierarchical structure allows users to easily navigate and browse, and can search using standard Boolean queries. However, with the emergence of massive multimedia databases, the traditional text-based search suffers from the following limitations (Djeraba, 2003; Shah et al., 2004):

- Manual annotations require too much time and are expensive to implement. As the number of media in a databases grows, the difficulty finding desired information increases. It becomes infeasible to manually annotate all attributes of the media content. Annotating a 60-minute video containing more than 100,000 images consumes a vast amount of time and expense.
- Manual annotations fail to deal with the discrepancy of subjective perception. The phrase "a picture is worth a thousand words" implies that the textual description is not sufficient for depicting subjective perception. Capturing all concepts, thoughts, and feelings for the content of any media is almost impossible.
- Some media contents are difficult to describe concretely in words. For example, a piece of melody without lyrics or an irregular organic shape cannot be expressed easily in textual form, but people expect to search media with similar contents based on examples they provide. In an attempt to overcome these difficulties, content-based retrieval employs content information to automatically index data with minimal human intervention.

APPLICATIONS

Content-based retrieval has been proposed by different communities for various applications. These include:

- **Medical Diagnosis:** The amount of digital medical images used in hospitals has increased tremendously. As images with the similar pathology-bearing regions can be found and interpreted, those images can be applied to aid diagnosis for image-based reasoning. For example, Wei & Li (2004) proposed a general framework for content-based medical image retrieval and constructed a retrieval system for locating digital mammograms with similar pathological parts.
- **Intellectual Property:** Trademark image registration has applied content-based retrieval techniques to compare a new candidate mark with existing marks to ensure that there is no repetition. Copyright protection also can benefit from content-based retrieval, as copyright owners are able to search and identify unauthorized copies of images on the Internet. For example, Wang & Chen (2002) developed a content-based system using hit statistics to retrieve trademarks.
- **Broadcasting Archives:** Every day, broadcasting companies produce a lot of audiovisual data. To deal with these large archives, which can

contain millions of hours of video and audio data, content-based retrieval techniques are used to annotate their contents and summarize the audiovisual data to drastically reduce the volume of raw footage. For example, Yang et al. (2003) developed a content-based video retrieval system to support personalized news retrieval.

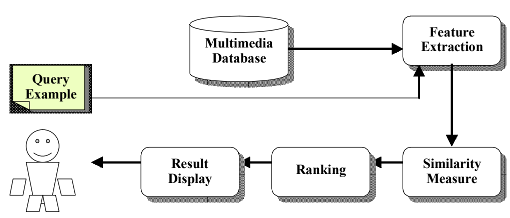
- **Information Searching on the Internet:** A large amount of media has been made available for retrieval on the Internet. Existing search engines mainly perform text-based retrieval. To access the various media on the Internet, content-based search engines can assist users in searching the information with the most similar contents based on queries. For example, Hong & Nah (2004) designed an XML scheme to enable content-based image retrieval on the Internet.

DESIGN OF CONTENT-BASED RETRIEVAL SYSTEMS

Before discussing design issues, a conceptual architecture for content-based retrieval is introduced and illustrated in Figure 1.

Content-based retrieval uses the contents of multimedia to represent and index the data (Wei & Li, 2004). In typical content-based retrieval systems, the contents of the media in the database are extracted and described by multi-dimensional feature vectors, also called descriptors. The feature vectors of the media constitute a feature dataset. To retrieve desired data, users submit query examples to the retrieval system. The system then represents these examples with feature vectors. The distances (i.e., similarities) between the feature vectors of the query example and those of the media in the feature dataset are then computed and ranked. Retrieval is conducted by applying an indexing scheme to provide an efficient way to search the media database. Finally, the system ranks the search results and then returns the top search results that are the most similar to the query examples.

Figure 1. A conceptual architecture for content-based retrieval



For the design of content-based retrieval systems, a designer needs to consider four aspects: feature extraction and representation, dimension reduction of feature, indexing, and query specifications, which will be introduced in the following sections.

FEATURE EXTRACTION AND REPRESENTATION

Representation of media needs to consider which features are most useful for representing the contents of media and which approaches can effectively code the attributes of the media. The features are typically extracted off-line so that efficient computation is not a significant issue, but large collections still need a longer time to compute the features. Features of media content can be classified into low-level and high-level features.

Low-Level Features

Low-level features such as object motion, color, shape, texture, loudness, power spectrum, bandwidth, and pitch are extracted directly from media in the database (Djeraba, 2002). Features at this level are objectively derived from the media rather than referring to any external semantics. Features extracted at this level can answer queries such as "finding images with more than 20% distribution in blue and green color," which might retrieve several images with blue sky and green grass (see Picture 1). Many effective approaches to low-level feature extraction have been developed for various purposes (Feng et al., 2003; Guan et al., 2001).

High-Level Features

High-level features are also called semantic features. Features such as timbre, rhythm, instruments, and events involve different degrees of semantics contained in the media. High-level features are supposed