

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, retrieval and management techniques become more important in order 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. Through textual description, media is managed and retrieved based on the classification of subject or semantics. This hierarchical structure, like yellow pages, allows users to easily navigate and browse, or search using standard Boolean queries. However, with the emergence of massive multimedia databases, the traditional text-based search suffers from the following limitations (Wei, Li, & Wilson, 2006):

Manual annotations require too much time and are expensive to implement. As the number of media in a database grows, the difficulty in 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, “an image says more than a thousand words,” implies that the textual description is sufficient for depicting subjective perception. To capture all concepts, thoughts, and feelings for the content of any media is almost impossible.

Some media contents are difficult to concretely describe in words. For example, a piece of melody without lyric or irregular organic shape cannot easily be expressed in textual form, but people expect to search media with similar contents based on examples they provided.

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 applications include:

- **Medical diagnosis:** The medical community is currently developing picture archiving and communication systems (PACS), which integrate imaging modalities and interfaces with hospital and departmental information systems in order to manage the storage and distribution of images to radiologists, physicians, specialists, clinics, and imaging centers. A crucial requirement in PACS is to provide an efficient search function to access desired images. 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 and Li (2006) proposed a content-based retrieval system for locating mammograms with similar pathological characteristics.
- **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 can also benefit from content-based retrieval as copyright owners are able to search and identify unauthorized copies of images on the Internet. For example, Jiang, Ngoa, and Tana (2006) developed a content-based system using adaptive selection of visual features for trademark image retrieval.
- **Broadcasting archives:** Every day, broadcasting companies produce a lot of audio-visual 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 audio-visual data to drastically reduce the volume of raw footage. For example, Lopez and Chen (2006) developed a content-based video retrieval system to support news and sports retrieval.

- **Multimedia searching on the Internet:** Although a large amount of multimedia has been made available on the Internet for retrieval, 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 media with the most similar contents based on queries. For example, Khan (2007) designed a framework for image annotation and used ontology to enable content-based image retrieval on the Internet.

FRAMEWORK OF CONTENT-BASED RETRIEVAL SYSTEMS

The retrieval framework, as shown in Figure 1m can be divided into off-line feature extraction and online retrieval. In the off-line feature extraction, the contents of the data in the database are pre-processed, extracted and described with a feature vector, also called a descriptor. A feature vector for each datum is stored alongside with its corresponding audio/video/image data in the database. In the online retrieval, the user can submit a query example to the retrieval system to search for desired data. The similarities between the feature vectors of the query example and those of the data in the database are computed and ranked. Retrieval is conducted by applying an indexing scheme to provide an efficient way of searching the database. Finally, the system ranks the similarity and returns the data that are most similar to the query example. If the user is not satisfied with the initial search results, human-centered computing is introduced into an interactive search process. The user can provide relevance information to the retrieval system in order to search further (following the arrows on the dashed lines in Figure 1). This interactive search process can be repeated until the user is satisfied with the search results or unwilling to offer any further feedback.

For the design of content-based retrieval system, 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 and meaningful 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 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 color, shape, texture, object motion, loudness, power spectrum, bandwidth, and pitch, are extracted directly from media in the database (Liu, Zhang, Lu, & Ma, 2007). Features at this level are objectively derived from the media themselves, 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, such as Picture 1. Many effective approaches to low-level feature extraction have been developed for various purposes (Feng et al., 2003; Russ, 2006).

High-Level Features

High-level features, which are also called semantic features, such as timbre, rhythm, instruments, and events, involve different degrees of semantics contained in the media. High-level features are supposed to deal with the semantic queries, such as the query “finding a picture of water” or “searching for Mona Lisa Smile.” The latter query contains higher-degree semantics than the former. As water in images displays the homogeneous texture represented in low-level features, such a query is easier to process. To retrieve the latter query, the retrieval system requires prior knowledge

5 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/content-based-multimedia-retrieval/17410

Related Content

New Internet Protocols for Multimedia Transmission

Michael Welzl (2008). *Multimedia Technologies: Concepts, Methodologies, Tools, and Applications* (pp. 1634-1642).

www.irma-international.org/chapter/new-internet-protocols-multimedia-transmission/27182

Using an Interactive Feedback Tool to Enhance Pronunciation in Language Learning

F. Zhang (2008). *Multimedia Technologies: Concepts, Methodologies, Tools, and Applications* (pp. 1042-1057).

www.irma-international.org/chapter/using-interactive-feedback-tool-enhance/27138

Online Education and Cultural Background

Elizabeth Koh (2009). *Encyclopedia of Multimedia Technology and Networking, Second Edition* (pp. 1080-1085).

www.irma-international.org/chapter/online-education-cultural-background/17520

A Simple Prediction Method for Progressive Image Transmission

Chin-Chen Chang, Guang-Xue Xiao and Tung-Shou Chen (2002). *Distributed Multimedia Databases: Techniques and Applications* (pp. 262-272).

www.irma-international.org/chapter/simple-prediction-method-progressive-image/8626

Radio Resource Management Strategies for HSDPA-Enhanced UMTS Networks

Dirk Staehle and Andreas Mäder (2009). *Handbook of Research on Wireless Multimedia: Quality of Service and Solutions* (pp. 31-54).

www.irma-international.org/chapter/radio-resource-management-strategies-hsdpa/22019