M

Multimedia Representation

Bo Yang
Bowie State University, USA

INTRODUCTION

Multimedia Information Processing: Promises and Challenges

In recent years, the rapid expansion of multimedia applications, partly due to the exponential growth of the Internet, has proliferated over the daily life of computer users (Yang & Hurson, 2006). The integration of wireless communication, pervasive computing, and ubiquitous data processing with multimedia database systems has enabled the connection and fusion of distributed multimedia data sources. In addition, the emerging applications, such as smart classroom, digital library, habitat/environment surveillance, traffic monitoring, and battlefield sensing, have provided increasing motivation for conducting research on multimedia content representation, data delivery and dissemination, data fusion and analysis, and content-based retrieval. Consequently, research on multimedia technologies is of increasing importance in computer society. In contrast with traditional text-based systems, multimedia applications usually incorporate much more powerful descriptions of human thought—video, audio, and images (Karpouzis, Raouzaiou, Tzouveli, Iaonnou, & Kollias, 2003; Liu, Bao, Yu, & Xu, 2005; Yang & Hurson, 2005). Moreover, the large collections of data in multimedia systems make it possible to resolve more complex data operations such as imprecise query or content-based retrieval. For instance, the image database systems may accept an example picture and return the most similar images of the example (Cox, Miller, & Minka, 2000; Hsu, Chua, & Pung, 2000; Huang, Chang, & Huang, 2003). However, the conveniences of multimedia applications come with challenges to the existing data management schemes:

- **Efficiency:** Multimedia applications generally require more resources; however, the storage space and processing power are limited in many practical systems, for example, mobile devices and wireless networks (Yang & Hurson, 2005). Due to the large data volume and complicated operations of multimedia applications, new methods are needed to facilitate efficient representation, retrieval, and processing of multimedia data while considering the technical constraints.
- **Semantic Gap:** There is a gap between user perception of multimedia entities and physical representation/access mechanism of multimedia data. Users often browse and desire to access multimedia data at the object level ("entities" such as human beings, animals, or buildings). However, the existing multimedia retrieval systems tend to access multimedia data based on their lower-level features ("characteristics" such as color patterns and textures), with little regard to combining these features into data objects. This representation gap often leads to higher processing cost and unexpected retrieval results. The representation of multimedia data according to human's perspective is one of the focuses in recent research activities; however, few existing systems provide automated identification or classification of objects from general multimedia collections.
- **Heterogeneity:** The collections of multimedia data are often diverse and poorly indexed. In a distributed environment, because of the autonomy and heterogeneity of data sources, multimedia data objects are often represented in heterogeneous formats. The difference in data formats further leads to the difficulty of incorporating multimedia data objects under a unique indexing framework.
- **Semantic Unawareness:** The present research on content-based multimedia retrieval is based on feature vectors—features are extracted from audio/video streams or image pixels, empirically or heuristically, and combined into vectors according to the application criteria. Because of the application-specific multimedia formats, the feature-based paradigm lacks scalability and accuracy.

Representation: The Foundation of Multimedia Data Management

Successful storage and access of multimedia data, especially in a distributed and heterogeneous database environment (Lim & Hurson, 2002), requires careful analysis of the following issues:

1. Efficient representation of multimedia data objects in databases,
2. Proper indexing architecture of the multimedia databases, and
3. Proper and efficient technique to browse and/or query data objects in multimedia database systems.

Multimedia representation focuses on efficient and accurate description of content information that facilitates multimedia information retrieval. Various approaches have been proposed to map multimedia data objects into computer-friendly formats; however, most of the proposed schemes cannot guarantee accuracy when performing content-based retrieval (Bourgeois, Mory, & Spies, 2003). Consequently, the key issue in multimedia representation is the trade-off between accuracy and efficiency (Li et al., 2003; Yu & Zhang, 2000).

As noted in the literature, multimedia information retrieval methods can be classified into three groups: query-by-keyword, query-by-example, and query-by-browsing (Kim & Kim, 2002). Each method is suitable for a specific application domain. Naturally, different query methods also require employment of different indexing models. Query-by-browsing is suitable for the novice user who has little knowledge about the multimedia data objects, so the aim of indexing is to compress multimedia data objects into small icons (compact representation) for fast browsing. For query-by-keyword, the keywords involve the complete semantic information about the multimedia data objects; thus it is suitable for application domains where the users have clear knowledge about the representation of the data. In the case of query-by-example, users are more interested in the effectiveness of the query processing in locating the most related images (nearest neighbors) to the given examples.

Among the aforementioned fundamental issues, multimedia representation provides the foundation for indexing, classification, and query processing. The suitable representation of multimedia entities has significant impact on the efficiency of multimedia indexing and retrieval (Karpouzis et al., 2003). For instance, object-level representation of multimedia data usually provides more convenient methods in content-based indexing than pixel-level representation (Bart, Huimin, & Huang, 2006). Similarly, queries are resolved within the representation domains of multimedia data, either at object-level or at pixel-level (Erol, Berkner, Joshi, & Hull, 2006). The nearest-neighbor searching schemes are usually based on the careful analysis of multimedia representation—the knowledge of data contents and organization in multimedia systems (Han, Toshihiko, & Kiyoharu, 2006).

BACKGROUND

Preliminaries

The major challenges of traditional multimedia indexing and query processing are caused by the feature-based representation of multimedia data contents. To analyze the impact of content representation, in this section, we will study the existing representation approaches and evaluate their effectiveness in describing multimedia contents.

Multimedia content representation is the process of mapping low-level perceptual features to high-level content information with high accuracy. It is the fundamental means for supporting effective multimedia information organization and retrieval. Several recent researches have proposed using either a generative statistical model, such as Markov Chain Monte Carlo model (Li & Wang, 2003), or a discriminative approach, such as Support Vector Machines model (Zhang et al., 2001) and Bayesian Point Machines model (Chang, Goh, Sychay, & Wu, 2003), to generate annotations for images. Some other studies employ clustering approach to classify multimedia objects into content-similar groups. The content-representation approaches are usually based on two assumptions:

- The set of content categories is known a priori. That is to say, the number of possible categories is fixed, and the content of each category is determined by the multimedia database administrator.