Multimedia Content Representation Technologies

Ali R. Hurson
The Pennsylvania State University, USA

Bo Yang
The Pennsylvania State University, USA

INTRODUCTION

Multimedia: 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 Internet users. 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 (Auffret, Foote, Li & Shahraray, 1999). 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, image database systems may accept an example picture and return the most similar images of the example (Cox, Miller & Minka, 2000, Huang, Chang & Huang, 2003). However, the conveniences of multimedia applications come at the expense of new challenges to the existing data management schemes:

- 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 (Lim & Hurson, 2002). Due to the large size of multimedia databases and complicated operations of multimedia applications, new methods are needed to facilitate efficient accessing and processing of multimedia data while considering the technological constraints (Bourgeois, Mory & Spies, 2003).
- There is a gap between user perception and physical representation 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 represent multimedia data based on their lower-level features ("characteristics" such as color patterns and textures), with less emphases on combining these features into objects (Hsu, Chua & Pung, 2000). This representation gap often leads to unexpected retrieval results. The representation of multimedia data according to a human's perspective is one of the focuses in recent research activities; however, no existing systems provide automated identification or classification of objects from general multimedia collections (Kim & Kim, 2002).
- The collections of multimedia data are often diverse and poorly indexed (Huang et al., 2002). In a distributed environment, due to the autonomy and heterogeneity of data sources, multimedia objects are often represented in heterogeneous formats (Kwon, Choi, Bisdikian & Naghshineh, 2003). The difference in data formats further leads to the difficulty of incorporating multimedia objects within a unique indexing framework (Auffret et al., 1999).
- Last but not least, present research on content-based multimedia retrieval is based on features. These features are extracted from the audio/video streams or image pixels, with the empirical or heuristic selection, and then combined into vectors according to the application criteria (Hershey & Movellan, 1999). Due to

the application-specific multimedia formats, this paradigm of multimedia data management lacks scalability, accuracy, efficiency and robustness (Westermann & Klas, 2003).

Representation: The Foundation of Multimedia Data Management

Successful storage and access of multimedia data, especially in a distributed heterogeneous database environment, require careful analysis of the following issues:

- Efficient representation of multimedia entities in databases
- Proper indexing architecture for the multimedia databases
- Proper and efficient technique to browse and/or query objects in multimedia database systems.

Among these three 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 (Huang et al., 2003). For instance, object-level representation usually provides more convenient content-based indexing on multimedia data than pixel-level representation (Kim & Kim, 2002). Similarly, queries are resolved within the representation domains of multimedia data, either at the object level or pixel level (Hsu et al., 2000). The nearest-neighbor searching schemes are usually based on careful analysis of multimedia representation – the knowledge of data contents and organization in multimedia systems (Yu & Zhang, 2000; Li et al., 2003).

The remaining part of this article is organized into three sections: First, we offer the background and related work. Then, we introduce the concepts of semantic-based multimedia representation approach and compare it with the existing non-semantic-based approaches. Finally, we discuss the future trends in multimedia representation and draw this article into a conclusion.

BACKGROUND

Preliminaries of Multimedia Representation

The main goal of multimedia representation is to obtain a concise content description during the analysis of multimedia objects. Representation approaches as advanced in the literature are classified into four groups: clustering-based, representative-region-based, decision-tree-based and annotation-based.

Clustering-Based Approach

The clustering-based approach recursively merges content-similar multimedia objects into clusters with human intervention or automated classification algorithms while obtaining the representation of these multimedia objects. There are two types of clustering schemes: *supervised* and *unsupervised* (Kim & Kim,

Figure 1. The decomposition of clusters