Multimedia Database Management Systems

ARIF GHAFOOR

School of Electrical & Computer Engineering, Purdue University, West Lafayette, Indiana

1. INTRODUCTION

Due to the heterogeneous nature of multimedia data, systems intended to store, transport, display and in general manage such data must have considerably more capabilities than conventional information-management systems. In this paper, we elaborate on issues pertaining to multimedia database management and discuss technical challenges involved in developing a general-purpose multimedia DBMS. We can view these issues as specification requirements for this DBMS and focus our discussion on a reference architecture for such a system. A list of prominent issues in multimedia database management follows.

Development of formal semantic modeling techniques for multimedia information, especially for video and image data. These models should be rich in capabilities for abstracting multimedia information and capturing semantics. They should be able to provide canonical representations of complex images, scenes, and events in terms of objects and their spatio-temporal behavior. These models need to be compared and evaluated.

Design of powerful indexing, searching, and organization methods for multimedia data. Search in multimedia databases can be computationally intensive, especially if content-based retrieval is needed for image and video data stored in compressed or uncompressed form.

Development of models for specifying the media synchronization/integration requirements. Integration of these models with the monomedia database schema will be required. Subsequently, in order to determine the synchronization requirements at retrieval time, transformation of these models into a metaschema is needed. This entails designing object-retrieval algorithms for the operating systems. Similarly, integration of these models with higher-level information abstractions such as hypermedia or object-oriented models may be required.

Designing formal multimedia query languages. These languages should have strong capabilities to express arbitrarily complex semantic and spatiotemporal schemas associated with composite multimedia information. They must support manipulation of content-based functions for multimedia objects.

Development of efficient data-placement schemas for physical storage management. These schemes are needed to manage real-time multimedia data, for both single and parallel disk systems.

Design and development of suitable architecture and operating system support. Heterogeneity of multimedia information dictates that the architecture of a general-purpose multimedia database management system must support a rich set of data-management and computational functionalities. The operating system must also support real-time requirements of multimedia data.

Management of distributed multimedia databases. In a networked environment extensive coordination and management capabilities are needed among the distributed sites to provide location-transparent access and support real-time delivery of data to distributed users.

2. A REFERENCE ARCHITECTURE FOR MULTIMEDIA DBMS

Many types of real-world knowledge can be represented by describing the interplay among objects (persons, buildings, events, etc.) in the course of time and their relationships in space. An application may need to store and access information about this knowledge that can be expressed by complex spatiotemporal logic. A video database is a typical replica of this real-world environment [Day et al. 1995]. The spatiotemporal specification process is reversed while composing multimedia information/documents [Little and Ghafoor 1993; Iino et al. 1994]. In this case, a user synthetically creates interplay among various monomedia objects (akin to physical objects), both in space and time. In multimedia databases these objects may represent individual data entities that serve as components of some multimedia document [Little and Ghafoor 1993; Stotts and Furuta 1989]. Furthermore, these entities/documents can be grouped together for efficient management and access.

In either case, it is essential that the user be able to identify and address different objects and be able to express their relations in time and space. These relations should be representable in a suitable structure that is powerful enough to specify higher-level contents and semantic abstractions. It is, therefore, desirable that a general framework for spatiotemporal and semantic modeling should be available. Such a framework can provide a reasonable approach to address the issues discussed in the previous section and can be used in designing a general-purpose multimedia database-management system. Various conceptual models for multimedia information have been proposed in the literature [Little and Ghafoor 1993; Stotts and Furuta 1989]. These models are either aimed at synchronization aspects of the multimedia data or are concerned with the browsing aspects of information. Irrespective of the media type, it is imperative that a model should be able to generate a clear specification of the meta-schema and it must be integrable with the underlying data models of various monomedia. Such a model must also facilitate the efficient development of high-level user interfaces.

From this discussion, we can perceive a reference architecture for a general-purpose multimedia DBMS as shown in Figure 1(a). The architecture consists of three layers: a monomedia database-management layer, a multimedia composition/management layer, and a user-interface layer. We now describe functionalities of each layer and discuss their role in managing the issues mentioned in the previous section.

The monomedia database-management layer provides the functionalities essential for managing individual media. The key aspects of each DBMS at this level include maintaining efficient indexing mechanisms and allowing users to develop semantics-based modeling and grouping of complex information associated with each medium. In this respect, this layer deals with the first two issues discussed in the previous section. The primary objective is to process content-based queries as discussed in the previous section. The primary objective is to process content-based queries and facilitate retrieval of appropriate pieces of monomedia data, such as video clips, parts of an image, or audio segments. For example, for video data, abstraction hierarchies in space and time, as shown in Figure 1(b), are needed in order to express events and allow content-based indexing and retrieval. Spatiotemporal logic can provide a formal framework for representing such events and building these hierarchical abstractions [Day et al. 1995]. Subsequently, this process can lead to schema definition for each monomedia database. Such schema may also contain some unconventional information, for example, the quality of service (QoS) parameters for presentation, such as speed, volume, resolution, delay bounds, and the like. The objective is to

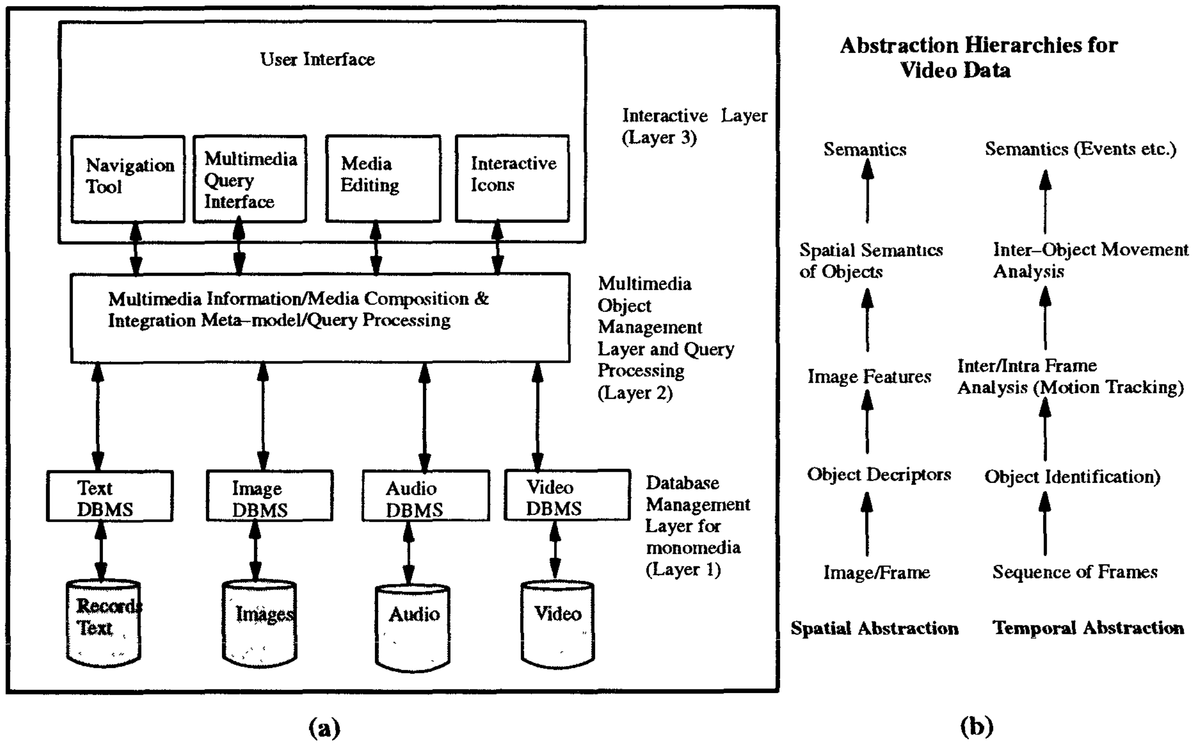

Figure 1. Reference architecture; (b) abstraction for video data.

allow retrieval of monomedia data with the desired quality, if possible. Accordingly, a suitable database language is needed at this level to manipulate hierarchical abstractions and to query about events of interest. It is important to mention that considerable computational demand can be placed at this layer, especially for image and video data, if a fully automated system is desired that is capable of generating content-based indices based on object recognition [Guidavada and Raghavan 1995]. Such computation can span a broad spectrum of effective and robust computer-vision and image-processing algorithms that are intermingled with data and entities at various levels of abstractions of Figure 1(b).

Another important function incorporated within this layer is the physical management of individual databases. Such management may require, as mentioned previously, efficient placement of data on a single/parallel disk system. For example, due to the time-variant characteristics of compressed video, it is difficult to predict the disk seek latencies. A poor data-placement strategy can result in significant degradation in quality due to excessive droppage of video frames during multiple concurrent sessions.

The primary objective of the middle layer is to deal with the third issue, the integration of monomedia to compose multimedia documents and to crosslink information stored in monomedia databases. There is a growing demand for management of multimedia documents and libraries, and the need for efficient integration models is one of the key research issues in developing a general-purpose multimedia DBMS. Integration of media can span multiple dimensions including space, time, and semantics. Therefore, this layer needs to maintain a multidimensional monomedia integration model, in form of a meta-schema, along with the QoS parameters associated with each medium. Basically, the layer processes users' queries for composite multimedia information and generates appropriate subqueries for the monomedia databases. Retrieval of the monomedia data is controlled by the integration model maintained by the meta-schema [Little and Ghafoor 1993]. It is important to note that for the purpose of consistency, this meta-schema needs to be interfaced with the schemas associated with the monomedia databases present at the lower layer, because both schemas share information about monomedia, including their content-based semantics and QoS parameters.

Another important function of this layer is to provide coordination among monomedia databases if they are distributed. As mentioned before, one of the major objectives in this case is to provide location-transparent access to different database sites and maintain synchronization among media streams originating from heterogeneous hosts. From the database point of view, information about the location of various objects of composite multimedia and their schema must be maintained by this layer.

The interactive layer consists of various user-interface facilities that support multimedia presentation functionalities such as display of images, playout of video clips or some audio segment, and the like. These interactive facilities may require some formal query language that can be identical to that used at the bottom layer, with some enhanced capabilities to manipulate composite multimedia information. Alternatively, some graphical query interface can be used for that purpose. Additional capabilities may include browsing, media editing, and so on.

3. DEVELOPMENT CONSIDERATIONS FOR THE REFERENCE ARCHITECTURE

The development of a general-purpose multimedia database system, including design of multilevel meta-schema, management of a large number of indices and physical databases, as well as interfacing with the operating system for real-time retrieval, can be a technologically daunting task. In this section, we focus on two development issues that are unique to multimedia databases and are crucial for realizing the reference architecture of Figure 1(a): spatiotemporal modeling of monomedia/composite multimedia and the suitability of two well-known data-modeling paradigms, relational versus object-oriented, for developing a general-purpose multimedia DBMS.

A number of attempts have been made to develop synchronization models for multimedia information. HyTime, one such model, has been recommended as an ISO standard. However, currently this model suffers from the drawback that the extraction of various spatiotemporal and content semantics from this model can be quite cumbersome. On the other hand, the Petri-net-based model proposed in Little and Ghafoor [1993] not only allows extraction of the desired semantics and generation of a database schema in a rather straightforward manner, but also has the additional advantage of pictorially illustrating synchronization aspects of the information. In this regard this model is unique and is thus also well suited for the third layer of the reference architecture, where visual orchestration of multimedia information can be highly desirable. The model uses a set of generalized temporal operators. An expanded version of the model also allows specification of spatial semantics of information [Iino et al. 1994]. The notion of generalized spatiotemporal relations has recently been extended to specify complex events in video data, a process that is almost a reversal of the multimedia composition process [Day et al. 1995]. Numerous extensions to this model have been proposed in the literature, including provision for user interaction, modeling of distributed object synchronization, development of synchronization protocols for communication, and the like. In summary, the Petri-net-based model in Little and Ghafoor [1993] can be a good candidate for spatiotemporal modeling and generation of meta-schemas for the reference architecture.

We now discuss the second development issue. As the application domain for a given multimedia application can be highly complex and ill-structured, both relational calculus and semantic-based object-oriented approaches need to be scrutinized. We can evaluate these approaches on two important criteria: their ability to express semantics associated with hierarchical abstractions, which need to be managed by almost all the monomedia DBMSs of the bottom layer of Figure 1(a), and their expressive power for the specification of meta-schema for media integration, which is managed by the middle layer. Hierarchical abstractions can result in considerable semantic heterogeneity, which has been a difficult probem for relational database models. On the other hand, the semantic-based object-oriented models can provide a richer set of abstractions that are particularly useful for the users to extract/define view of information at various levels of these hierarchies [Day et al. 1995]. From the media integration and composition point of view, the relational technology again provides no elegant mechanisms for maintaining complex logical structures associated with composite multimedia objects. Although some relational DBMSs support access to multimedia objects by using pointers to BLOBs (binary large objects), there is no provision for interactively accessing various portion of these objects, because a BLOB is treated as a single entity in its entirety. It has been demonstrated that the object-oriented technology can provide a powerful paradigm to meet the requirements of multimedia composition and developing semantic models [Day et al. 1995; Iino et al. 1994]. Its data and computational encapsulation features provide effective mechanisms for media synchronization.

REFERENCES

Day, Y. F., Dagtas, S., Iino, M., Khokhar, A., and Ghafoor, A. 1995. Object-oriented conceptual modeling of video data. In *Proceedings of the Eleventh IEEE International Conference on Data Engineering* (Taipei, Taiwan) March, 401–408.

Guidavada, V. N. and Raghavan, V. V., Eds. 1995. Content-based image retrieval systems. Special issue. *IEEE Comput. 28*, 9 (Sept).

IINO, M., DAY, Y. F., AND GHAFOOR, A. 1994. An object-oriented model for spatio-temporal synchronization of multimedia information. In *Proceedings of the First IEEE International Conference on Multimedia Computing and Systems* (Boston, MA) May, 110–119.

LITTLE, T. D. C. AND GHAFOOR, A. 1993. Interval-based conceptual models for time-dependent multimedia data. *IEEE Trans. Knowl. Data Eng. 5*, 4 (Aug), 551–563.

STOTTS, P. D. AND FURUTA, R. 1989. Petri-net-based hypertext: Document structure with browsing semantics. *ACM Trans. Office Autom. Syst. 7*, 1 (Jan.), 3–29.