An MPEG-7 Based Video Database Management System

Adnan Yazici\textsuperscript{1}, Öznur Yavuz\textsuperscript{1}, and Roy George\textsuperscript{2}

\textsuperscript{1} Dept. of Computer Engineering
METU, 06531, Ankara, Turkey
yazici@ceng.metu.edu.tr
\textsuperscript{2} Dept. of Computer Science
Clark Atlanta University
Atlanta, GA 30314, U.S.A.

8.1 Introduction

The availability of higher storage capacities and faster processing speeds have made it possible to store and operate on multimedia data resources such as images, graphics, 3D models, audio, speech, and video. The quantity of multimedia data available in digital archives, on the World Wide Web (WWW), and in personal and professional databases is growing rapidly. There is a need for systems that can index, store, and retrieve these media items for applications such as digital libraries, geographical information systems, video on demand services, teleconferencing, etc. However, developing these mechanisms is a difficult task due to the high volume and complexity of multimedia material, and a lack of adequate indexing standards. Since traditional database systems are inadequate in handling multimedia data, the development of multimedia databases has drawn both research and industry efforts. Although searching and querying on image databases have been studied intensively it is only in recent years, video database management systems have been researched.

In this research we develop a web-based video database management system that supports the semantic description of video content, and facilitates content-based querying on video data. Video entities are classified as Objects (cars, balls, etc.), Agent Objects (persons, teams, etc.), Activities (eating, shouting, etc.), and Events (eating an apple by John.). These artifacts and the spatial, temporal, and spatio-temporal properties of these entities are expressed using the description tools of MPEG-7 standard. The Data Description Language of MPEG-7 is based on the XML schema language [11, 15, 16] and is used to define multimedia objects in the video database system. Additionally, new description schemas may be defined by extending these description tools to address specific requirements of the system, espe-
cially to deal with the temporal, spatial, and spatio-temporal queries (object trajectories) that have to be supported by the system. Defining the spatial relationships between objects and agents objects crisply is an oversimplification and is not sufficient for deployable systems. We define a flexible spatial and spatio-temporal query model that retrieves the results based on fuzzy similarity measures. This generalized query model also retrieves exact results as a subset of the larger set of similarity based results. XML documents containing MPEG-7 descriptions are stored on text large object type columns as attributes of videos. All queries are performed using XML Path Language (XPath), which is a query notation for addressing and filtering the elements and text of XML documents. With our system it is possible to query multiple related videos defined under a single video group name.

There are a number of innovate contributions presented in this chapter. A video database system based on the spatio-temporal video model [21] was developed during this research. This model differs from the other extant ones in several ways. First, all entities and their properties are described using MPEG-7 descriptions. The XML documents containing MPEG-7 descriptions are stored in tree like DOM structure [7]. Queries are performed using XPath expressions which are applied to the nodes of DOM tree [7]. Second, this model supports a richer and more flexible query model in comparison to extant video models. The enhanced query model includes temporal relationships queries (starts, finishes, meets, equals, overlaps, during, and before) between events and directional trajectory queries (e.g., trajectories of an object starting from the left of a region and ending at the bottom of this region) of objects and agents. Querying on video groups and the definition of still and moving regions as inputs to the query model is also facilitated by the system.

The organization of the rest of the chapter is as follows: Section 8.2 explains the concepts and data structures of AVIS model [1], and the spatio-temporal video extensions to this model. Section 8.3 briefly overviews the MPEG-7 description tools used in the system and Section 8.4 gives the implementation details of the system. Query processing mechanisms and some of the XPath expressions used by the system are included in this section. Finally, Section 8.5 investigates possible extensions and concludes the study.

8.2 Related Work

There have been several previous research efforts on developing video models that permit advanced querying methodologies. These efforts include annotation-based models, physical level video segmentation approaches and object-based modeling approaches. Annotation-based video models allow free text or attribute/keyword annotations in order to describe semantic and spatial attributes of video data. The physical level video segmentation approach does not address semantic concepts, such as, objects, events, roles, players, etc., rather the video data is described as a stream of small segments with applica-