Preface

A few years ago, the main focus in multimedia research and development was on multimedia communications and multimedia authoring and presentation. Since more and more digital multimedia data in the forms of image, video, and audio are being captured and stored, the focus has shifted in last few years to efficient and effective storage and retrieval of multimedia information. A similar situation happened about thirty years ago when more and more alphanumeric data were stored in computer-readable form. This led to the development of database management systems (DBMSs), which are now one of the most popular computer applications and are used in almost every organization. Due to the differences between the characteristics and requirements of alphanumeric data and multimedia data, a traditional DBMS is not capable of effectively handling multimedia data. Thus new multimedia indexing and retrieval techniques are required.

This book provides comprehensive coverage on issues and current technologies in multimedia database management. It starts with a discussion on the main characteristics and requirements of multimedia data. It then describes general design issues of multimedia database management systems to meet these characteristics and requirements. We discuss techniques for the indexing and retrieval of text documents, audio, images, and video. The commonality among these techniques is that they all try to extract main features from raw media data, and then try to retrieve items relevant to the user's query based on the similarity or the distance between the extracted feature vectors of the stored items and the query. As multiple media types usually appear together in multimedia objects or documents and different techniques capture different features or contents of them, we discuss how to integrate various indexing and retrieval techniques for effective retrieval of multimedia documents. Since the features extracted from raw media data are usually represented by multidimensional vectors, it would be very time-consuming to calculate the similarity between the query feature vector and the feature vector of each of the stored items. We discuss various techniques and data structures so that the search and retrieval can be carried out efficiently.

Multimedia databases are usually accessed remotely oyer a network. Multimedia objects identified as relevant to the query must be retrieved from the server and transmitted to the client for presentation. A set of requirements involving factors including delay and delay jitter, called quality of service, must be met to achieve multimedia presentation in a smooth and timely fashion. We describe computer architecture, multimedia storage, operating system, and networking support to meet these requirements.

In traditional DBMSs, the main performance concern is efficiency (how long it takes to answer a query). In multimedia database management systems (MMDBMSs), efficiency is important too. In addition, retrieval effectiveness (ability to retrieve relevant items and ability to reject irrelevant items) becomes important. This is because MMDBMSs retrieve items based on similarity measured using a similarity metric instead of exact match. Since it is difficult to extract all features of multimedia items and design a similarity metric that exactly conforms to human judgment, it is likely that some items determined as relevant by the system are actually judged irrelevant by the user and some relevant items are not retrieved. Thus we also discuss performance measurement issues in MMDBMSs. Finally, we briefly describe current products, application development, and other issues such as security and standardization.

This book is intended for the following groups of readers:

• University students at senior levels and postgraduate students. Many universities around the world have started or will start to offer subjects related to multimedia technology and MMDBMSs. This book serves as a text for such subjects.

• System designers and developers who want to learn various issues and techniques in developing multimedia database management systems.

• Researchers who want to learn the current developments and new research directions in MMDBMSs.

• Other professionals who want to know technical issues and the current status of MMDBMSs.

I would like to thank Associate Professor Bob Bignall for his constant encouragement and support, and for proofreading all chapters for me. I thank Templar Hankinson for preparing the three graphs in Chapter 5, and Dengsheng Zhang and Shyh Wei Teng for correcting some errors in the original manuscript. I would also like to thank my colleagues and friends at Gippsland School of Computing and Information Technology of Monash University for their support during the last three years.

Dr. Ilias Petrounias of UMIST in Machester, United Kingdom, reviewed the book and provided many helpful comments and suggestions, for which I am most grateful.

I dedicate this book to my wife Fengxia, sons James and Colin, without whose support and understanding the writing of this book would not have been possible.

Chapter 1— Introduction

The initial focus in multimedia research and development was on multimedia communications and multimedia authoring and presentation [1–4]. In the past few years, more and more digital multimedia data in the forms of image, video, and audio has been captured and stored. There is now a strong research interest in efficient and effective storage and retrieval of multimedia information. A similar situation happened about thirty years ago when more and more alphanumeric data was stored in computer readable form. This led to the development of database management systems (DBMSs), which are now one of the most popular computer applications and are used in almost every organization. However, DBMSs cannot handle multimedia data effectively because of the differences between the characteristics of alphanumeric data and multimedia data. Therefore, new multimedia indexing and retrieval techniques are required.

The main purpose of this book is to describe issues and current technologies in multimedia indexing and retrieval. The area of multimedia indexing and retrieval is still in its early development stage. But it is expected that multimedia database management systems (MMDBMSs) will be as popular as current DBMSs in the near future [5–7].

This chapter first defines some important terms and concepts used throughout the book, and explains why traditional DBMS cannot handle multimedia information and why new technologies are required to support efficient and effective multimedia information retrieval. It then gives an overview of multimedia indexing and retrieval systems (MIRSs). Some expected capabilities and common applications of MIRSs are presented. The chapter concludes with an outline of the organization of the following chapters.

1.1— Some Important Definitions

To avoid any confusion or misunderstanding, this section provides definitions for some important terms and concepts used in this book.

1.1.1— Media Types and Multimedia

Media refer to the types of information or types of information representation, such as alphanumeric data, images, audio, and video. There are many ways to classify media. Common classifications are based on physical formats and media relationships with time. In this book, we classify media based on whether or not there are time dimensions to them. This convention leads to two classes of media: *static* and *dynamic* (or *time continuous).*

Static media do not have a time dimension, and their contents and meanings do not depend on the presentation time. Static media include alphanumeric data, graphics, and still images.

Dynamic media have time dimensions, and their meanings and correctness depend on the rate at which they are presented. Dynamic media include animation, audio, and video. These media have their intrinsic unit intervals or rates. For example, to convey a perceptually smooth movement, video must be played back at 25 frames per second (or 30 frames, depending on the video system used). Similarly, when we play back a recorded voice message or music, only one playback rate is natural or sensible. Playback at a slower or faster rate distorts the meaning or the quality of the sound. Because these media must be played back continuously at a fixed rate, they are often called *continuous media.* They are also called *isochronous media* because of the fixed relationship between each media unit and time.

Multimedia refers to a collection of media types used together. It is implied that at least one media type is not alphanumeric data (i.e., at least one media type is image, audio, or video). In this book, "multimedia" is used as an adjective—so we will specifically say multimedia information, multimedia data, multimedia system, multimedia communications, multimedia applications, and so forth. Multimedia data refers to the computer-readable representation of multiple media types. Multimedia information refers to the information conveyed by multiple media types. Sometimes, multimedia information and multimedia data are used interchangeably.

We sometimes use multimedia or media item and object to refer to any autonomous entity in an MIRS that can be queried, retrieved, and presented. The term "object" may not be properly defined in the technical object-oriented (OO) sense. The context should make it clear whether it is used in a general sense or refers to a properly defined object in an OO approach.

1.1.2— Databases and DBMSs

In the literature, databases and DBMSs are sometimes used interchangeably. In this book, database refers to a collection or repository of data or media items. We use DBMS to refer to the entire system that manages the database.

1.1.3— Text Document Information Retrieval

Automated information retrieval (IR) systems were developed to help manage the huge amount of scientific literature that has been created since the 1940s [8, 9]. The main function of an IR system is to store and manage a large number of text documents in a way so that documents relevant to user queries can be retrieved quickly. Note that the acronym IR specifically refers to *text document retrieval*, although the full term *information retrieval* can refer to retrieval of any type of information. We will discuss IR techniques in Chapter 4.

1.1.4— Multimedia Indexing and Retrieval

DBMSs retrieve items based on structured data using exact matching. IR is also called text-based retrieval. Content-based retrieval refers to retrieval based on actual media features such as color and shape, instead of text annotation of the media item. Content-based retrieval is normally based on similarity instead of an exact match between a query and a set of database items.

MIRS refers to a basic system providing multimedia information retrieval using a combination of DBMS, IR, and content-based retrieval techniques. In an MIRS, some issues such as versioning and security control may not be fully implemented. A fully fledged MIRS is called a multimedia DBMS (MMDBMS).

1.1.5— Feature Extraction, Content Representation, and Indexing

In MIRSs, one of the most important issues is feature extraction or content representation (what are the main features or contents in a multimedia item). Feature extraction may be an automatic or semiautomatic process. In some of the content-based retrieval literature, feature extraction is also called indexing. We follow this convention in this book. When the term "index" is used as a noun, it refers to a data structure or to the organization of extracted features for efficient search and retrieval.

1.2— Need for MIRS

The need for MIRS can be explained by the following three facts. First, more and more multimedia data is being captured and stored. In order to use the information contained in this data, an efficient and effective indexing and retrieval system is required. Second, multimedia data has special characteristics and requirements that are significantly different from alphanumeric data. The traditional DBMS, therefore, is not suitable for handling multimedia data. Third, although IR techniques can help in multimedia retrieval, they alone are not adequate to handle multimedia data effectively.

1.2.1— Proliferation of Multimedia Data and Its Characteristics

We are currently faced with an explosion of multimedia information. For example, a large amount of images and video are being created and stored on the Internet. Many paintings and pictures in printed form are being converted to digital form for easy processing, distribution, and preservation. Pictures from TV news and newspapers are also being converted into digital form for easy maintenance and preservation. A large number of medical images are being captured every day and satellites are producing many more. This trend is going to continue with the advancement of storage and digital technologies. Creating a mere repository for this ever-increasing amount of multimedia information is of little use, however. It will be impossible to fully use this multimedia information unless it is organized for rapid retrieval on demand.

Not only is an increasing amount of data being stored, but also the types of data and its characteristics are different from alphanumeric data. We discuss different media types and their characteristics in Chapter 2. We list the main characteristics of multimedia data below:

• Multimedia data, especially audio and video, is very data intensive. For example, a 10-minute video sequence of reasonable quality requires about 1.5 GB of storage without compression.

• Audio and video have a temporal dimension and they must be played out at a fixed rate to achieve the desired effect.

• Digital audio, image, and video are represented in a series individual sample values and lack obvious semantic structure for computers to automatically recognize the contents.

• Many multimedia applications require simultaneous presentation of multiple media types in a spatially and temporally coordinated way.

• The meaning of multimedia data is sometimes fuzzy and subjective. For example, different people may interpret the same picture in entirely different ways.

• Multimedia data is information rich. Many parameters are required to represent its content adequately.

1.2.2— DBMSs and Their Role in Handling Multimedia Data

DBMSs are now well developed and used widely for structured data. The dominant DBMSs are relational database management systems (RDBMSs). In RDBMSs, information is organized in tables or relations. The rows of the table correspond to information item or records, while the columns correspond to attributes. The structured query language (SQL) is used to create such tables and to insert and retrieve information from them.

We use a simple example to show how to use SQL to create a table and insert and retrieve information from it. Suppose we want to create a table containing student records consisting of the student number, name, and address. The following statement is used:

```
create table STUDENT (
      stu# integer,
      name char(20),
      address char(100) );
```

The above statement creates an empty table as shown in Table 1.1. When we want to insert student records into the table, we use the SQL *insert* command as follows:

```
insert into STUDENT values (10, "Lew, Tom", "2 Main St.,
Churchill,Australia");
```

The above statement will insert a row in the STUDENT table as shown in Table 1.2. More student records can be inserted into the table using similar statements.

Table 1.1
The Initial STUDENT Table

stu#	name	address

Table 1.2
The STUDENT Table After One Record Inserted

stu#	name	address
10	Lew, Tom	2 Main St., Churchill, Australia

Information in the table is retrieved using the SQL *select* command. For example, if we want to retrieve the name of a student with student number 32, we use the following query statement:

```
select name
from STUDENT
where stu#=32
```

Attributes in a RDBMS have fixed types with fixed widths. In the above example, the attribute stu# is an integer type of fixed length of 32 bits. Thus RDBMSs are well suited for handling numeric data and short alphanumeric strings.

To support large variable fields in a RDBMS, a concept called binary large objects (BLOBs) was introduced. A BLOB is a large bit string of variable length. For example, if we want to store students' pictures in the above student record example, we can create a table using the following statement:

```
create table STUDENT (
      stu# integer,
      name char(20),
      address char(100)
      picture BLOB);
```

BLOBs are normally just bit strings and operations such as comparison can not be carried out on them. That is, a RDBMS does not know the contents or semantics of a BLOB. All it knows is a block of data.

Another type of DBMSs is object-oriented database management systems (OODBMSs). Note that a detailed coverage of object-oriented techniques is beyond the scope of this book. The reader is referred to [10, 12, 13] for a complete coverage. OODBMSs combine database capabilities (such as store and search) and object-oriented features (encapsulation, inheritance, and object identity). One common approach is to combine object-oriented features with a relational database. The combined system is called an object relational database system. In such a system, objects are properly defined in the object-oriented sense. That is, each object contains properties or attributes and methods or functions used to manipulate the properties. For example, we can define a class of type called IMAGE as follows:

```
create type IMAGE (
private
      size integer,
      resolution integer,
      content float[],
public
    ...
);
```

We then declare picture as a type of IMAGE that can be used in a table as follows:

```
create table STUDENT (
      stu# integer,
      name char(20),
      address char(100)
      picture IMAGE);
```

The main difference between the BLOB and the object is that the object is properly defined, including its properties and allowed operations on the properties, while the BLOB is not.

The concepts of BLOBs and objects are a step toward handling multimedia data [5, 10–12]. But BLOBs are used just to store large data. While objects contain some simple attributes, mANY more capabilities should be developed to handle content-based multimedia retrieval. Some of the required capabilities are as follows:

• Tools, to automatically, or semiautomatically extract contents and features contained in multimedia data;

• Multidimensional indexing structures, to handle multimedia feature vectors;

• Similarity metrics, for multimedia retrieval instead of exact match;

• Storage subsystems, redesigned to cope with the requirements of large size and high bandwidth and meet realtime requirements;

• The user interface, designed to allow flexible queries in different media types and provide multimedia presentations.

The above capabilities and related issues are the focus of this book.

1.2.3— IR Systems and Their Role in Multimedia Retrieval

In addition to DBMSs, there is another type of information management system that focuses on text document retrieval. This type of system is called an information retrieval (IR) system [8, 9]. IR techniques are important in multimedia information management systems for two main reasons. First, there exist a large number of text documents in many organizations such as libraries. Text is a very important information source of an organization. To use information stored in these documents, an efficient and effective IR system is needed. Second, text can be used to annotate other media such as audio, images, and video. Conventional IR techniques can be used for multimedia information retrieval. However, the use of IR for handling multimedia data has the following limitations:

• The annotation is commonly a manual process and time consuming;

• Text annotation is incomplete and subjective;

• IR techniques cannot handle queries in forms other than text (such as audio and images);

• Some multimedia features such as image texture and object shapes are difficult, if not impossible, to describe using text.

1.2.4— Integrated Approach to Multimedia Information Indexing and Retrieval

From the above discussion we see that DBMSs and IR cannot fully meet the requirements of multimedia indexing and retrieval, so new techniques to handle special characteristics of multimedia data are required. Nevertheless, we recognize that DBMSs and IR can play important roles in MMDBMSs. Parts of multimedia data, such as the creation date and author of a multimedia document, are structured. This structured data can be handled with DBMS techniques. Text annotation is still a powerful method for capturing the contents of multimedia data, so IR techniques have an important role to play.

To summarize, an integrated approach combining DBMSs, IR, and specific techniques for handling multimedia data is required to develop an efficient and effective MIRS.

1.3— An Overview of the MIRS

Figure 1.1 provides an overview of MIRS operation. Information items in the database are preprocessed to extract features and semantic contents and are indexed based on these features and semantics. During information retrieval, a user's query is processed and its main features are extracted. These features are then compared with the features or index of each information item in the database. Information items whose features are most similar to those of the query are retrieved and presented to the user.