Multimedia Systems, Standards, and Networks

edited by

ATUL PURI
TSUHAN CHEN

Signal Processing and Communications Series

Multimedia Systems, Standards, and Networks

edited by

Atul Puri

AT&T Labs

Red Bank, New Jersey

Tsuhan Chen

Carnegie Mellon University

Pittsburgh, Pennsylvania

MARCEL DEKKER, INC. NEW YORK • BASEL

ISBN: 0-8247-9303-X

This book is printed on acid-free paper.

Headquarters
Marcel Dekker, Inc.
270 Madison Avenue, New York, NY 10016
tel: 212-696-9000; fax: 212-685-4540

Eastern Hemisphere Distribution
Marcel Dekker AG
Hutgasse 4, Postfach 812, CH-4001 Basel, Switzerland
tel: 41-61-261-8482; fax: 41-61-261-8896

World Wide Web
http://www.dekker.com

The publisher offers discounts on this book when ordered in bulk quantities. For more information, write to Special Sales/Professional Marketing at the headquarters address above.

Current printing (last digit):
10 9 8 7 6 5 4 3

PRINTED IN THE UNITED STATES OF AMERICA

Preface

We humans, being social creatures, have historically felt the need for increasingly sophisticated means to express ourselves through, for example, conversation, stories, pictures, entertainment, social interaction, and collaboration. Over time, our means of expression have included grunts of speech, storytelling, cave paintings, smoke signals, formal languages, stone tablets, printed newspapers and books, telegraphs, telephones, phonographs, radios, theaters and movies, television, personal computers (PCs), compact disc (CD) players, digital versatile disc (DVD) players, mobile phones and similar devices, and the Internet.

Presently, at the dawn of the new millennium, information technology is continuously evolving around us and influencing every aspect of our lives. Powered by high-speed processors, today's PCs, even inexpensive ones, have significant computational capabilities. These machines are capable of efficiently running even fairly complex applications, whereas not so long ago such tasks could often be handled only by expensive mainframe computers or dedicated, expensive hardware devices. Furthermore, PCs when networked offer a low-cost collaborative environment for business or consumer use (e.g., for access and management of corporate information over intranets or for any general information sharing over the Internet). Technological developments such as web servers, database systems, Hypertext Markup Language (HTML), and web browsers have considerably simplified our access to and interaction with information, even if the information resides in many computers over a network. Finally, because this information is intended for consumption by humans, it may be organized not only in textual but also in aural and/or visual forms.

Who Needs This Book?

Multimedia Systems, Standards, and Networks is about recent advances in multimedia systems, standards, and networking. This book is for you if you have ever been interested in efficient compression of images and video and want to find out what is coming next; if you have any interest in upcoming techniques for efficient compression of speech or music, or efficient representation of graphics and animation; if you have heard about existing or evolving ITU-T video standards as well as Moving Picture Experts Group (MPEG) video and audio standards and want to know more; if you have ever been curious about the space needed for storage of multimedia on a disc or bandwidth issues in transmission of multimedia over networks, and how these problems can be addressed by new coding standards; and finally (because it is not only about efficient compression but also about effective playback systems) if you want to learn more about flexible composition and user interactivity, over-the-network streaming, and search and retrieval.

What Is This Book About?

This is not to say that efficient compression is no longer important—in fact, this book pays a great deal of attention to that topic—but as compression technology undergoes standardization, matures, and is deployed in multimedia applications, many other issues are becoming increasingly relevant. For instance, issues in system design for synchronized playback of several simultaneous audio-visual streams are important. Also increasingly important is the capability for enhanced interaction of user with the content, and streaming of the same coded content over a variety of networks. This book addresses all these facets mainly by using the context of two recent MPEG standards. MPEG has a rich history of developing pioneering standards for digital video and audio coding, and its standards are currently used in digital cable TV, satellite TV, video on PCs, high-definition television, video on CD-ROMs, DVDs, the Internet, and much more. This book addresses two new standards, MPEG-4 and MPEG-7, that hold the potential of impacting many future applications, including interactive Internet multimedia, wireless videophones, multimedia search/browsing engines, multimedia-enhanced e-commerce, and networked computer video games. But before we get too far, it is time to briefly introduce a few basic terms.

So what is multimedia? Well, the term *multimedia* to some conjures images of cinematic wizardry or audiovisual special effects, whereas to others it simply means video with audio. Neither of the two views is totally accurate. We use the term *multimedia* in this book to mean digital multimedia, which implies the use of several digitized media simultaneously in a synchronized or related manner. Examples of various types of media include speech, images, text/graphics, audio, video, and computer animation. Furthermore, there is no strict requirement that all of these different media ought to be simultaneously used, just that more than one media type may be used and combined with others as needed to create an interesting multimedia presentation.

What do we mean by a *multimedia system*? Consider a typical multimedia presentation. As described, it may consist of a number of different streams that need to be continuously decoded and synchronized for presentation. A multimedia system is the entity that actually performs this task, among others. It ensures proper decoding of individual media streams. It ties the component media contained in the multimedia stream. It guarantees proper synchronization of individual media for playback of a presentation. A multimedia

system may also check for and enforce intellectual property rights with respect to multimedia content.

Why do we need *multimedia standards*? Standards are needed to guarantee interoperability. For instance, a decoding device such as a DVD player can decode multimedia content of a DVD disc because the content is coded and formatted according to rules understood by the DVD player. In addition, having internationally uniform standards implies that a DVD disc bought anywhere in the world may be played on any DVD player. Standards have an important role not only in consumer electronics but also in multimedia communications. For example, a videotelephony system can work properly only if the two endpoints that want to communicate are compatible and each follows protocols that the other can understand. There are also other reasons for standards; e.g., because of economies of scale, establishment of multimedia standards allows devices, content, and services to be produced inexpensively.

What does *multimedia networking* mean? A multimedia application such as playing a DVD disc on a DVD player is a stand-alone application. However, an application requiring downloading of, for example, MP3 music content from a Web site to play on a hardware or software player uses networking. Yet another form of multimedia networking may involve playing streaming video where multimedia is chunked and transmitted to the decoder continuously instead of the decoder having to wait to download all of it. Multimedia communication applications such as videotelephony also use networking. Furthermore, a multiplayer video game application with remote players also uses networking. In fact, whether it relates to consumer electronics, wireless devices, or the Internet, multimedia networking is becoming increasingly important.

What Is in This Book?

Although an edited book, *Multimedia Systems, Standards, and Networks* has been painstakingly designed to have the flavor of an authored book. The contributors are the most knowledgeable about the topic they cover. They have made numerous technology contributions and chaired various groups in development of the ITU-T H.32x, H.263, or ISO MPEG-4 and MPEG-7 standards.

This book comprises 22 chapters. Chapters 1, 2, 3, and 4 contain background material including that on the ITU-T as well as ISO MPEG standards. Chapters 5 and 6 focus on MPEG-4 audio. Chapters 7, 8, 9, 10, and 11 describe various tools in the MPEG-4 Visual standard. Chapters 12, 13, 14, 15, and 16 describe important aspects of MPEG-4 Systems standard. Chapters 17, 18, and 19 discuss multimedia over networks. Chapters 20, 21, and 22 address multimedia search and retrieval as well as MPEG-7. We now elaborate on the contents of individual chapters.

Chapter 1 traces the history of technology and communication standards, along with recent developments and what can be expected in the future.

Chapter 2 presents a technical overview of the ITU-T H.323 and H.324 standards and discusses the various components of these standards.

Chapter 3 reviews the ITU-T H.263 (or version 1) standard as well as the H.263 version 2 standard. It also discusses the H.261 standard as the required background material for understanding the H.263 standards.

Chapter 4 presents a brief overview of the various MPEG standards to date. It thus addresses MPEG-1, MPEG-2, MPEG-4, and MPEG-7 standards.

Chapter 5 presents a review of the coding tools included in the MPEG-4 natural audio coding standard.

Chapter 6 reviews synthetic audio coding and synthetic natural hybrid coding (SNHC) of audio in the MPEG-4 standard.

Chapter 7 presents a high-level overview of the visual part of the MPEG-4 visual standard. It includes tools for coding of natural as well as synthetic video (animation).

Chapter 8 is the first of two chapters that deal with the details of coding natural video as per the MPEG-4 standard. It addresses rectangular video coding, scalability, and interlaced video coding.

Chapter 9 is the second chapter that discusses the details of coding of natural video as per the MPEG-4 standard. It also addresses coding of arbitrary-shape video objects, scalability, and sprites.

Chapter 10 discusses coding of still-image texture as specified in the visual part of the MPEG-4 standard. Both rectangular and arbitrary-shape image textures are supported.

Chapter 11 introduces synthetic visual coding as per the MPEG-4 standard. It includes 2D mesh representation of visual objects, as well as definition and animation of synthetic face and body.

Chapter 12 briefly reviews various tools and techniques included in the systems part of the MPEG-4 standard.

Chapter 13 introduces the basics of how, according to the systems part of the MPEG-4 standard, the elementary streams of coded audio or video objects are managed and delivered.

Chapter 14 discusses scene description and user interactivity according to the systems part of the MPEG-4 standard. Scene description describes the audiovisual scene with which users can interact.

Chapter 15 introduces a flexible MPEG-4 system based on Java programming language; this system exerts programmatic control on the underlying fixed MPEG-4 system.

Chapter 16 presents the work done within MPEG in software implementation of the MPEG-4 standard. A software framework for 2D and 3D players is discussed mainly for the Windows environment.

Chapter 17 discusses issues that arise in the transport of general coded multimedia over asynchronous transfer mode (ATM) networks and examines potential solutions.

Chapter 18 examines key issues in the delivery of coded MPEG-4 content over Internet Protocol (IP) networks. The MPEG and Internet Engineering Task Force (IETF) are jointly addressing these as well as other related issues.

Chapter 19 introduces the general topic of delivery of coded multimedia over wireless networks. With the increasing popularity of wireless devices, this research holds significant promise for the future.

Chapter 20 reviews the status of research in the general area of multimedia search and retrieval. This includes object-based as well as semantics-based search and filtering to retrieve images and video.

Chapter 21 reviews the progress made on the topic of image search and retrieval within the context of a digital library. Search may use a texture dictionary, localized descriptors, or regions.

Chapter 22 introduces progress in MPEG-7, the ongoing standard focusing on content description. MPEG-7, unlike previous MPEG standards, addresses search/retrieval and filtering applications, rather than compression.

Now that you have an idea of what each chapter covers, we hope you enjoy *Multimedia Systems, Standards, and Networks* and find it useful. We learned a great deal—and had a great time—putting this book together. Our heartfelt thanks to all the contributors for their enthusiasm and hard work. We are also thankful to our management, colleagues, and associates for their suggestions and advice throughout this project. We would like to thank Trista Chen, Fu Jie Huang, Howard Leung, and Deepak Turaga for their assistance in compiling the index. Last, but not least, we owe thanks to B. J. Clarke, J. Roh, and M. Russell along with others at Marcel Dekker, Inc.

Atul Puri
Tsuhan Chen

Contributors

Olivier Avaro Deutsche Telekom-Berkom GmbH, Darmstadt, Germany

Vahe Balabanian Nortel Networks, Nepean, Ontario, Canada

Mahesh Balakrishnan Philips Research, Briarcliff Manor, New York

Andrea Basso Broadband Communications Services Research, AT&T Labs, Red Bank, New Jersey

Stefano Battista bSoft, Macerata, Italy

Serge J. Belongie Computer Science Division, EECS Department, University of California at Berkeley, Berkeley, California

Chad Carson Computer Science Division, EECS Department, University of California at Berkeley, Berkeley, California

Shih-Fu Chang Department of Electrical Engineering, Columbia University, New York, New York

Tsuhan Chen Carnegie Mellon University, Pittsburgh, Pennsylvania

Xuemin Chen General Instrument, San Diego, California

Leonardo Chiariglione Television Technologies, CSELT, Torino, Italy

Mehmet Reha Civanlar Speech and Image Processing Research Laboratory, AT&T Labs, Red Bank, New Jersey

Yining Deng Electrical and Computer Engineering Department, University of California at Santa Barbara, Santa Barbara, California

Gianluca De Petris CSELT, Torino, Italy

Gianluca Di Cagno Services and Applications, CSELT, Torino, Italy

Peter K. Doenges Evans & Sutherland, Salt Lake City, Utah

F. Dufaux Compaq, Cambridge, Massachusetts

Touradj Ebrahimi Signal Processing Laboratory, Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland

Alexandros Eleftheriadis Department of Electrical Engineering, Columbia University, New York, New York

Gerard Fernando Sun Microsystems, Menlo Park, California

Yuval Fisher Institute for Nonlinear Science, University of California at San Diego, La Jolla, California

David A. Forsyth Computer Science Division, EECS Department, University of California at Berkeley, Berkeley, California

Guido Franceschini Services and Applications, CSELT, Torino, Italy

Jean Gelissen Nederlandse Phillips Bedrijven, Eindhoven, The Netherlands

Bernhard Grill Fraunhofer Geselshaft IIS, Erlangen, Germany

Barry G. Haskell AT&T Labs, Red Bank, New Jersey

Kaynam Hedayat Brix Networks, Billerica, Massachusetts

Carsten Herpel Thomson Multimedia, Hannover, Germany

Jürgen Herre Fraunhofer Geselshaft IIS, Erlangen, Germany

Caspar Horne Mediamatics, Inc., Fremont, California

Qian Huang AT&T Labs, Red Bank, New Jersey

Thomas Huang Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, Illinois

Sergey Ioffe Computer Science Division, EECS Department, University of California at Berkeley, Berkeley, California

James D. Johnston AT&T Labs, Florham Park, New Jersey

Rob H. Koenen Multimedia Technology Group, KPN Research, Leidschendam, The Netherlands

Youngjik Lee ETRI Switching & Transmission Technology Laboratories, Taejon, Korea

Shipeng Li Microsoft Research China, Beijing, China

Weiping Li Optivision, Inc., Palo Alto, California

Jie Liang Texas Instruments, Dallas, Texas

Zvi Lifshitz Triton R&D Ltd., Jerusalem, Israel

Ajay Luthra General Instrument, San Diego, California

Wei-Ying Ma Hewlett-Packard Laboratories, Palo Alto, California

Jitendra Malik Computer Science Division, EECS Department, University of California at Berkeley, Berkeley, California

Bangalore S. Manjunath Electrical and Computer Engineering Department, University of California at Santa Barbara, Santa Barbara, California

Y. Nakaya Hitachi Ltd., Tokyo, Japan

Chiu Yeung Ngo Video Communications, Philips Research, Briarcliff Manor, New York

Joern Ostermann AT&T Labs, Red Bank, New Jersey

Fernando Pereira Instituto Superior Técnico/Instituto de Telecommunicações, Lisbon, Portugal

Eric Petajan Lucent Technologies, Murray Hill, New Jersey

G.D. Petris CSELT, Torino, Italy

Atul Puri AT&T Labs, Red Bank, New Jersey

Schuyler R. Quackenbush AT&T Labs, Florham Park, New Jersey

Hayder Radha Philips Research, Briarcliff Manor, New York

Ganesh Rajan General Instrument, San Diego, California

Dipankar Raychaudhuri C&C Research Laboratories, NEC USA, Inc., Princeton, New Jersey

Daniel J. Reininger C&C Research Laboratories, NEC USA, Inc., Princeton, New Jersey

Takashi Sato Philips Research, Briarcliff Manor, New York

Richard Schaphorst Delta Information Systems, Horsham, Pennsylvania

Eric D. Scheirer Machine Listening Group, MIT Media Laboratory, Cambridge, Massachusetts

Robert L. Schmidt AT&T Labs, Red Bank, New Jersey

Behzad Shahraray AT&T Labs, Red Bank, New Jersey

Julien Signès Research and Development, France Telecom Inc., Brisbane, California

Iraj Sodagar Sarnoff Corporation, Princeton, New Jersey

Gary J. Sullivan Picture Tel Corporation, Andover, Massachusetts

Viswanathan Swaminathan Sun Microsystems, Menlo Park, California

Raj Talluri Texas Instruments, Dallas, Texas

Peter van Beek Sharp Laboratories of America, Camas, Washington

Liam Ward Teltec Ireland, DCU, Dublin, Ireland

Jae-Woo Yang ETRI Switching & Transmission Technology Laboratories, Taejon, Korea

Ya-Qin Zhang Microsoft Research China, Beijing, China

Contents

1

Communication Standards: Götterdämmerung?

Leonardo Chiariglione
CSELT, Torino, Italy

I. INTRODUCTION

Communication standards are at the basis of civilized life. Human beings can achieve collective goals through sharing a common understanding that certain utterances are associated with certain objects, concepts, and all the way up to certain intellectual values. Civilization is preserved and enhanced from generation to generation because there is an agreed mapping between certain utterances and certain signs on paper that enable a human being to leave messages to posterity and posterity to revisit the experience of people who have long ago departed.

Over the centuries, the simplest communication means that have existed since the remotest antiquity have been supplemented by an endless series of new ones: printing, photography, telegraphy, telephony, television, and the new communication means such as electronic mail and the World Wide Web.

New inventions made possible new communication means, but before these could actually be deployed some agreements about the meaning of the "symbols" used by the communication means was necessary. Telegraphy is a working communication means only because there is an agreement on the correspondence between certain combinations of dots and dashes and characters, and so is television because there is an agreed procedure for converting certain waveforms into visible and audible information. The ratification and sometimes the development of these agreements—called standards—are what standards bodies are about. Standards bodies exist today at the international and national levels, industry specific or across industries, tightly overseen by governments or largely independent.

Many communication industries, among these the telecommunication and broadcasting industries, operate and prosper thanks to the existence of widely accepted standards. They have traditionally valued the role of standards bodies and have often provided their best personnel to help them achieve their goal of setting uniform standards on behalf of their industries. In doing so, they were driven by their role of "public service" providers,

---

Götterdämmerung: Twilight of the Gods. See, e.g., *http://walhall.com/*

a role legally sanctioned in most countries until very recently. Other industries, particularly the consumer electronics and computer industry, have taken a different attitude. They have "defined" communication standards either as individual companies or as groups of companies and then tried to impose their solution on the marketplace. In the case of a successful outcome, they (particularly the consumer electronics industry) eventually went to a standards body for ratification.

The two approaches have been in operation for enough time to allow some comparisons to be drawn. The former has given stability and constant growth to its industries and universal service to the general citizenship, at the price of a reduced ability to innovate: the telephone service is ubiquitous but has hardly changed in the past 100 years; television is enjoyed by billions of people around the world but is almost unchanged since its first deployment 60 years ago. The latter, instead, has provided a vibrant innovative industry. Two examples are provided by the personal computer (PC) and the compact disc. Both barely existed 15 years ago, and now the former is changing the world and the latter has brought spotless sound to hundreds of millions of homes. The other side of the coin is the fact that the costs of innovation have been borne by the end users, who have constantly struggled with incompatibilities between different pieces of equipment or software ("I cannot open your file") or have been forced to switch from one generation of equipment to the next simply because some dominant industry decreed that such a switch was necessary.

Privatization of telecommunication and media companies in many countries with renewed attention to the cost–benefit bottom line, the failure of some important standardization projects, the missing sense of direction in standards, and the lure that every company can become "the new Microsoft" in a business are changing the standardization landscape. Even old supporters of formal standardization are now questioning, if not the very existence of those bodies, at least the degree of commitment that was traditionally made to standards development.

The author of this chapter is a strong critic of the old ways of formal standardization that have led to the current diminished perception of its role. Having struggled for years with incompatibilities in computers and consumer electronics equipment, he is equally adverse to the development of communication standards in the marketplace. He thinks the time has come to blend the good sides of both approaches. He would like to bring his track record as evidence that a Darwinian process of selection of the fittest can and should be applied to standards *making* and that having standards is good to expand existing business as well as to create new ones. All this should be done not by favoring any particular industry, but working for all industries having a stake in the business.

This chapter revisits the foundations of communication standards, analyzes the reasons for the decadence of standards bodies, and proposes a framework within which a reconstruction of standardization on new foundations should be made.

II. COMMUNICATION SYSTEMS

Since the remotest antiquity, language has been a powerful communication system capable of conveying from one mind to another simple and straightforward as well as complex and abstract concepts. Language has not been the only communication means to have accompanied human evolution: body gesture, dance, sculpture, drawing, painting, etc. have all been invented to make communication a richer experience.

Writing evolved from the last two communication means. Originally used for point-to-point communication, it was transformed into a point-to-multipoint communication means by amanuenses. Libraries, starting with the Great Library of Alexandria in Egypt, were used to store books and enable access to written works.

The use of printing in ancient China and, in the West, Gutenberg's invention brought the advantage of making the reproduction of written works cheaper. The original simple system of book distribution eventually evolved to a two-tier distribution system: a network of shops where end users could buy books. The same distribution system was applied for newspapers and other periodicals.

Photography enabled the automatic reproduction of a natural scene, instead of hiring a painter. From the early times when photographers built everything from cameras to light-sensitive emulsions, this communication means has evolved to a system where films can be purchased at shops that also collect the exposed films, process them, and provide the printed photographs.

Postal systems existed for centuries, but their use was often restricted to kings or the higher classes. In the first half of the 19th century different systems developed in Europe that were for general correspondence use. The clumsy operational rules of these systems were harmonized in the second half of that century so that prepaid letters could be sent to all countries of the Universal Postal Union (UPU).

The exploitation of the telegraph (started in 1844) allowed the instant transmission of a message composed of Latin characters to a distant point. This communication system required the deployment of an infrastructure—again two-tier—consisting of a network of wires and of telegraph offices where people could send and receive messages. Of about the same time (1850) is the invention of facsimile, a device enabling the transmission of the information on a piece of paper to a distant point, even though its practical exploitation had to wait for another 100 years before effective scanning and reproduction techniques could be employed. The infrastructure needed by this communication system was the same as the telephony's.

Thomas A. Edison's phonograph (1877) was another communication means that enabled the recording of sound for later playback. Creation of the master and printing of disks required fairly sophisticated equipment, but the reproduction equipment was relatively inexpensive. Therefore the distribution channel developed in a very similar way as for books and magazines.

If the phonograph had allowed sound to cross the barriers of time and space, telephony enabled sound to overcome the barriers of space in virtually no time. The simple point-to-point model of the early years gave rise to an extremely complex hierarchical system. Today any point in the network can be connected with any other point.

Cinematography (1895) made it possible for the first time to capture not just a snapshot of the real world but a series of snapshots that, when displayed in rapid succession, appeared to reproduce something very similar to real movement to the eye. The original motion pictures were later supplemented by sound to give a complete reproduction to satisfy both the aural and visual senses.

The exploitation of the discovery that electromagnetic waves could propagate in the air over long distances produced wireless telegraphy (1896) and sound broadcasting (1920). The frequencies used at the beginning of sound broadcasting were such that a single transmitter could, in principle, reach every point on the globe by suitably exploiting propagation in the higher layers of atmosphere. Later, with the use of higher frequencies,