

MULTIMEDIA COMPUTING



Gerald Friedland
Ramesh Jain

"This book fills a critical need in multimedia computing by offering a comprehensive, authoritative, and holistic review of key principles, state-of-the-art technologies, and open issues in this vibrant field. Authored by two renowned pioneers and educators, it will be a wonderful book for students and anyone interested in learning more about multimedia."

– Shih-Fu Chang, Department of Electrical Engineering and Computer Science,
Columbia University

"The multimedia field needed a textbook, and it is finally here. This is a comprehensive textbook covering a diverse set of topics in modern multimedia. Each chapter is very accessible; this book will be an excellent resource for beginning graduate students and for practitioners."

– Professor Mubarak Shah, Center for Research in Computer Vision,
University of Central Florida

"This is the long-awaited textbook on multimedia. The authors, both senior members of the multimedia community, have managed to produce a book that is useful not only to students and practitioners but also to many of us in the field. The book is comprehensive in its coverage of the subject, and it is structured in an easy-to-read way. I particularly like the fact that the exercises are meant to solicit the readers to think beyond the actual content of the book."

– Professor Nicu Sebe, University of Trento

"Teaching multimedia is challenging because it encompasses concepts and technology from many areas, including human perception, digital representation of continuous and discrete data and behavior, synchronization in space and time, and distributed real-time software/hardware systems. Applications range from multimedia authoring, analysis, and retrieval to interactive entertainment and distributed collaboration. *Multimedia Computing* covers a wide range of material suitable for a one- to two-semester introduction to multimedia. It covers the fundamentals of human perception that produce multimedia experiences through the systems and applications used to produce these experiences. Friedland and Jain do an excellent job of presenting enough details and enough guidance to further sources for students to understand the scope of the field and to learn more about topics of interest."

– Lawrence A. Rowe, Professor Emeritus, EECS Department,
University of California, Berkeley

"*Multimedia Computing* is indeed an excellent resource – a one-stop resource – for students, young researchers, and professionals. The authors have provided the basics from the physics concepts of sound and light to user interface design for multimedia systems. A must-read book."

– B. Prabhakaran, Professor and ACM Distinguished Scientist, University of Texas at Dallas

Humans are the best functioning example of multimedia communication and computing – that is, we understand information and experiences through the unified perspective offered by our five senses. This innovative textbook presents emerging techniques in multimedia computing from an experiential perspective in which each medium – audio, images, text, and so on – is a strong component of the complete, integrated exchange of information or experience. The authors' goal is to present current techniques in computing and communication that will lead to the development of a unified and holistic approach to computing using heterogeneous data sources.

Gerald Friedland and Ramesh Jain introduce the fundamentals of multimedia computing, describing the properties of perceptually encoded information, presenting common algorithms and concepts for handling it, and outlining the typical requirements for emerging applications that use multifarious information sources. Designed for advanced undergraduate and beginning graduate courses, the book will also serve as an introduction for engineers and researchers interested in understanding the elements of multimedia and their role in building specific applications.

CAMBRIDGE
UNIVERSITY PRESS
www.cambridge.org

ISBN 978-0-521-76451-3



9 780521 764513 >

Cover image courtesy of Claudiad / istockphoto.com

Cover design by James F. Brisson

FRIEDLAND
JAIN
MULTIMEDIA
COMPETING

CAMBRIDGE

Multimedia Computing

GERALD FRIEDLAND

International Computer Science Institute, Berkeley, California

RAMESH JAIN

University of California, Irvine



CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

32 Avenue of the Americas, New York, NY 10013-2473, USA

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning, and research at the highest international levels of excellence.

www.cambridge.org

Information on this title: www.cambridge.org/9780521764513

© Gerald Friedland and Ramesh Jain 2014

This publication is in copyright. Subject to statutory exception and to the provisions of relevant collective licensing agreements, no reproduction of any part may take place without the written permission of Cambridge University Press.

First published 2014

Printed in the United States of America

A catalog record for this publication is available from the British Library.

Library of Congress Cataloging in Publication data

Friedland, Gerald, 1978–

Multimedia computing / Gerald Friedland, Ramesh Jain.
pages cm

Includes bibliographical references and index.

ISBN 978-0-521-76451-3 (hardback)

1. Multimedia systems. I. Jain, Ramesh, 1949– II. Title.

QA76.575.F76 2014

006.7–dc23 2014002250

ISBN 978-0-521-76451-3 Hardback

Additional resources for this publication at <http://www.mm-creole.org>

Cambridge University Press has no responsibility for the persistence or accuracy of URLs for external or third-party Internet Web sites referred to in this publication and does not guarantee that any content on such Web sites is, or will remain, accurate or appropriate.

MULTIMEDIA COMPUTING

The word “multimedia” is often associated with specific applications from entertainment to Web design to video to music. This textbook presents emerging techniques in multimedia computing from an experiential perspective in which each medium – audio, images, text, and so on – is a strong component of the complete, integrated exchange of information or experience. Humans are the best functioning example of multimedia communication and computing – that is, we understand information and experiences through the unified perspective the combination of our senses offers. The goal of this book is to present current techniques in computing and communication that will lead to the development of a unified and holistic approach to computing using heterogeneous data sources.

The authors introduce the fundamentals of multimedia computing, describing the properties of perceptually encoded information, presenting common algorithms and concepts for handling it, and outlining the typical requirements for emerging applications that use multifarious information sources. Designed for advanced undergraduate and beginning graduate courses, this book will also serve as an introduction for engineers and researchers interested in understanding the elements of multimedia and their role in building specific applications.

Gerald Friedland is the director of audio and multimedia research at the International Computer Science Institute, a research laboratory affiliated with the University of California, Berkeley.

Ramesh Jain is a Donald Bren Professor in information and computer sciences at the University of California, Irvine, where he is doing research in EventWeb and experiential computing for developing and building social life networks.

Preface

The project to write a textbook on multimedia computing started a few years ago, when the coauthors independently realized that a book that addresses the basic concepts related to the increasing volume of multimedia data in different aspects of communications in the computing age is needed.

Digital computing started with processing numbers, but very soon after its start it began dealing with symbols of other kinds and developed computational approaches for dealing with alpha-numeric data. Efforts to use computers to deal with audio, visual, and other perceptual information did not become successful enough to be used for any applications until about the 1980s. Only slowly, computer graphics, audio processing, and visual analysis started becoming feasible. First was the ability to store large volumes of audiovisual data, then displaying or rendering it, then distributing it, and later processing and analyzing it. For that reason, it took until about the 1990s for the term *multimedia* to grow popular in computing.

While different fields have emerged around acoustic, visual, and natural text content that specializes in these data types, multimedia computing deals with documents holistically, taking into account all media available. Dominated by the availability of electronic sensors, multimedia communication is currently focused on visual and audio, followed by metadata (such as GPS) and touch. Multimedia computing deals with multiple media holistically because the purpose of documents that contain multiple media is to communicate information. Information about almost all real-world events and objects must be captured using multiple sensors as each sensor only captures one aspect of the information of interest. The challenge for multimedia computing systems is to integrate the different information streams into a coherent view. Humans do this every moment of their lives from birth and are therefore often used as a baseline when building multimedia systems. Therefore it's not surprising that, slowly but surely, all computing is becoming multimedia. The use of multimedia data in computing has grown even more rapidly than imagined just a few years ago with the installation of cameras in cell phones in combination with the ability to share multimedia documents in social networks easily.

It was at this point that we felt that most of the books introducing multimedia to students and practitioners discussed technology from a perspective that has now become obsolete. We wanted to introduce multimedia from a more current, yet sustainable, perspective.

In this book, we adopt a holistic approach in introducing the field of multimedia computing. Instead of introducing multimedia based on the media types (image, audio, video), we present multimedia as a fundamental and unique discipline that takes advantage of all media available to solve problems. This results in a unified introduction to the field by focusing on the fundamental techniques and mathematical foundations of multimedia computing as a process in information extraction from diverse types of signals. For instance, the book introduces lossy compression by presenting the principles of quantization and differential coding, before showing how these principles are applied to compression of image, audio, and video.

Having said that, covering all aspects of different types of signals used in multimedia computing in a book is not practical. This would require covering large parts of mathematics, physics, physiology, psychology, electrical engineering, and computer science aspects of different types of signals and related topics. We therefore decided to adopt a different approach: introduce the main concepts in a “capsule” form and provide pseudo-code for algorithms such as ADPCM encoding or the Paeth Predictor. We decided to include many research papers, blogs, and books giving details and detailed perspectives on different topics under further readings. We will keep these updated in a companion Web version of the book at <http://www.mm-creole.org>. Advanced and detailed literature references are provided that allow the reader to explore the topics beyond the coverage of this book.

We want our book to be used as a textbook for advanced undergraduate or graduate students in computer science and computer engineering. Students in other areas might also benefit, but they must have good preparation in programming, data structures, and data management. Many computer science, computer engineering, and electrical and computer engineering programs offer a course in multimedia. This book is targeted to be suitable to support these classes.

We intend to make this a “living book.” Based on reader feedback, development of new methods and technology, and emerging applications, we will keep updating the book. As we are finishing the last words for this version we are well aware that this will not be the end of the project.

We therefore hope you enjoy reading this book and start engaging with this exciting field and, while you do, don’t hesitate to contact us to participate in the evolution of this book.

Gerald Friedland Ramesh Jain
Berkeley and Irvine
November 18, 2013

Acknowledgments

A textbook project is a long one and requires significant efforts by several people. This book would not have been possible without the encouragement of many colleagues at ACM Multimedia who reminded us, year by year, of the necessity of this book. These include especially Klara Nahrstedt, Balakrishnan Prabhakaran, and Abdulmoteleb El Saddik. Moreover, we want to thank the various students who provided us with feedback during the course of this project. We want to thank our editors Lauren Cowles and David Jou at Cambridge University Press. Furthermore, we want to thank Joan Taylor for editing early versions of our nonnative English creations that we dared to call text. Last but not least, we want to thank our families, the Friedlands and the Jains, for bearing long weekends and nights of work.

Contents

Preface..... page xi

Acknowledgments..... xiii

1. Introduction 1

 Organization of This Book 3

 Recommended Reading 4

2. Multimedia: A Definition 6

 Communication in Human Society 7

 Evolution of Computing and Communication Technology 8

 Why *Multimedia* Computing? 10

 Multimedia: A More Formal Definition 13

 Recommended Reading 13

 Exercises 14

3. Elements of Multimedia Computing 15

 Experience and Information 15

 Objects and Events 17

 Perception 18

 Perceptual Cycle 19

 Challenges in Multimedia Systems 21

 Semantic Gap 22

 Context and Content 23

 Recommended Reading 25

 Exercises 26

4. Introduction to Sensors 28

 Types of Sensors 28

 Properties of Sensors 31

 Digitization 32

 Recommended Reading 34

 Exercises 35

5. Sound	36
The Physics of Sounds	36
Observed Properties of Sound	38
Recording and Reproduction of Sound	39
Microphones	39
Reproduction of Sound	41
Production of Sound	43
Exercises	49
6. Light	51
What Is Light?	51
Observed Properties of Light	53
Recording Light	54
Reproducing Light	58
Perception of Light	60
Color Spaces	61
Light Production	64
Recommended Reading	64
Exercises	66
7. Multimedia Documents	67
What Is a Document?	67
Evolving Nature of Documents	69
Stages in Document Creation	71
Basic Elements of a Multimedia Authoring Environment	75
Representation of a Multimedia Document	77
Current Authoring Environments	78
Recommended Reading	79
Exercises	80
8. Multimodal Integration and Synchronization	82
Multimodal Integration	82
Sensor Integration in Multimedia Computing	86
Introduction to Synchronization	86
Content Synchronization	87
Temporal Synchronization	88
Synchronization Levels in a System	90
Specification of Synchronization	90
Deadlines	91
Spatial Synchronization	91
Recommended Reading	93
Exercises	94

9. Multimedia Systems	95
Components of a Multimedia System	95
Different Configurations of Multimedia Nodes	99
QoS and QoE	99
Different Modes of Media Delivery	99
Live Video	103
Emerging Systems	105
Recommended Reading	105
Exercises	107
10. The Human Factor	108
Principles of User Interface Design	108
Privacy	116
Security	118
Safety	119
Recommended Reading	120
Exercises	122
11. Fundamentals of Compression	124
Run-Length Coding	124
Information Content and Entropy	125
Compression Algorithms	128
Weakness of Entropy-Based Compression Methods for Multimedia Data	138
Recommended Reading	138
Exercises	139
12. Lossy Compression	141
Mathematical Foundation: Vector Quantization	141
Perceptual Quantization	144
Differential Coding	147
Recommended Reading	153
Standards	154
Exercises	154
13. Advanced Perceptual Compression	156
Discrete Fourier Transform (DFT)	156
Discrete Cosine Transform (DCT)	158
The Convolution Theorem	160
JPEG	161
Psychoacoustics	164
MP3	168
Perceptual Video Compression	170
Recommended Reading	171
Exercises	172

14. Speech Compression	174
Properties of a Speech Coder	174
Linear Predictive Coding (LPC)	175
CELP	182
GSM	185
Recommended Reading	186
Exercises	187
15. Multimedia Information Retrieval	188
Databases	189
Structured, Unstructured, and Semistructured Data	191
Classic Information Retrieval and Search	192
From Information Retrieval to Multimedia Information	
Retrieval	198
Content-Based Retrieval	199
Recommended Reading	205
Exercises	207
16. Signal Processing Primer	209
Sampling and Quantization	209
Levels of Computation	210
Thresholding	212
Linear Operations	214
Nonlinear Filters	216
Filter by Example	220
Graphical Filters	222
Notions for Visual Filters	222
Edges	226
Segmentation Using Split and Merge	228
Recommended Reading	230
Standards	232
Exercises	232
17. Multimedia Content Analysis	233
From Monomedia to Multimedia Content Analysis	234
Setup of a Multimedia Content Analysis Experiment	236
Features	238
Supervised Learning	248
Unsupervised Modeling	260
Error Measurement and Evaluation	262
Final Remarks	266
Recommended Reading	269
Exercises	270

18. Content Analysis Systems	272
Speech Activity Detection	272
Large Vocabulary Automatic Speech Recognition	273
Optical Character Recognition	275
Speaker Recognition	276
Speaker Diarization	278
Face Detection and Face Recognition	280
Acoustic Event Detection	281
Visual Object Recognition and Tracking	282
Multimedia Content Analysis Systems	286
Recommended Reading	287
Exercises	288
19. Content and Context	290
Connecting Data and Users	292
Content and Context	293
Context-Only Image Search: Commercial Systems	297
Device Parameters	297
Recommended Reading	302
Exercises	304
20. Future Topics	305
How to Predict the Future	305
Future Topics in Multimedia	306
Recommended Reading	315
Exercises	315
<i>Index</i>	317

Plates follow page 50.

