
Ze-Nian Li • Mark S. Drew •
Jiangchuan Liu

Fundamentals of Multimedia

Third Edition

 Springer

Contents

Part I Introduction and Multimedia Data Representations

1	Introduction to Multimedia	3
1.1	What is Multimedia?	3
1.1.1	Components of Multimedia	4
1.2	Multimedia: Past and Present	5
1.2.1	Early History of Multimedia	5
1.2.2	Hypermedia, WWW, and Internet	9
1.2.3	Multimedia in the New Millennium	13
1.3	Multimedia Software Tools: A Quick Scan	16
1.3.1	Music Sequencing and Notation	17
1.3.2	Digital Audio	17
1.3.3	Graphics and Image Editing	18
1.3.4	Video Editing	18
1.3.5	Animation	19
1.3.6	Multimedia Authoring	20
1.3.7	Multimedia Broadcasting	21
1.4	The Future of Multimedia	22
1.5	Exercises	24
	References	25
2	A Taste of Multimedia	27
2.1	Multimedia Tasks and Concerns	27
2.2	Multimedia Presentation	28
2.3	Data Compression	35
2.4	Multimedia Production	38
2.5	Multimedia Sharing and Distribution	39
2.6	Some Useful Editing and Authoring Tools	41
2.6.1	Adobe Premiere	42
2.6.2	HTML5 Canvas	45
2.6.3	Adobe Director	46
2.6.4	Adobe XD	48

2.7	Exercises	51
	References	55
3	Graphics and Image Data Representations	57
3.1	Graphics and Image Data Types	57
3.1.1	1-Bit Images	57
3.1.2	8-Bit Gray-Level Images	58
3.1.3	Image Data Types	62
3.1.4	24-Bit Color Images	62
3.1.5	Higher Bit-Depth Images	62
3.1.6	8-Bit Color Images	63
3.1.7	Color Lookup Tables (LUTs)	65
3.2	Popular File Formats	69
3.2.1	GIF	69
3.2.2	JPEG	73
3.2.3	PNG	75
3.2.4	TIFF	75
3.2.5	Windows BMP	75
3.2.6	Windows WMF	76
3.2.7	Netpbm Format	76
3.2.8	EXIF	76
3.2.9	HEIF	76
3.2.10	PS and PDF	77
3.2.11	PTM	78
3.3	Exercises	79
	References	81
4	Color in Image and Video	83
4.1	Color Science	83
4.1.1	Light and Spectra	83
4.1.2	Human Vision	85
4.1.3	Spectral Sensitivity of the Eye	85
4.1.4	Image Formation	86
4.1.5	Camera Systems	88
4.1.6	Gamma Correction	88
4.1.7	Color-Matching Functions	91
4.1.8	CIE Chromaticity Diagram	92
4.1.9	Color Monitor Specifications	96
4.1.10	Out-of-Gamut Colors	96
4.1.11	White Point Correction	97
4.1.12	XYZ to RGB Transform	99
4.1.13	Transform with Gamma Correction	99

4.1.14	L*a*b* (CIELAB) Color Model	100
4.1.15	More Color Coordinate Schemes	102
4.1.16	Munsell Color Naming System	102
4.2	Color Models in Images	102
4.2.1	RGB Color Model for Displays	102
4.2.2	Multi-sensor Cameras	103
4.2.3	Camera-Dependent Color	103
4.2.4	Subtractive Color: CMY Color Model	105
4.2.5	Transformation from RGB to CMY	105
4.2.6	Undercolor Removal: CMYK System	106
4.2.7	Printer Gamuts	106
4.2.8	Multi-ink Printers	107
4.3	Color Models in Video	108
4.3.1	Video Color Transforms	108
4.3.2	YUV Color Model	109
4.3.3	YIQ Color Model	111
4.3.4	YCbCr Color Model	112
4.4	Exercises	113
	References	117
5	Fundamental Concepts in Video	119
5.1	Analog Video	119
5.1.1	NTSC Video	122
5.1.2	PAL Video	125
5.1.3	SECAM Video	125
5.2	Digital Video	126
5.2.1	Chroma Subsampling	126
5.2.2	CCIR and ITU-R Standards for Digital Video	128
5.2.3	High Definition TV (HDTV)	129
5.2.4	Ultra-High-Definition TV (UHDTV)	130
5.3	Video Display Interfaces	131
5.3.1	Analog Display Interfaces	131
5.3.2	Digital Display Interfaces	133
5.4	360° Video	135
5.4.1	Equirectangular Projection (ERP)	136
5.4.2	Other Projections	137
5.5	3D Video and TV	138
5.5.1	Cues for 3D Percept	138
5.5.2	3D Camera Models	139
5.5.3	3D Movie and TV Based on Stereo Vision	140
5.5.4	The Vergence–Accommodation Conflict	141
5.5.5	Autostereoscopic (Glasses-Free) Display Devices	142
5.5.6	Disparity Manipulation in 3D Content Creation	143

5.6	Video Quality Assessment (VQA)	145
5.6.1	Objective Assessment	145
5.6.2	Subjective Assessment	146
5.6.3	Other VQA Metrics	146
5.7	Exercises	146
	References	148
6	Basics of Digital Audio	151
6.1	Digitization of Sound	151
6.1.1	What Is Sound?	151
6.1.2	Digitization	152
6.1.3	Nyquist Theorem	155
6.1.4	Signal-to-Noise Ratio (SNR)	156
6.1.5	Signal-to-Quantization-Noise Ratio (SQNR)	158
6.1.6	Linear and Nonlinear Quantization	159
6.1.7	Audio Filtering	164
6.1.8	Audio Quality versus Data Rate	164
6.1.9	Synthetic Sounds	165
6.2	MIDI: Musical Instrument Digital Interface	167
6.2.1	MIDI Overview	167
6.2.2	Hardware Aspects of MIDI	171
6.2.3	Structure of MIDI Messages	173
6.2.4	MIDI-to-WAV Conversion	177
6.2.5	General MIDI	177
6.2.6	MIDI 2.0	177
6.3	Quantization and Transmission of Audio	178
6.3.1	Coding of Audio	178
6.3.2	Pulse Code Modulation	178
6.3.3	Differential Coding of Audio	180
6.3.4	Lossless Predictive Coding	182
6.3.5	DPCM	186
6.3.6	DM	189
6.3.7	ADPCM	190
6.4	Exercises	191
	References	195

Part II Multimedia Data Compression

7	Lossless Compression Algorithms	199
7.1	Introduction	199
7.2	Basics of Information Theory	200
7.3	Run-Length Coding	203

7.4	Variable-Length Coding (VLC)	203
7.4.1	Shannon–Fano Algorithm	204
7.4.2	Huffman Coding	205
7.4.3	Adaptive Huffman Coding	210
7.5	Dictionary-Based Coding	215
7.6	Arithmetic Coding	220
7.6.1	Basic Arithmetic Coding Algorithm	221
7.6.2	Scaling and Incremental Coding	224
7.6.3	Integer Implementation	229
7.6.4	Binary Arithmetic Coding	229
7.6.5	Adaptive Arithmetic Coding	230
7.7	Lossless Image Compression	233
7.7.1	Differential Coding of Images	233
7.7.2	Lossless JPEG	234
7.8	Exercises	236
	References	239
8	Lossy Compression Algorithms	241
8.1	Introduction	241
8.2	Distortion Measures	242
8.3	The Rate-Distortion Theory	243
8.4	Quantization	244
8.4.1	Uniform Scalar Quantization	244
8.4.2	Nonuniform Scalar Quantization	247
8.4.3	Vector Quantization	249
8.5	Transform Coding	250
8.5.1	Discrete Cosine Transform (DCT)	251
8.5.2	Karhunen–Loève Transform*	266
8.6	Wavelet-Based Coding	269
8.6.1	Introduction	269
8.6.2	Continuous Wavelet Transform*	274
8.6.3	Discrete Wavelet Transform*	277
8.7	Wavelet Packets	288
8.8	Embedded Zerotree of Wavelet Coefficients	289
8.8.1	The Zerotree Data Structure	290
8.8.2	Successive Approximation Quantization	292
8.8.3	EZW Example	293
8.9	Set Partitioning in Hierarchical Trees (SPIHT)	296
8.10	Exercises	296
	References	300

9	Image Compression Standards	301
9.1	The JPEG Standard	301
9.1.1	Main Steps in JPEG Image Compression	302
9.1.2	JPEG Modes	311
9.1.3	A Glance at the JPEG Bitstream	314
9.2	The JPEG 2000 Standard	315
9.2.1	Main Steps of JPEG 2000 Image Compression*	316
9.2.2	Adapting EBCOT to JPEG 2000	325
9.2.3	Region-of-Interest Coding	325
9.2.4	Comparison of JPEG and JPEG 2000 Performance	326
9.3	The JPEG-LS Standard	327
9.3.1	Prediction	329
9.3.2	Context Determination	331
9.3.3	Residual Coding	331
9.3.4	Near-Lossless Mode	332
9.4	Bi-Level Image Compression Standards	332
9.4.1	The JBIG Standard	332
9.4.2	The JBIG2 Standard	333
9.5	Exercises	336
	References	338
10	Basic Video Compression Techniques	341
10.1	Introduction to Video Compression	341
10.2	Video Compression Based on Motion Compensation	342
10.3	Search for Motion Vectors	343
10.3.1	Sequential Search	344
10.3.2	2D Logarithmic Search	345
10.3.3	Hierarchical Search	347
10.4	H.261	349
10.4.1	Intra-Frame (I-Frame) Coding	351
10.4.2	Inter-Frame (P-Frame) Predictive Coding	351
10.4.3	Quantization in H.261	352
10.4.4	H.261 Encoder and Decoder	353
10.4.5	A Glance at the H.261 Video Bitstream Syntax	355
10.5	H.263	357
10.5.1	Motion Compensation in H.263	357
10.5.2	Optional H.263 Coding Modes	358
10.5.3	H.263+ and H.263++	360
10.6	Exercises	362
	References	364

11	MPEG Video Coding: MPEG-1, 2, 4, and 7	365
11.1	Overview	365
11.2	MPEG-1	366
11.2.1	Motion Compensation in MPEG-1	366
11.2.2	Other Major Differences from H.261	368
11.2.3	MPEG-1 Video Bitstream	371
11.3	MPEG-2	372
11.3.1	Supporting Interlaced Video	374
11.3.2	MPEG-2 Scalabilities	377
11.3.3	Other Major Differences from MPEG-1	383
11.4	MPEG-4	385
11.4.1	Overview of MPEG-4	385
11.4.2	Video Object-Based Coding in MPEG-4	388
11.4.3	Synthetic Object Coding in MPEG-4	401
11.4.4	MPEG-4 Parts, Profiles, and Levels	409
11.5	MPEG-7	410
11.5.1	Descriptor (D)	412
11.5.2	Description Scheme (DS)	414
11.5.3	Description Definition Language (DDL)	417
11.6	Exercises	417
	References	420
12	Modern Video Coding Standards: H.264, H.265, and H.266	423
12.1	Overview	423
12.2	H.264	423
12.2.1	Motion Compensation	425
12.2.2	Integer Transform	428
12.2.3	Quantization and Scaling	430
12.2.4	Examples of H.264 Integer Transform and Quantization	432
12.2.5	Intra-Coding	435
12.2.6	In-loop Deblocking Filtering	436
12.2.7	Entropy Coding	437
12.2.8	Context-Adaptive Variable Length Coding (CAVLC)	440
12.2.9	Context-Adaptive Binary Arithmetic Coding (CABAC)	442
12.2.10	H.264 Profiles	444
12.2.11	H.264 Scalable Video Coding (SVC)	446
12.2.12	H.264 Multiview Video Coding (MVC)	447
12.3	H.265	448
12.3.1	Motion Compensation	449
12.3.2	Integer Transform	453

12.3.3	Quantization and Scaling	454
12.3.4	Intra-Coding	455
12.3.5	Discrete Sine Transform (DST)	455
12.3.6	In-Loop Filtering	457
12.3.7	Entropy Coding	458
12.3.8	Special Coding Modes	458
12.3.9	H.265 Profiles	459
12.3.10	H.265 Extensions	460
12.4	H.266	462
12.4.1	Motion Compensation	463
12.4.2	Adaptive Multiple Transforms	465
12.4.3	Non-separable Secondary Transform	465
12.4.4	In-Loop Filtering	468
12.4.5	Tools for High Dynamic Range (HDR) Video	469
12.4.6	Tools for 360° Video	471
12.4.7	H.266 Performance Report	473
12.5	Exercises	474
	References	477
13	Basic Audio Compression Techniques	479
13.1	ADPCM in Speech Coding	480
13.1.1	ADPCM	480
13.1.2	G.726 ADPCM, G.727-9	480
13.2	Vocoders	483
13.2.1	Phase Insensitivity	483
13.2.2	Channel Vocoder	484
13.2.3	Formant Vocoder	485
13.2.4	Linear Predictive Coding	486
13.2.5	CELP	489
13.2.6	Hybrid Excitation Vocoders*	495
13.3	Open Source Speech Codecs*	498
13.3.1	Speex	499
13.3.2	Opus	500
13.4	Exercises	502
	References	504
14	MPEG Audio Compression	505
14.1	Psychoacoustics	506
14.1.1	Equal-Loudness Relations	506
14.1.2	Frequency Masking	508
14.1.3	Temporal Masking	513
14.2	MPEG Audio	515
14.2.1	MPEG Layers	515
14.2.2	MPEG Audio Strategy	516

14.2.3	MPEG Audio Compression Algorithm	517
14.2.4	MPEG-2 AAC (Advanced Audio Coding)	523
14.2.5	MPEG-4 Audio	525
14.3	Other Audio Codecs	526
14.3.1	Ogg Vorbis	526
14.4	MPEG-7 Audio and Beyond	528
14.5	Further Exploration	529
14.6	Exercises	529
	References	530

Part III Multimedia Communications and Networking

15	Network Services and Protocols for Multimedia Communications	535
15.1	Protocol Layers of Computer Communication Networks	535
15.2	Local Area Network (LAN) and Access Networks	536
15.2.1	LAN Standards	538
15.2.2	Ethernet Technology	538
15.2.3	Access Network Technologies	540
15.3	Internet Technologies and Protocols	543
15.3.1	Network Layer: IP	543
15.3.2	Transport Layer: TCP and UDP	545
15.3.3	Network Address Translation (NAT) and Firewall	550
15.4	Multicast Extension	552
15.4.1	Router-Based Architectures: IP Multicast	552
15.4.2	Non Router-Based Multicast Architectures	554
15.5	Quality of Service (QoS) and Quality of Experience (QoE)	555
15.5.1	QoS and QoE for Multimedia Communications	556
15.5.2	Internet QoS Architecture: IntServ and DiffServ	560
15.5.3	Network Softwarization and Virtualization: SDN and NVF	564
15.5.4	Rate Control and Buffer Management	565
15.6	Protocols for Multimedia Transmission and Interaction	567
15.6.1	HyperText Transfer Protocol (HTTP)	567
15.6.2	Real-Time Transport Protocol (RTP)	569
15.6.3	RTP Control Protocol (RTCP)	571
15.6.4	Real-Time Streaming Protocol (RTSP)	571
15.7	Case Study: Internet Telephony	573
15.7.1	Signaling Protocols: H.323 and Session Initiation Protocol (SIP)	574
15.8	Further Exploration	578

15.9	Exercises	578
	References	580
16	Internet Multimedia Content Distribution	583
16.1	Proxy Caching	584
16.1.1	Sliding-Interval Caching	585
16.1.2	Prefix Caching and Segment Caching	587
16.1.3	Rate-Split Caching and Work-Ahead Smoothing	588
16.2	Content Distribution Networks (CDNs)	592
16.2.1	Request Routing and Redirection	592
16.2.2	Representative: Akamai Streaming CDN	594
16.3	Broadcast/Multicast Video Distribution	596
16.3.1	Smart TV and Set-Top Box (STB)	596
16.3.2	Scalable Broadcast/Multicast VoD	598
16.3.3	Multi-rate Broadcast/Multicast for Heterogeneous Users	602
16.4	Application-Layer Multicast and Peer-to-Peer Streaming	605
16.4.1	Application-Layer Multicast Tree	605
16.4.2	Representative: End-System Multicast (ESM)	606
16.4.3	Peer-to-Peer Mesh Overlay	608
16.4.4	Representative: CoolStreaming	610
16.5	Web-Based Media Streaming	613
16.5.1	Dynamic Adaptive Streaming over HTTP (DASH)	615
16.5.2	Common Media Application Format (CMAF)	617
16.5.3	Web Real-Time Communication (WebRTC)	619
16.6	Exercises	621
	References	624
17	Multimedia Over Wireless and Mobile Networks	627
17.1	Characteristics of Wireless Channels	627
17.1.1	Path Loss	627
17.1.2	Multipath Fading	628
17.2	Wireless Networking Technologies	630
17.2.1	Cellular Wireless Mobile Networks: 1G–5G	631
17.2.2	Wireless Local Area Networks (WLANs)	640
17.2.3	Bluetooth and Short-Range Technologies	643
17.3	Multimedia Over Wireless Channels	644
17.3.1	Error Detection	645
17.3.2	Error Correction	648
17.3.3	Error-Resilient Coding	652
17.3.4	Error Concealment	657

17.4	Mobility Management	660
17.4.1	Network Layer Mobile IP	661
17.4.2	Link-Layer Handoff Management	663
17.5	Further Exploration	665
17.6	Exercises	666
	References	667
18	Cloud Computing for Multimedia Services	671
18.1	Cloud Computing Overview	672
18.1.1	Representative Storage Service: Amazon S3	676
18.1.2	Representative Computation Service: Amazon EC2	678
18.2	Multimedia Cloud Computing	681
18.3	Multimedia Content Sharing over Cloud	681
18.3.1	Impact of Globalization	684
18.3.2	Case Study: Netflix	685
18.4	Multimedia Computation Offloading	687
18.4.1	Requirements for Computation Offloading	688
18.4.2	Service Partitioning for Video Processing	689
18.5	Interactive Cloud Gaming	690
18.5.1	Workload and Delay in Cloud Gaming	691
18.5.2	Implementation and Deployment	694
18.6	Edge Computing and Serverless Computing for Multimedia	696
18.6.1	Mobile Edge Computing	696
18.6.2	Serverless Computing for Video Processing	698
18.7	Further Exploration	700
18.8	Exercises	700
	References	702

Part IV Human-Centric Interactive Multimedia

19	Online Social Media Sharing	707
19.1	Representatives of Social Media Services	708
19.1.1	User-Generated Content (UGC)	708
19.1.2	Online Social Networking (OSN)	709
19.2	User-Generated Media Content Sharing	710
19.2.1	YouTube Video Format and Meta-Data	710
19.2.2	Characteristics of YouTube Video	711
19.2.3	Small-World in YouTube Videos	714
19.2.4	YouTube from a Partner’s View	716
19.2.5	Crowdsourced Interactive Livecast	719
19.3	Media Propagation in Online Social Networks	722

19.3.1	Sharing Patterns of Individual Users	723
19.3.2	Video Propagation Structure and Model	724
19.3.3	Video Watching and Sharing Behaviors	727
19.4	Mobile Video Clip Sharing	727
19.4.1	Mobile Interface Characteristics	728
19.4.2	Video Clip Popularity	729
19.4.3	Lifespan and Propagation	729
19.5	Further Exploration	731
19.6	Exercises	731
	References	733
20	Augmented Reality and Virtual Reality	737
20.1	Defining Augmented Reality and Virtual Reality	737
20.2	Workflow of Augmented Reality	740
20.2.1	Sensory Data Collection	740
20.2.2	Localization and Alignment	741
20.2.3	World Generation and Emission	741
20.3	Early Foundational Systems and Applications	742
20.4	Enabling Hardware and Infrastructure	745
20.4.1	Graphics Processing Unit (GPU)	745
20.4.2	Global Positioning System (GPS)	747
20.4.3	Networking for Multiple Users	748
20.5	Modern Augmented Reality Systems and Applications	749
20.6	Limitations and Challenges	752
20.6.1	Color Perception	753
20.6.2	Depth Perception	754
20.6.3	Localization	755
20.6.4	Information Presentation	756
20.6.5	Social Acceptance	756
20.7	Further Exploration	757
20.8	Exercises	757
	References	758
21	Content-Based Retrieval in Digital Libraries	763
21.1	How Should We Retrieve Images?	763
21.2	Synopsis of Early CBIR Systems	766
21.3	C-BIRD—An Early Experiment	768
21.3.1	Color Histogram	768
21.3.2	Color Density and Color Layout	770
21.3.3	Texture Layout	770
21.3.4	Search by Illumination Invariance	773
21.3.5	Search-by-Object Model	774
21.4	Quantifying Search Results	777
21.5	Key Technologies in Current CBIR Systems	780

21.5.1	Robust Image Features and Their Representation	780
21.5.2	User Feedback and Collaboration	782
21.5.3	Other Post-processing Techniques	783
21.5.4	Visual Concept Search	784
21.5.5	Feature Learning with Convolutional Neural Networks	785
21.5.6	Database Indexing	788
21.6	Querying on Videos	789
21.7	Querying on Videos Based on Human Activity—A Case Study	792
21.7.1	Modeling Human Activity Structures	793
21.7.2	Experimental Results	795
21.8	Quality-Aware Mobile Visual Search	795
21.8.1	Quality-Aware Method	798
21.8.2	Experimental Results	799
21.9	Deep Incremental Hashing Network*	800
21.9.1	Problem Definition	801
21.9.2	Descriptions of DIHN	801
21.9.3	Experimental Results	804
21.10	Exercises	804
	References	806
Index	811