

Digital Signal Compression

With clear and easy-to-understand explanations, this book covers the fundamental concepts and coding methods of signal compression, while still retaining technical depth and rigor. It contains a wealth of illustrations, step-by-step descriptions of algorithms, examples, and practice problems, which make it an ideal textbook for senior undergraduate and graduate students, as well as a useful self-study tool for researchers and professionals.

Principles of lossless compression are covered, as are various entropy coding techniques, including Huffman coding, arithmetic coding, run-length coding, and Lempel–Ziv coding. Scalar and vector quantization, and their use in various lossy compression systems, are thoroughly explained, and a full chapter is devoted to mathematical transformations, including the Karhunen–Loeve transform, discrete cosine transform (DCT), and wavelet transforms. Rate control is covered in detail, with several supporting algorithms to show how to achieve it. State-of-the-art transform and subband/wavelet image and video coding systems are explained, including JPEG2000, SPIHT, SBHP, EZBC, and H.264/AVC. Also, unique to this book is a chapter on set partition coding that sheds new light on SPIHT, SPECK, EZW, and related methods.

William A. Pearlman is a Professor Emeritus in the Electrical, Computer, and Systems Engineering Department at the Rensselaer Polytechnic Institute (RPI), where he has been a faculty member since 1979. He has more than 35 years of experience in teaching and researching in the fields of information theory, data compression, digital signal processing, and digital communications theory, and he has written about 250 published works in these fields. He is a Fellow of the IEEE and the SPIE, and he is the co-inventor of two celebrated image compression algorithms: SPIHT and SPECK.

Amir Said is currently a Master Researcher at Hewlett-Packard Laboratories, where he has worked since 1998. His research interests include multimedia communications, coding and information theory, image and video compression, signal processing, and optimization, and he has more than 100 publications and 20 patents in these fields. He is co-inventor with Dr. Pearlman of the SPIHT image compression algorithm and co-recipient, also with Dr. Pearlman, of two Best Paper Awards, one from the IEEE Circuits and Systems Society and the other from the IEEE Signal Processing Society.

Digital Signal Compression

Principles and Practice

WILLIAM A. PEARLMAN

Rensselaer Polytechnic Institute, New York

AMIR SAID

Hewlett-Packard Laboratories, Palo Alto, California



CAMBRIDGE
UNIVERSITY PRESS

Cambridge University Press & Assessment
 978-0-521-89982-6 — Digital Signal Compression
 William A. Pearlman, Amir Said
 Frontmatter
[More Information](#)



Shaftesbury Road, Cambridge CB2 8EA, United Kingdom
 One Liberty Plaza, 20th Floor, New York, NY 10006, USA
 477 Williamstown Road, Port Melbourne, VIC 3207, Australia
 314–321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre, New Delhi – 110025, India
 103 Penang Road, #05–06/07, Visioncrest Commercial, Singapore 238467

Cambridge University Press is part of Cambridge University Press & Assessment, a department of the University of Cambridge.

We share the University's mission to contribute to society through the pursuit of education, learning and research at the highest international levels of excellence.

www.cambridge.org

Information on this title: www.cambridge.org/9780521899826

© Cambridge University Press & Assessment 2011

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 & Assessment.

First published 2011

A catalogue record for this publication is available from the British Library

Library of Congress Cataloging-in-Publication data

Pearlman, William A. (William Abraham)

Digital signal compression : principles and practice / William A. Pearlman, Amir Said.
 p. cm.

Includes bibliographical references and index.

ISBN 978-0-521-89982-6 (hardback)

1. Data compression (Telecommunication) 2. Signal processing – Digital techniques.

I. Said, Amir. II. Title.

TK5102.92.P43 2011

005.74'6–dc23

2011012972

ISBN 978-0-521-89982-6 Hardback

Additional resources for this publication at www.cambridge.org/9780521899826

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

Cambridge University Press & Assessment
978-0-521-89982-6 — Digital Signal Compression
William A. Pearlman, Amir Said
Frontmatter
[More Information](#)

To Eleanor
To Celli and Ricardo

Contents

	<i>Preface</i>	<i>page</i> xv
	<i>Acknowledgments</i>	xix
1	Motivation	1
	1.1 The importance of compression	1
	1.2 Data types	2
	1.2.1 Symbolic information	2
	1.2.2 Numerical information	3
	1.3 Basic compression process	4
	1.4 Compression applications	5
	1.5 Design of compression methods	6
	1.6 Multi-disciplinary aspect	8
	Note	8
	References	9
2	Book overview	10
	2.1 Entropy and lossless coding	10
	2.2 Quantization	11
	2.3 Source transformations	12
	2.3.1 Prediction	12
	2.3.2 Transforms	13
	2.4 Set partition coding	16
	2.5 Coding systems	17
	2.5.1 Performance criteria	17
	2.5.2 Transform coding systems	18
	2.5.3 Subband coding systems	19
	2.6 Distributed source coding	20
	Notes	21
	References	22
3	Principles of lossless compression	23
	3.1 Introduction	23
	3.2 Lossless source coding and entropy	23

3.3	Variable length codes	28
3.3.1	Unique decodability and prefix-free codes	28
3.3.2	Construction of prefix-free codes	28
3.3.3	Kraft inequality	30
3.4	Optimality of prefix-free codes	32
3.4.1	Sources with memory	36
3.5	Concluding remarks	37
	Problems	37
	References	40
4	Entropy coding techniques	41
4.1	Introduction	41
4.2	Huffman codes	41
4.3	Shannon–Fano–Elias codes	47
4.3.1	SFE code examples	48
4.3.2	Decoding the SFE code	49
4.4	Arithmetic code	50
4.4.1	Preliminaries	50
4.4.2	Arithmetic encoding	51
4.4.3	Arithmetic decoding	53
4.5	Run-length codes	55
4.6	Alphabet partitioning: modified Huffman codes	57
4.6.1	Modified Huffman codes	57
4.6.2	Alphabet partitioning	58
4.7	Golomb code	60
4.8	Dictionary coding	63
4.8.1	The LZ78 code	64
4.8.2	The LZW algorithm	65
4.8.3	The LZ77 coding method	67
4.9	Summary remarks	72
	Problems	72
	Notes	75
	References	76
5	Lossy compression of scalar sources	77
5.1	Introduction	77
5.2	Quantization	77
5.2.1	Scalar quantization	77
5.2.2	Uniform quantization	81
5.3	Non-uniform quantization	87
5.3.1	High rate approximations	89
5.4	Companding	91
5.4.1	Distortion at high rates	93

Contents

ix

5.5	Entropy coding of quantizer outputs	95
5.5.1	Entropy coded quantizer characteristics	98
5.5.2	Null-zone quantization	99
5.6	Bounds on optimal performance	101
5.6.1	Rate-distortion theory	102
5.6.2	The Gish–Pierce bound	104
5.7	Concluding remarks	107
5.8	Appendix: quantization tables	107
	Problems	109
	Note	113
	References	114
6	Coding of sources with memory	116
6.1	Introduction	116
6.2	Predictive coding	116
6.2.1	Optimal linear prediction	117
6.2.2	DPCM system description	120
6.2.3	DPCM coding error and gain	121
6.3	Vector coding	122
6.3.1	Optimal performance bounds	122
6.3.2	Vector (block) quantization (VQ)	129
6.3.3	Entropy constrained vector quantization	135
6.4	Tree-structured vector quantization	141
6.4.1	Variable length TSVQ coding	144
6.4.2	Pruned TSVQ	145
6.5	Tree and trellis codes	146
6.5.1	Trellis codes	148
6.5.2	Encoding and decoding of trellis codes	150
6.5.3	Codevector alphabets	152
6.6	Trellis coded quantization (TCQ)	152
6.6.1	Entropy-coded TCQ	154
6.6.2	Improving low-rate performance in TCQ	155
6.7	Search algorithms	155
6.7.1	M-algorithm	155
6.7.2	The Viterbi algorithm	158
6.8	Concluding remarks	160
	Problems	160
	Notes	163
	References	164
7	Mathematical transformations	166
7.1	Introduction	166
7.1.1	Transform coding gain	169

7.2	The optimal Karhunen–Loeve transform	171
7.2.1	Optimal transform coding gain	172
7.3	Suboptimal transforms	172
7.3.1	The discrete Fourier transform	172
7.3.2	The discrete cosine transform	173
7.3.3	The Hadamard–Walsh transform	174
7.4	Lapped orthogonal transform	175
7.4.1	Example of calculation of transform coding gain	178
7.5	Transforms via filter banks	179
7.6	Two-dimensional transforms for images	181
7.7	Subband transforms	184
7.7.1	Introduction	184
7.7.2	Coding gain of subband transformation	187
7.7.3	Realizable perfect reconstruction filters	192
7.7.4	Orthogonal wavelet transform	194
7.7.5	Biorthogonal wavelet transform	199
7.7.6	Useful biorthogonal filters	204
7.7.7	The lifting scheme	205
7.7.8	Transforms with integer output	208
7.8	Concluding remarks	211
	Problems	212
	Notes	214
	References	216
8	Rate control in transform coding systems	218
8.1	Rate allocation	218
8.1.1	Optimal rate allocation for known quantizer characteristics	220
8.1.2	Realizing the optimal rate allocation	223
8.1.3	Fixed level quantization	225
8.1.4	Optimal bit allocation for arbitrary set of quantizers	226
8.1.5	Building up to optimal rates for arbitrary quantizers	228
8.1.6	Transform coding gain	230
8.2	Subband rate allocation	233
8.2.1	Practical issues	237
8.2.2	Subband coding gain	239
8.3	Algorithms for rate allocation to subbands	241
8.4	Conclusions	242
	Problems	242
	Notes	243
	References	244
9	Transform coding systems	245
9.1	Introduction	245

Contents

xi

9.2	Application of source transformations	245
9.2.1	Model-based image transform coding	246
9.2.2	Encoding transform coefficients	249
9.3	The JPEG standard	251
9.3.1	The JPEG baseline system	252
9.3.2	Detailed example of JPEG standard method	256
9.4	Advanced image transform coding: H.264/AVC intra coding	259
9.5	Concluding remarks	262
	Problems	262
	Notes	263
	References	264
10	Set partition coding	265
10.1	Principles	265
10.1.1	Partitioning data according to value	267
10.1.2	Forming partitions recursively: square blocks	270
10.1.3	Binary splitting	274
10.1.4	One-dimensional signals	276
10.2	Tree-structured sets	276
10.2.1	A different wavelet transform partition	279
10.2.2	Data-dependent thresholds	282
10.2.3	Adaptive partitions	283
10.3	Progressive transmission and bitplane coding	285
10.4	Applications to image transform coding	286
10.4.1	Block partition coding and amplitude and group partitioning (AGP)	287
10.4.2	Enhancements via entropy coding	289
10.4.3	Traversing the blocks	289
10.4.4	Embedded block coding of image wavelet transforms	291
10.4.5	A SPECK coding example	291
10.4.6	Embedded tree-based image wavelet transform coding	297
10.4.7	A SPIHT coding example	299
10.4.8	Embedded zerotree wavelet (EZW) coding	302
10.4.9	Group testing for image wavelet coding	306
10.5	Conclusion	306
	Problems	307
	Notes	310
	References	311
11	Subband/wavelet coding systems	313
11.1	Wavelet transform coding systems	313
11.2	Generic wavelet-based coding systems	317
11.3	Compression methods in wavelet-based systems	318

11.4	Block-based wavelet transform set partition coding	320
11.4.1	Progressive resolution coding	321
11.4.2	Quality-progressive coding	323
11.4.3	Octave band partitioning	326
11.4.4	Direct bit-embedded coding methods	328
11.4.5	Lossless coding of quantizer levels with adaptive thresholds	329
11.4.6	Tree-block coding	331
11.4.7	Coding of subband subblocks	332
11.4.8	Coding the initial thresholds	333
11.4.9	The SBHP method	335
11.4.10	JPEG2000 coding	336
11.4.11	The embedded zero-block coder (EZBC)	343
11.5	Tree-based wavelet transform coding systems	347
11.5.1	Fully scalable SPIHT	347
11.5.2	Resolution scalable SPIHT	349
11.5.3	Block-oriented SPIHT coding	352
11.6	Rate control for embedded block coders	354
11.7	Conclusion	356
	Notes	357
	References	359
12	Methods for lossless compression of images	361
12.1	Introduction	361
12.2	Lossless predictive coding	362
12.2.1	Old JPEG standard for lossless image compression	362
12.3	State-of-the-art lossless image coding and JPEG-LS	364
12.3.1	The predictor	364
12.3.2	The context	365
12.3.3	Golomb–Rice coding	366
12.3.4	Bias cancellation	366
12.3.5	Run mode	367
12.3.6	Near-lossless mode	368
12.3.7	Remarks	368
12.4	Multi-resolution methods	368
12.5	Concluding remarks	369
	Problems	370
	Notes	371
	References	372
13	Color and multi-component image and video coding	373
13.1	Introduction	373

Contents	xiii
13.2 Color image representation	374
13.2.1 Chrominance subsampling	376
13.2.2 Principal component space	377
13.3 Color image coding	378
13.3.1 Transform coding and JPEG	378
13.3.2 Wavelet transform systems	380
13.4 Multi-component image coding	383
13.4.1 JPEG2000	383
13.4.2 Three-dimensional wavelet transform coding	384
13.4.3 Video coding	389
13.5 Concluding remarks	395
Notes	395
References	396
14 Distributed source coding	398
14.1 Slepian–Wolf coding for lossless compression	398
14.1.1 Practical Slepian–Wolf coding	400
14.2 Wyner–Ziv coding for lossy compression	404
14.2.1 Scalar Wyner–Ziv coding	406
14.2.2 Probability of successful reconstruction	407
14.3 Concluding remarks	411
Problems	411
Notes	412
References	413
<i>Index</i>	414

Preface

This book is an outgrowth of a graduate level course taught for several years at Rensselaer Polytechnic Institute (RPI). When the course started in the early 1990s, there were only two textbooks available that taught signal compression, Jayant and Noll¹ and Gersho and Gray.² Certainly these are excellent textbooks and valuable references, but they did not teach some material considered to be necessary at that time, so the textbooks were supplemented with handwritten notes where needed. Eventually, these notes grew to many pages, as the reliance on published textbooks diminished. The lecture notes remained the primary source even after the publication of the excellent book by Sayood,³ which served as a supplement and a source of some problems. While the Sayood book was up to date, well written, and authoritative, it was written to be accessible to undergraduate students, so lacked the depth suitable for graduate students wanting to do research or practice in the field. The book at hand teaches the fundamental ideas of signal compression at a level that both graduate students and advanced undergraduate students can approach with confidence and understanding. The book is also intended to be a useful resource to the practicing engineer or computer scientist in the field. For that purpose and also to aid understanding, the 40 algorithms listed under *Algorithms* in the Index are not only fully explained in the text, but also are set out step-by-step in special algorithm format environments.

This book contains far more material than can be taught in a course of one semester. As it was being written, certain subjects came to light that begged for embellishment and others arose that were needed to keep pace with developments in the field. One example of this additional material, which does not appear in any other textbook, is Chapter 14 on “Distributed source coding,” a subject which has received considerable attention lately. The intent was to present the fundamental ideas there, so that the student can understand and put into practice the various methods being proposed that use this technique.

The two longest chapters in the book are Chapters 10 and 11, entitled “Set partition coding” and “Subband/wavelet coding systems,” respectively. They were actually the first chapters written and were published previously as a monograph in two parts.⁴ The versions appearing here are updated with some minor errors corrected. Being the inventors of SPIHT and proponents and pioneers of set partition coding, we felt that its fundamental principles were not expounded in the technical literature. Considering the great interest in SPIHT, as evidenced by the thousands of inquiries received by us over the years since its origin in 1995 (at this writing 94,200 hits on Google), we

were eager to publish a true tutorial on the fundamental concepts of this algorithm. We believe that Chapter 10 fulfills this intent. Other books usually present only the SPIHT algorithm, almost always by working an example without revealing the underlying principles. Chapter 11 describes more image wavelet coding systems than any other book, including the JPEG2000 standard, fully scalable SPIHT, SBHP, and EZBC. The last three are set partition coders, while JPEG2000 contains auxiliary algorithms that use set partitioning. Furthermore, this chapter explains how to embed features of scalability and random access in coded bitstreams.

Besides distributed source coding, some preliminary material are also firsts in this book. They are: analysis of null-zone quantization, rate allocation algorithms, and the link between filters and wavelets. The aforementioned link is explained in Chapter 7 on “Mathematical transformations,” in a way that requires only some knowledge of discrete-time Fourier transforms and linear system analysis. The treatment avoids the concepts of functional analysis and the use of polyphase transforms with little compromise of rigor. The intent was to make the book accessible to advanced undergraduates, who would likely not have exposure to these subjects. Also to serve this purpose, prior exposure to information theory is not a prerequisite, as the book teaches the relevant aspects needed to grasp the essential ideas.

One criticism that might be levied at this book is its emphasis on compression of images. Certainly, that reflects the main field of research and practice of the authors. However, image compression is possible only by understanding and applying the principles of compression that pertain to all source data. In fact, the material of the first eight chapters is generic and dimension independent. The notation is one-dimensional for the most part, and the generalization to higher dimensions is mostly obvious and hence unnecessary. Although the applications are mostly to images in the remainder of the book, except for the generic Chapter 14, the corresponding one-dimensional signal methods are mentioned when applicable and even included in some problems. The standards and state-of-the-art systems of image compression are treated in some detail, as they represent the most straightforward application of basic principles. The standard speech and audio coding systems require additional complications of perceptual and excitation models, and echo cancellation. Their inclusion would make the book too long and cumbersome and not add much to its primary objective. Nevertheless, the material on image compression systems in Chapters 9, 11, and 12 is comprehensive enough to meet the secondary objective of serving as a good tutorial and reference for image compression researchers and practitioners.

Chapter 12 treats the subject of lossless image compression. Lossless image compression is used only for archiving of image data. There seems to be no call for lossless compression of any sources for the purpose of multimedia communication, as the data transfer would be too slow or require too much bandwidth or both. For example, there is no compression of audio or speech in the WAV storage format for compact discs. MP3 is a compressed format for audio and speech transmission and recording; the compressed format of JPEG is standard in every digital camera for consumer use; all digital video is compressed by MPEG or ITU standard formats. Images seem to be the only sources

that are subject to lossless compression. The standard methods described in Chapter 12 serve as good examples of how basic principles are put into practice.

The book did not seem complete without a chapter on how compression is applied to three-dimensional sources, such as color images, volume medical images, and video. At the time of writing, there are no other textbooks that teach three-dimensional wavelet coding methods. Therefore, we wrote Chapter 13 with the intent to show the reader how the methods of the earlier chapters are extended to these higher dimensional sources. We purposely omitted detailed descriptions of video standards. We just explained the general framework in these systems upon which compression is applied and a little about the compression methods, which are mostly covered in detail in earlier chapters.

We urge potential buyers or browsers to read Chapters 1 and 2, which discuss the motivation to learn signal compression and take a brief tour through the book. This book turned out to be different in many respects from what was taught in the RPI course. Roughly, the coverage of that course was all of Chapters 3, 4, 5, and 6, which was deemed essential material. One can be selective in Chapter 7, for example, skipping the lapped orthogonal transform and some parts of Section 7.7, “Subband transforms,” especially the detailed development of the connection between wavelet theory and FIR filters. Likewise, in Chapter 8, some of the rate allocation algorithms may be skipped, as well as the detailed derivations of optimal rate allocation and coding gain. One can skim through Section 9.3 in the next chapter, and skip Section 9.4 on H.264/AVC intra coding, which did not exist when the course was last taught. Time may not allow anything in Chapter 10, other than set partitioning for SPIHT and the accompanying coding example, and in Chapter 11 only a sketch of the JPEG2000 coding system. Lossless image compression in Chapter 12 could be covered earlier in the course, perhaps after Chapter 4, “Entropy coding techniques.” Certainly, there is enough material to accommodate different choices of emphasis and objective.

For students with a background in information theory and signal processing or those more interested in computer science or actual implementation, an instructor may skip some of the preliminary material in the early chapters and teach all of the rate allocation algorithms in Chapter 8 and cover Chapters 10 and 11 in their entirety. Chapter 11 contains much practical material on implementation of coding systems. In fact, we think that approach would be appropriate for a short course.

The book contains many figures and problems. The problems in many cases have to be worked using a calculation and plotting program, such as MATLAB, and sometimes by making a computer program. Some datasets and basic software in C or C++ and MATLAB m-files, will be made freely available on the course website: <http://www.cambridge.org/pearlman>. Also freely available on the website are Powerpoint animations of the SPIHT and SPECK algorithms. Figures and problem solutions will be made available to instructors.

Notes

1. N. S. Jayant and P. Noll, *Digital Coding of Waveforms*, Prentice-Hall 1984.
2. A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*, Kluwer Academic Publishers 1992.
3. K. Sayood, *Introduction to Data Compression*, Morgan Kaufmann Publishers 1996, 2000, 2006.
4. Look under **Books** in <http://www.cipr.rpi.edu/~pearlman/>.

Acknowledgments

I wish to acknowledge my debt and gratitude over the years to all my students from whom I have learned so much. Without the experience of interacting with them, I could not have written this book. There is not enough space to list all the contributors and their contributions, as I would like. Instead, I shall mention only those who were directly involved in the content of this book. I wish to thank Sungdae Cho, who created most of the video and three-dimensional SPIHT software and who encouraged me to write this book. He also created the SPIHT animation that is available on the book's website. I am grateful to Ying Liu, who developed the higher dimensional SBHP algorithms and graciously helped me with my programming difficulties. My gratitude goes out to Asad Islam, who developed the SPECK algorithm, and Xiaoli Tang, who extended it to higher dimensions. I also wish to thank Matthias Narroschke, who created the SPECK animation while I was visiting the University of Hannover, and Emmanuel Christophe, who developed resolution-scalable SPIHT, while he was a visiting student to RPI from the University of Toulouse and TeSA (Telecom Spatiales Aeronautiques). I also wish to thank Professor James Fowler of Mississippi State University for checking my explanation of bisection in two-dimensional block partitioning. I am grateful to Alessandro Dutra, who helped me write the solutions manual for the book's problems.

A special acknowledgment goes to Amir Said, my co-author and long-time collaborator, who, while a doctoral student at RPI working under Professor John Anderson, came to me to do a special project on image compression. From this project, by virtue of Amir's intellectual brilliance, creativity, and prodigious programming skills, the SPIHT algorithm was born. There are no words that adequately express my thanks to him.

Lastly, I wish to thank my wife, Eleanor, who suffered from having an often distracted and inattentive husband during more than three years while I was writing this book. This book is dedicated to her, as I could not manage without her love and support.

William A. Pearlman