

THIRD EDITION

# Introduction to Data Compression

**Khalid Sayood**

University of Nebraska

**Morgan Kaufmann Publishers**

*An Imprint of Elsevier*

S I T Library  
Valachil, Mangalore



Accn No: 012993

Srinivas Institute of Technology

Acc. No.: 12993

Call No.:

**Introduction to Data Compression, 3/e**  
**Sayood**

Morgan Kaufmann Publishers  
*An Imprint of Elsevier*  
500 Sansome Street, Suite 400,  
San Francisco, CA 94111

© 2006 by Elsevier Inc.

Original ISBN: 978-0-12-620862-7

All rights reserved.

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means—electronic, mechanical, photocopying, recording, or otherwise—without the prior written permission of the publisher.

First Printed in India 2006  
Reprinted 2007, 2008 (Twice)  
Reprint, 2009

Indian Reprint ISBN: 978-81-312-0624-9

This edition is for sale in Bangladesh, Bhutan, India, Maldives, Nepal, Pakistan, Sri Lanka and designated countries in South-East Asia through Elsevier (Singapore) Pte. Ltd. Sale and purchase of this book outside of these countries is unauthorised by the publisher.

Published by Elsevier, a division of Reed Elsevier India Private Limited,  
Logix Park, 1st Floor, A-4 & 5, Sector-16, Noida-201301, UP (India)

Printed and bound in India by Sheel Print-N-Pack  
D-132, Hosiery Complex, Phase II Extn., NOIDA 201 305.

# Contents

<b>Preface</b>	<b>xvii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Compression Techniques	3
1.1.1 Lossless Compression	4
1.1.2 Lossy Compression	5
1.1.3 Measures of Performance	5
1.2 Modeling and Coding	6
1.3 Summary	10
1.4 Projects and Problems	11
<b>2 Mathematical Preliminaries for Lossless Compression</b>	<b>13</b>
2.1 Overview	13
2.2 A Brief Introduction to Information Theory	13
2.2.1 Derivation of Average Information ★	18
2.3 Models	23
2.3.1 Physical Models	23
2.3.2 Probability Models	23
2.3.3 Markov Models	24
2.3.4 Composite Source Model	27
2.4 Coding	27
2.4.1 Uniquely Decodable Codes	28
2.4.2 Prefix Codes	31
2.4.3 The Kraft-McMillan Inequality ★	32
2.5 Algorithmic Information Theory	35
2.6 Minimum Description Length Principle	36
2.7 Summary	37
2.8 Projects and Problems	38
<b>3 Huffman Coding</b>	<b>41</b>
3.1 Overview	41
3.2 The Huffman Coding Algorithm	41
3.2.1 Minimum Variance Huffman Codes	46
3.2.2 Optimality of Huffman Codes ★	48
3.2.3 Length of Huffman Codes ★	49
3.2.4 Extended Huffman Codes ★	51

3.3	Nonbinary Huffman Codes ★	55
3.4	Adaptive Huffman Coding	58
	3.4.1 Update Procedure	59
	3.4.2 Encoding Procedure	62
	3.4.3 Decoding Procedure	63
3.5	Golomb Codes	65
3.6	Rice Codes	67
	3.6.1 CCSDS Recommendation for Lossless Compression	67
3.7	Tunstall Codes	69
3.8	Applications of Huffman Coding	72
	3.8.1 Lossless Image Compression	72
	3.8.2 Text Compression	74
	3.8.3 Audio Compression	75
3.9	Summary	77
3.10	Projects and Problems	77
<b>4</b>	<b>Arithmetic Coding</b>	<b>81</b>
4.1	Overview	81
4.2	Introduction	81
4.3	Coding a Sequence	83
	4.3.1 Generating a Tag	84
	4.3.2 Deciphering the Tag	91
4.4	Generating a Binary Code	92
	4.4.1 Uniqueness and Efficiency of the Arithmetic Code	93
	4.4.2 Algorithm Implementation	96
	4.4.3 Integer Implementation	102
4.5	Comparison of Huffman and Arithmetic Coding	109
4.6	Adaptive Arithmetic Coding	112
4.7	Applications	112
4.8	Summary	113
4.9	Projects and Problems	114
<b>5</b>	<b>Dictionary Techniques</b>	<b>117</b>
5.1	Overview	117
5.2	Introduction	117
5.3	Static Dictionary	118
	5.3.1 Digram Coding	119
5.4	Adaptive Dictionary	121
	5.4.1 The LZ77 Approach	121
	5.4.2 The LZ78 Approach	125
5.5	Applications	133
	5.5.1 File Compression—UNIX <code>compress</code>	133
	5.5.2 Image Compression—The Graphics Interchange Format (GIF)	133
	5.5.3 Image Compression—Portable Network Graphics (PNG)	134
	5.5.4 Compression over Modems—V.42 bis	136