



GILBERT HELD

# DATA AND IMAGE COMPRESSION

Tools and Techniques

Fourth Edition

 WILEY

---

# DATA AND IMAGE COMPRESSION

Tools and Techniques

Fourth Edition

---

**Gilbert Held**

*4-Degree Consulting,  
Macon, Georgia, USA*

and

**Thomas R. Marshall**

*(Software Author, Second  
and Third Editions)*

**JOHN WILEY & SONS LTD**

Chichester · New York · Brisbane · Toronto · Singapore

---

# CONTENTS

---

Preface	xi
Acknowledgements	xiii
<b>1 Rationale and Utilization</b>	<b>1</b>
1.1 Logical compression	2
1.2 Physical compression	3
1.3 Compression benefits	4
1.4 Terminology	5
Compression efficiency	6
Compression methods	7
1.5 Communications applications	12
1.6 Data storage applications	16
1.7 Other applications	17
1.8 Data compression and information transfer	18
BISYNC communications	19
<b>2 Data Codes and Compression-indicating Characters</b>	<b>34</b>
2.1 Data codes	35
Boudot code	35
BCD and EBCDIC codes	35
The ASCII code	39
2.2 Selecting compression-indicating characters	39
Logical compression-indicating characters	41
Physical compression-indicating characters	43
<b>3 Character-oriented Compression Techniques</b>	<b>47</b>
3.1 Null suppression	48
Technique overview	48
Programming examples	50
File naming conventions	50
Compression program	52
Decompression program	53
Technique variations	59

3.2 Bit mapping	60
Encoding process	60
Hardware considerations	61
Suppression efficiency	61
Bitmap variations	65
Technique constraints	66
3.3 Run length	68
Operation	68
Encoding process	69
Decoding	71
Utilization	72
Programming examples	77
Decompression	83
3.4 Half-byte packing	90
Encoding format and technique efficiency	92
Encoding process	96
Decoding	99
Programming examples	99
Encoding	100
Decompression	107
Extended half-byte	109
BASIC compression program	109
BASIC decompression program	118
Encoding variations and efficiency considerations	125
3.5 Diatomic encoding	128
Operation	128
Air frequency of occurrence	129
Communications hardware implementation	133
Programming examples	137
Decompression	143
Encoding considerations	148
3.6 Byte pair encoding	150
Expansion	156
Efficiency	157
3.7 Pattern substitution	157
The pattern table	159
Encoding process	159
Patterns in programming Languages	161
3.8 Forms-mode operation	163
Transmission	165
Modified forms-mode method	167
<b>4 Statistical Encoding</b>	<b>170</b>
4.1 Information theory	170
Entropy examples	173
4.2 Huffman coding	176
Prefix property of code	176
Code construction considerations	181
Information requirements	184
HUFF1.C	186

DHUFF.C	194
4.3 Shannon-Fano coding	195
Code ambiguity	200
Efficiency comparison	201
4.4 Comma codes	206
4.5 Arithmetic coding	208
Overview	209
Operation	209
Decoding	214
Computer models	214
4.6 Adaptive compression	217
The fixed compression table	218
Adaptive compression	222
Coded example	224
Sixpack program	229
Finite window compression	230
Circular buffer array	231
Hash table	231
Doubly linked lists	231
Adaptive Huffman coding	240
Bit packing	241
4.7 MNP compression	241
MNP Class 5	241
MNP Class 7 enhanced data compression	245
<b>5 Facsimile Compression</b>	248
5.1 Group 3 fax operational overview	248
5.2 Relative encoding	250
Telemetry compression	250
Facsimile techniques	252
5.3 Modified Huffman codes	255
5.4 Modified read coding	259
<b>6 Dictionary Based String Compression</b>	265
6.1 String compression efficiencies	266
6.2 LZ77 compression	269
Encoding	269
Decoding	270
Coding problems	272
6.3 LZSS compression	272
LZSS modifications	273
6.4 LZ78 compression	274
Operation	274
Coding tradeoffs	275
6.5 LZW compression	280
Operation	280
Decoding	282
Implementation issues	283

6.6 BTLZ (V.42 bis) compression	284
Dictionary pruning	287
Decoding algorithm	288
Configuration and negotiation parameters	288
Efficiency limitations	289
LZ variations	289
6.7 Multilevel coding and other efficiency improvements	290
LZH compression	290
The AR archive program	291
LZARI compression	292
<b>7 Image Compression</b>	<b>311</b>
7.1 Graphics overview	312
Raster programs	312
Vector programs	312
Value of image compression	313
7.2 Image compression techniques	314
Character based	314
Statistical	315
Dictionary based coding	315
Binary graphics	316
Predictive coding	316
Adaptive Bilevel Image Compression	316
JPEG	317
Object model	318
Pattern matching	318
Fractal coding	319
7.3 GIF file formats	319
GIF Signature	320
Screen Descriptor	321
Global Color Map	322
Image Descriptor	323
Local Color Map/Table	324
Raster Data	324
LZW algorithm	324
GIF89 extensions	326
Proposed GIF specification modules	327
7.4 JPEG file formats	330
7.5 Image tools	331
GIF coding	331
JPEG4.ZIP	338
GIF2JPG/JPG2GIF	341
Image Alchemy	342
PRINTGF	343
<b>8 Communications Software-linkage Considerations</b>	<b>344</b>
8.1 Compression routine placement	344
Software considerations	345

8.2 Timing considerations	348
Total time	349
Flow control	350
Routine linkage	351

## **9 Compression-performing Hardware and Software Product Overview**

9.1 Hardware products	353
Asynchronous data compressors	353
V.42 bis modems	356
Synchronous data compressors	356
Compression DSU	357
The Time Machine	358
Multifunctional compression devices	358
9.2 Software products	362
Teleprocessing monitor compression products	360
Database compression	364
Personal computer file compression	366
Archive utility programs	366

## **APPENDIX A DATANALYSIS Program Descriptions and Listings**

A.1 FORTRAN program: operational description	389
A.2 Main routine	389
A.3 Program transferability	393
A.4 Variable assignments	394
A.5 BASIC program: operational description	395
A.6 Main routine	397
A.7 Variable assignments	400
A.8 FORTRAN DATANALYSIS program listing	401
A.9 BASIC language DATANALYSIS program listing	410

## **APPENDIX B SHRINK Program Descriptions and Listings**

B.1 MERGEC.BAS and MERGED.BAS	417
B.2 Program operations	424

References	427
Further Reading	429
Index	432