

データの圧縮と効率化

圧縮と伸張

ファイルサイズを小さくするための技術 ()

圧縮したデータを元の状態に復元する技術 ()

音声、画像、動画データはデータサイズが大きくなりがち

⇒ファイル形式そのものに圧縮技術を使っているものが多い

圧縮率=データを圧縮した際に、圧縮後のデータが元のデータのどのくらいの情報量に減ったかを表す割合。「圧縮後のサイズ」の「元のサイズ」に対する割合を百分率で表すことが多い。つまり

50MB のファイルが 5MB に圧縮された場合、 $5 / 50 \dots () \%$

「削減されたサイズ」の「元のサイズ」に対する割合とする場合もある。

50MB のファイルが 5MB に圧縮された場合、 $(50-5) / 50 \dots () \%$

プロパティ	値
イメージ	
大きさ	500 x 495
幅	500 ピクセル
高さ	495 ピクセル
ビットの深さ	24
ファイル	
名前	7.bmp
項目の種類	BMP ファイル
フォルダーのパス	G:\Document\#INFOMATION#gaz
作成日時	2020/07/23 21:48
更新日時	2020/07/23 21:48
サイズ	725 KB
属性	A

プロパティ	値
イメージ	
大きさ	500 x 495
幅	500 ピクセル
高さ	495 ピクセル
ビットの深さ	8
ファイル	
名前	7.gif
項目の種類	GIF ファイル
フォルダーのパス	G:\Document\#INFOMATION#gaz
作成日時	2020/07/23 21:50
更新日時	2020/07/23 21:50
サイズ	95.3 KB
属性	A

プロパティ	値
元の場所	
撮影日時	
イメージ	
大きさ	500 x 495
幅	500 ピクセル
高さ	495 ピクセル
ビットの深さ	32
ファイル	
名前	7.png
項目の種類	PNG ファイル
フォルダーのパス	G:\Document\#INFOMATION#gaz
作成日時	2020/07/23 21:50
更新日時	2020/07/23 21:50
サイズ	432 KB
属性	A

プロパティ	値
イメージ	
イメージID	
大きさ	500 x 495
幅	500 ピクセル
高さ	495 ピクセル
水平方向の解像度	96 dpi
垂直方向の解像度	96 dpi
ビットの深さ	24
圧縮	
解像度の単位	
色の表現	
圧縮ビット/ピクセル	
ファイル	
名前	7.jpg
項目の種類	JPG ファイル
フォルダーのパス	G:\Document\#INFOMATION#gaz.com...
作成日時	2020/07/23 21:50
更新日時	2020/07/23 21:50
サイズ	60.7 KB
属性	A

7.Bmp	7.Gif	7.Jpg	7.png
W500px,H495px	W500px,H495px	W500px,H495px	W500px,H495px
24bit	8bit	24bit	32bit
725kB	95kB	61kB	432kB
圧縮率	13%	8.4%	60%

画像圧縮・ファイル形式
実験結果

音声 非圧縮 wav⇒圧縮 mp3

画像 非圧縮 bmp⇒圧縮 jpg、gif、png

動画 (非圧縮)avi⇒圧縮 mpeg、mp4、mov

メディアタイプ	圧縮型	拡張子	ファイル	ファイルサイズ	圧縮率%
音声	MIDI	.mid	c.mid	170B	—
	WAVE	.wav	c.wav	483kB	—
	MP 3	.mp3	c.mp3	44.4kB	
画像	BMP	.bmp	7.bmp	725kB	—
	GIF	.gif	7.gif	95kB	
	JPEG	.jpg	7.jpg	61kB	
	PNG	.png	7.png	432kB	
動画	AVI	.avi	sample.avi	11.6MB	—
	MP4	.mp4	sample.mp4		
	MP4	.mp4	sample_low.mp4		

可逆圧縮と非可逆圧縮

圧縮した情報を元に戻すと完全に元の情報と同じものになる ➡ ()

伸張した時に完全な形では元の情報に戻らない。高い圧縮率 ➡ ()

文書ファイルなど圧縮には、プログラムを使い圧縮ファイル形式に変換。圧縮ファイルはそのまま使えないので、利用時に伸張する必要がある。

主な圧縮ファイル形式には、() 形式や LZH 形式があるが、マルチメディアデータとは異なり、() 圧縮方式を使う必要がある。

ランレングス圧縮：

AAAAABBBBBBBBBBAAA …元 文字数 ()

A5B9A3 (5A9B3A) 文字数 ()

問題点 ABCDDDD …元 文字数 ()

A1B1C1D4 (1A1B1C4D) 文字数 ()

圧縮結果が大きくなる ➡ 同じデータの繰り返しのみ圧縮

解決策 ABCDD2 文字数 ()

例題 AAAABCCCCCDEFG …元 文字数 ()

通常の圧縮 A4B2C6D1E1F1G1 文字数 ()

解決策圧縮 AA2BB1CC3DEFG 文字数 ()

発展改善策 4A2B6C-4DEFG 文字数 ()

- が付いた (負の) 長さデータは連続しないデータの長さ

“A” が 4、“B” が 2、“C” が 6、圧縮できない “DEFG” の 4 文字がある

メディアデータの非可逆圧縮

JPEG：(画質に影響しない) データの一部を捨てる…RGB データを別の変数「輝度」「赤方向の色相」「青方向の色相」に置き換えるなど難解なアルゴリズム

MP3：音の聞こえ易さの違い(周波数ごとの最小可聴値)や大きな音が鳴った際に、その直前直後や近い周波数の小さな音が聞こえにくくなる現象等の人間の聴覚心理を利用した圧縮を行う

MP4：現時点標準の動画形式。ほぼ全ての動画プレーヤー対応

B	B	B	B	B	B	B	B
B	B	B	B	B	B	B	B
B	B	A	A	A	A	B	B
B	B	A	A	A	A	B	B
B	B	B	B	B	B	B	B
A	A	A	A	A	A	B	B
B	B	A	A	A	A	B	B
B	B	B	B	B	B	B	B

問題：左図のデータ(8×8ビット)の **Aの部分**を0、**Bの部分**を1として、次の約束に従って1行ごとに圧縮するとデータ量は何ビットになるか。また、圧縮率は何パーセントになるか。小数点以下第一位まで求めなさい。

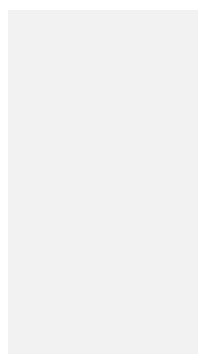
(1) **最初のビット：Aではじまる場合は0、Bではじまる場合は1とする。**

(2) 次の3ビット：最初のビットと同じ文字が続く個数を表す。
ただし、「個数-1」として表現する。

(3) 文字が変わるたびに、(2)と同様に3ビットで何個続くかを表す。

注：この問題で解を得る途中で53bit、64bitという値が登場するが、これらは便宜上53.0、64.0と解釈し有効数字を3桁と考える。解である82.8の小数第1位の信頼性を保つためである。

1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1
1	1	0	0	0	0	1	1
1	1	0	0	0	0	1	1
1	1	1	1	1	1	1	1
0	0	0	0	0	0	1	1
1	1	0	0	0	0	1	1
1	1	1	1	1	1	1	1



ビット
ビット
ビット =
ビット
ビット =
ビット =
ビット =
ビット =

圧縮率 = $\frac{\text{圧縮データ量}}{\text{元画像データ量}}$