

情報 I

- 1-8 データの圧縮と効率化実験
- 1-8-1 圧縮と伸張
- 1-8-2 可逆圧縮と非可逆圧縮
- 1-8-3 ランレングス圧縮



1-8-1 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** ）
- 圧縮したデータを元の状態に復元する技術（**伸張・展開・解凍**）
- 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、 $5/50 \cdots 10\%$

「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、 $(50-5)/50 \cdots 90\%$



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
作成日時	2020/07/23 21
更新日時	2020/07/23 21
サイズ	725 KB

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
作成日時	2020/07/23 21
更新日時	2020/07/23 21
サイズ	95.3 KB

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	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%INFC
作成日時	2020/07/23 21:50
更新日時	2020/07/23 21:50
サイズ	432 KB

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
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:%Docume
作成日時	2020/07/23
更新日時	2020/07/23
サイズ	60.7 KB

	7.Bmp	7.Gif	7.png	7.jpg
サイズ	725kB	95kB	432kB	61kB
ビットの深さ	24bit	8bit	32bit	24bit
解像度	W500px,H495px	W500px,H495px	W500px,H495px	W500px,H495px
圧縮率	—			

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

実験結果

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	—
	MP3	.mp3	c.mp3	44.4kB	9.19%
画像	BMP	.bmp	7.bmp	725kB	—
	GIF	.gif	7.gif	95kB	13%
	JPEG	.jpg	7.jpg	61kB	8.4%
	PNG	.png	7.png	432kB	60%
動画	AVI	.avi	sample.avi	11.6MB	—
	ZIP	.zip	sample.zip	6.82MB	58.80%
	MP4	.mp4	sample.mp4	10.5MB	90.5%
	MP4	.mp4	sample_low.mp4	833kB	7.01%



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イヤー対応**



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

実験結果

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	—
	MP3	.mp3	c.mp3	44.4kB	9.19%
画像	BMP	.bmp	7.bmp	725kB	—
	GIF	.gif	7.gif	95kB	13%
	JPEG	.jpg	7.jpg	61kB	8.4%
	PNG	.png	7.png	432kB	60%
動画	AVI	.avi	sample.avi	11.6MB	—
	ZIP	.zip	sample.zip	6.82MB	58.80%
	MP4	.mp4	sample.mp4	10.5MB	90.5%
	MP4	.mp4	sample_low.mp4	833kB	7.01%



1-8-2 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()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。
- 文字や数値データの場合、マルチメディアデータとは異なり、()方式を使う。

語群

可逆圧縮

ZIP

可逆圧縮

伸張・解凍

非可逆圧縮



1-8-3 ランレングス圧縮(連長圧縮)・・・可逆

連続したデータを、そのデータ一つ分と連続した長さで表現

- AAAAABBBBBBBBBBAAA 文字数(17)
- A5B9A3 (5A9B3A) 文字数(6)
- 問題点 ABCDDDD 文字数(7)
- A1B1C1D4 (1A1B1C4D) 文字数(8)

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(6)
- 例題 AAAABBCCCCCDEFG 文字数(16)
- 通常圧縮 A4B2C6D1E1F1G1 文字数(14)
- 解決策圧縮 AA2BB1CC3DEFG 文字数(13)
- 発展改善策 4A2B6C-4DEFG 文字数(11)

- 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



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

問題: 左図のデータ(8×8ビット)のAの部分で1, Bの部分で0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よ。

- (1)最初のビット:Aではじまる場合は1, Bではじまる場合は0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。

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

1 111 4 ビット
1 111 4 ビット
1 001 011 001 10ビット
1 001 011 001 10ビット
1 111 4 ビット
0 101 001 7 ビット
0 101 001 7 ビット
1 111 4 ビット

$$\begin{aligned}
 \text{圧縮率} &= \frac{\text{圧縮データ量}}{\text{元画像データ量}} \\
 &= \frac{4 \times 4 + 10 \times 2 + 7 \times 2}{8 \times 8} \\
 &= \frac{50}{64} \\
 &= 0.781 \\
 &= 78\%
 \end{aligned}$$

