

情報科学概論A

2017/05/19 第05回目
情報の効率的なデータ表現
データ圧縮と情報量

音データ

標本化

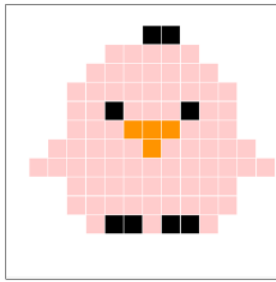
量子化

符号化

動画

- 画像と音の両方を合わせたもの
- 少しずつ動いている画像を
短い時間間隔で切り替えて投影
- イメージ的に「パラパラ漫画」
- 1秒間に切り替える画像の数を
「フレームレート」といい
「fps」という単位で表す

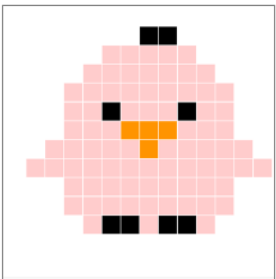
ラスタ形式の無圧縮画像



```
000000000003C00001550
00055540015555001757
50015A95005565541555
55501555500155550007
DF4000000000000000
```

- 3色で14*14の画像データ容量
- $14 * 14 * 2\text{bit/dot} = 392\text{bit}$

データ圧縮：ランレングス法



```
0 が 14個
0 が 7個, 3 が 2個, 0 が 5個
0 が 5個, 1 が 5個, 0 が 4個
0 が 3個, 1 が 7個, 0 が 2個
...続く
0 が 14個
```

- 色 2bit + 個数 4bit = 6bit/section
- $47\text{section} * 6\text{bit/section} = 282\text{bit}$

データ圧縮

- 可逆圧縮
 - 元情報を保ったままデータ量を減らす
 - 元のデータに戻すことができる
- 非可逆圧縮
 - 人間があまり強く認識できない成分を削除しデータ量を減らす
 - 元のデータに戻すことができない

日常生活で目にする圧縮形式

可逆圧縮

- ファイル
 - ZIP, LZH, RAR
- 画像
 - PNG
- 音楽
 - FLAC, ALAC

非可逆圧縮

- 画像
 - JPEG
- 音声
 - MP3, AAC
- 動画
 - MP4, MOV

データ圧縮に利用されている技術

- ランレングス法
- シャノン符号化
 - 圧縮率はよくない
 - 現在ではほぼ使われていない
 - ただこの符号化を元にして
様々な符号化手法が生み出された
- ハフマン符号化

事象の符号化

事象	年間確率	符号化
快晴	15%	
晴れ	43%	
曇り	15%	
雨	20%	
雷	2%	
雪	5%	

ハフマン符号化

事象	年間確率	符号化
快晴	15%	
晴れ	43%	
曇り	15%	
雨	20%	
雷	2%	
雪	5%	

- (1) 確率が高いものから順番に並べる
- (2) 確率が 1 番低いのと 2 番目に低いのを結合
- (3) 結合したものをグループにし確率を合算する
- (4) (2)に戻る。