

Tömörítési eljárások

jpeg, gif, GSM, compress, pkzip, (arc, rar ...) redundáns adatok:
tömörítés sikeres!

Veszteség nélküli és veszteséges eljárások

Futási hossz:

Pl. a 127 alatti karakter: darabszám a következő ismétlődő karakterre

128 feletti karakter: x-128 darab különböző karakter

Nem mindig hatásos: pl. egy ABABAB...

Delta kódolás:

csak a különbség: kevesebb bit!

Lineáris Prediktív Kódolás (LPC)

Lempel-Ziv-Welch eljárás alapötlet: szótár létrehozása

- minden bejövő adat: új szó a maximális egyezésű szótárelemmel
- tömörített kódot a maximális egyezésű szótárelem
- szótár mérete tipikusan 2^{12} - 2^{16}
- ha szótár megtelik → letörlik (csak az induló szótár marad meg ASCII!)

- visszaállítás/dekódolás: szótár újra építése
- adatfolyamban!!
- minden kód rekurzívan helyettesítődik a prefix kódjával+ a követő karakterrel
- HW megvalósítás

Huffman-kódolás:

Bejövő jelek eloszlása előzetesen pontosan ismert

sorbarendezés előfordulásuk valószínűségében

gyakoriak - kevés bit

ritkák - sok bit

rekurzívan a két legkisebb valószínűségű jel helyett új jelet vezet be, a két jel

valószínűségének együttes valószínűségével

Kódolás: pl. nagyobb valószínűségű jel 0, a kisebb 1

Visszalépünk egyet a rendezésben, s.í.t.

Prefix kódolás

Dekódolás: táblázattal
Egyértelmű kódhatárok!

Kérdés: Mekkora a példában szereplő karaktersorozat és annak Huffman kódolásának Shannon-entrópiája?

A statikus kódtábla felépítéséhez ismerni kell a jelet!

Aritmetikai kódolás

JPEG kódolás

Veszteséges kódolás (l. pl. wavelet) - hosszú kutatómunka, több ajánlás

JPEG Baseline coding

Kép felosztása 8×8 -as területekre (lokálisan adaptív)

Kódolás blokkonként: Karhunen-Loeve (főkomponens) transzformáció lenne a legjobb, de FFT is jó, és egyszerűbb:

Ne legyen komplex kimenet: diszkrét cosinus transzformációt (DCT) alkalmazunk

DCT: mintasorrend 1, 2, 3, 4, 5, 6, 7, 8, 7, 6, 5, 4, 3, 2, 1+FFT → szimmetria miatt valós lesz! Azaz 8×8 valós értékből 8×8 valós értéket készít!

A DCT értékeket kvantáljuk (max. 11 bit!):

kvantálási tábla:

Lineáris sorozattá alakítás:

egymás mellé kerülnek, a homogén, egyszínű felületek:
AC komponensek futási hossz kóddal tömörítve

DC komponensek: delta kóddal tömörítve (lassan váltooznak)

MPEG: mozgó JPEG