

Informatika – Komprese dat

Radim Farana

Podklady předmětu Informatika
pro akademický rok 2007/2008

Obsah

- Komprese dat, základní principy a použití.
- Bezeztrátová komprese,
 - Slovníkové metody.
 - Aritmetické metody.
- Ztrátová komprese.
- Fraktální komprese.

Komprese dat

- **Bezeztrátová** komprese - odstranění redundance:
 - Huffmanův kód,
 - Slovníkové metody,
 - opakování znaků (RLE),
 - Lempel-Ziv 1977, 1978, Lempel-Ziv-Welch 1984,
 - Aritmetická komprese
- **Ztrátová** komprese:
 - zvuku, obrazu, ...

Huffmanův kód

Triviální případ

Zpráva	$P(i)$	kód
1	>	0
2	<	1

Redukovaná abeceda

Zpráva	$P(i)$	redukce	kód	expanze
1	»	1	0	0
2	>	2,3	1	10
3	<			11

Postup:

- seřazení podle pravděpodobnosti,
- postupná redukce a oprava pořadí,
- přiřazení znaků 0, 1 a zpětná expanze.

Příklad

Zpráva	A	B	C	D	E
$P(i)$	0,4	0,3	0,1	0,1	0,1
1. redukce	0,4	0,3	0,2	0,1	
2. redukce	0,4	0,3	0,3		
3. redukce	0,6	0,4			
znaky	0	1			
kód	1	00	011	0100	0101

Problémy:

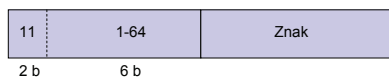
- definice pořadí zpráv pro stejnou $P(i)$,
- zařazení skupin se stejnou $P(i)$,
- pořadí přiřazení znaků 0, 1.

Varianty komprese

- **Statická komprese** – dvouprůchodová,
 - zjištění četnosti jednotlivých znaků,
 - sestavení optimálního kódu,
 - komprese souboru.
- **Dynamická komprese** – jednaprůchodová,
- průběžné upřesňování stromu kódu.

Opakování znaků

- Sekvence znaků kódována jako:
 - zarážka, opakovaný znak, počet
- Výskyt zarážky vede na tvar:
 - zarážka, zarážka, 1
- **RLE – Run Length Encoding** (součást formátu PCX) – sloučení zarážky a počtu:



LZ, LZW



Abraham Lempel



Jacob Ziv

Terry A. Welch

http://www.hpl.hp.com/about/bio/abraham_lempe.html

* 27. 11. 1931 Tiberias, Izrael
http://www.marcofoundation.org/pages/followed/followed_00000000.htm

● 1977 LZ, vylepšena 1978

- vyhledávání opakujícího se řetězce v kruhovém zásobníku, kódování s využitím devátého bitu:
0xxxxxxx – nekomprimovaný znak
1xxxxxxx – komprese: odkaz na začátek řetězce a jeho délka

● 1984 LZW

- Využití hašování (hešování) – moderní metody rychlého vyhledávání.

Aritmetická komprese

1. Znak R

$$D_1 = D_0 + \frac{m_R}{n}, L_0 = 0 + \frac{2}{10} = \frac{2}{10}$$

$$H_1 = D_1 + \frac{m_C}{n}, L_1 = \frac{2}{10} + \frac{4}{10} = \frac{6}{10}$$

$$L_1 = H_1 - D_1 = \frac{6}{10} - \frac{2}{10} = \frac{4}{10}$$

2. Znak C

$$D_2 = D_1 + \frac{m_C}{n}, L_2 = \frac{2}{10} + \frac{0}{10} + \frac{4}{10} = \frac{20}{100}$$

$$H_2 = D_2 + \frac{m_S}{n}, L_2 = \frac{20}{100} + \frac{2}{10} + \frac{4}{10} = \frac{28}{100}$$

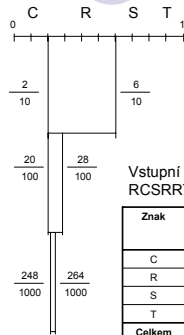
$$L_2 = H_2 - D_2 = \frac{28}{100} - \frac{20}{100} = \frac{8}{100}$$

3. Znak S

$$D_3 = D_2 + \frac{m_S}{n}, L_3 = \frac{20}{100} + \frac{6}{10} + \frac{8}{1000} = \frac{248}{1000}$$

$$H_3 = D_3 + \frac{m_T}{n}, L_3 = \frac{248}{1000} + \frac{2}{1000} + \frac{8}{100} + \frac{264}{1000} = \frac{264}{1000}$$

$$L_3 = H_3 - D_3 = \frac{264}{1000} - \frac{248}{1000} = \frac{16}{1000}$$



Vstupní soubor:
RCSRRTCSTR

Znak	Počet znaků n_{znak}	Počet menších m_{znak}
C	2	0
R	4	2
S	2	6
T	2	8
Celkem	10	

Vícenásobná komprese

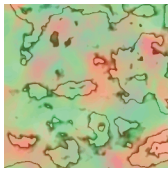
- **Primární** komprese – výchozí,
 - vhodné jsou slovníkové algoritmy.
- **Sekundární** komprese – následná,
 - vhodný Huffmanův kód, aritmetická komprese.
- Další komprese nepřináší užitek.
- **Kvalita** komprese,
 - kompresní poměr (komprimovaný soubor oproti původnímu).

Ztrátová komprese

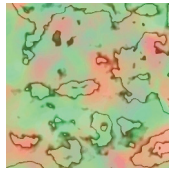
- DCT (Discrete Cosine Transform):
 - signál je nahrazen složením řady harmonických funkcí s teoreticky nekonečným počtem frekvencí.
 - využití kódování informace o barvě pomocí složek jasu a barevnosti (Y, cblue, cred)
 $Y = 0,3 R + 0,59 G + 0,11 B$
 $cblue = -0,17 R - 0,33 G + 0,5 B + 128$
 $cred = 0,5 R - 0,42 G - 0,08 B + 128$
 - určení převládající barvy v poli 8 x 8 bodů.

Komprese JPEG

Původní obrázek ve formátu BMP
(velikost 66 814 B)



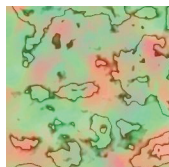
Bezeztrátová komprese JPEG
(velikost 56 463 B),



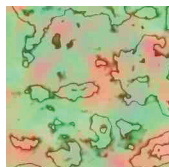
Informační obsah

Malé ztráty informace

Odstranění vysokých frekvencí (šum)
(velikost 13 816 B)



Zmenšení na desetinu obsahu
(velikost 5 696 B),



Velké ztráty informace

Odstranění detailů 5 % velikosti
(velikost **2 596 B**)



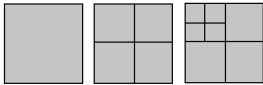
Zmenšení na 3 % velikosti
(velikost **2 001 B**),



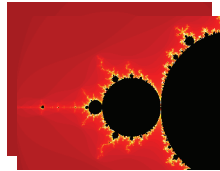
Fraktální komprese

- Hledání podobných tvarů

- uložení informace o tvaru, posunutí, otočení, změně velikosti a barvy.

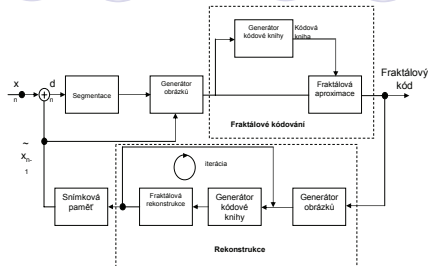


- velmi náročná komprese, rychlá dekomprese.



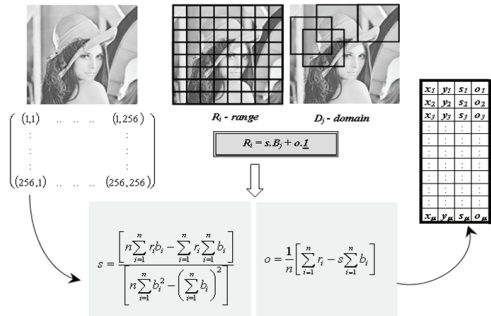
Benoit B. Mandelbrot
* 20. 11. 1924 Warsaw, Poland
(French mathematician)
<http://www.math.yale.edu/mandelbrot>

Algoritmus fraktálové komprese



Implementace fraktálového kodéru do standardu MPEG

Postup fraktálového kódování



Obr.2 Postup fraktálového kódování obrazu

Postup dekódování obrazu

