

Multimediální systémy

Jan Outrata



KATEDRA INFORMATIKY
UNIVERZITA PALACKÉHO V OLMOUCI

přednášky

1 Úvod

Historie, multimediální data a systémy, klasifikace.

2 Získání obsahu

Digitalizace analogového signálu, Fourierova transformace, vzorkovací věta (Sampling theorem), filtrace.

3 Reprezentace a formáty, barvy

Reprezentace a formáty obrazu, videa, audia a grafiky, problém barev, barevné prostory (modely).

4 Zpracování obsahu (authoring), komprese

Authoring a nástroje, základní pojmy z teorie informace, kódování a komprese, základní bezztrátové a ztrátové kompresní metody.

5 Komprese obrazu a videa

Bezeztrátové, JPEG, wavelet, fractal, predikce pohybu, standardy pro kompresi videa (MPEG).

6 Komprese audia a grafiky

PCM, psychoakustika, modely zvuku, standardy pro kompresi audia (MPEG, Dolby, MIDI), topological surgery, progresivní síť, standardy pro kompresi grafiky.

Anotace

Předmět poskytuje základy problematiky digitálního videa a audia a jeho distribuce po datové síti a pomocí digitální televize (DVB).

V první části jsou probírány reprezentace a formáty digitálního obrazu, videa a audia, jejich získávání, zpracování, komprese a distribuce po datové síti a také tvorba DVD. Zmíněna je i problematika DRM. Ve druhé části se pak studenti seznámí s problematikou digitální televize (DVB). Obsahem cvičení je praktické zacházení s obrazem, videem a audiem od jejich získání, přes editaci a kompresi až po distribuci - studenti si zkusí vytvořit DVD a provozovat streamovací server.

Předpoklady: Základní znalosti problematiky kódování a ztrátové komprese a počítačových sítí.



- Havaladar P., Medioni G.: *Multimedia Systems: Algorithms, Standards, and Industry Practices*. Course Technology, 2009. ISBN 978-1418835941
- Savage T. M., Vogel K. E.: *An Introduction to Digital Multimedia*. Jones & Bartlett Publishers, 2008. ISBN 978-0763750527
- Hwang J.-N.: *Multimedia Networking: From Theory to Practice*. Cambridge University Press, 2009. ISBN 978-0521882040
- Taylor J., Johnson M. R., Crawford Ch. G.: *Velký průvodce DVD*. Grada, 2007. ISBN 978-80-247-1721-0
- Wallace K.: *VoIP Bez předchozích znalostí*. Computer Press, 2007. ISBN 978-80-251-1458-2



Úvod

- „multimedia“ (1990) ~ hraní počítačových her, poslouchání hudby, sledování filmů a videí, videokonference, ... cokoliv s počítačem zahrnující obraz a zvuk?
- *multimedia* = zahrnující různé typy médií: text, obraz, video, audio → **multimediální obsah/data**

Multimediální systém

= systém pro vytvoření, uložení, manipulaci a distribuci multimediálního obsahu

- např. nahrávací a přehrávací technika obrazu, audia a videa (foťáky, kamery, rádio a TV s možností uložení a volby), prezentace (slajdy), interaktivní komunikační systémy, PC hry, modelovací systémy, ..., webový prohlížeč – multimediální hardware a software
- využití oblastí zpracování a analýzy textu a audio/video signálu, 2D/3D grafiky, teorie informace a kódování, sítí, databází, uživatelských rozhraní, ... pro praktická řešení získávání informací, komunikace, zábavy atd.

Vývoj

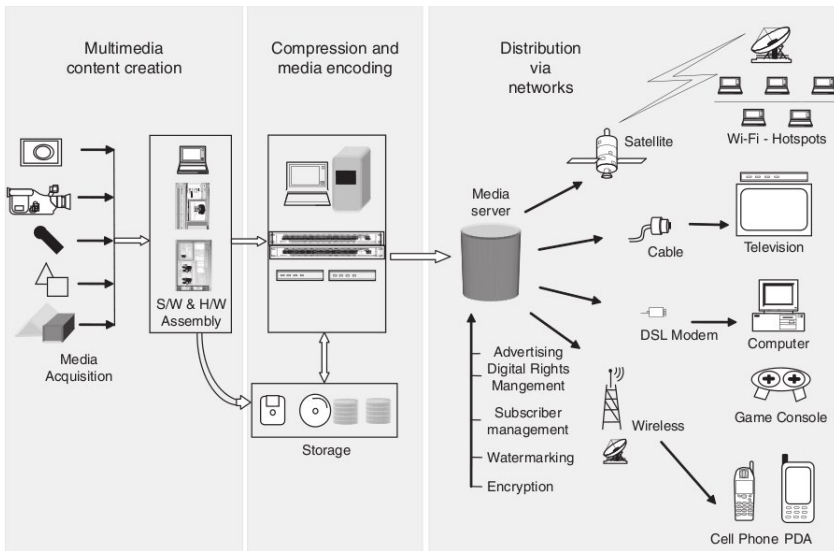
- malby v jeskyních (tisíce let před letopočtem)
- malby a text na papíře (stovky let před letopočtem) a v knihách (stovky let) šířené poštou (–1800)
- foto, audio, video na fotofilmech, deskách a magnetických páskách distribuované poštou, rádiem a TV (–1980)
- digitální obsah všeho druhu na optických a elektronických médiích s distribucí po sítích (internetu, dodnes)



- zvyšování množství informací a velikosti obsahu, zrychlování distribuce
- po audio průmyslu (CD) kombinace textu, obrazu, videa a audia (multimediální CD – informační i hry), od interakce člověka s PC k interakci mezilidské
- **multimediální revoluce** (posledních 30 let) = rozvoj metod a zařízení pro tvorbu a uložení (digitálního) obsahu, jejich dostupnosti a použitelnosti, rozvoj sítí a internetu a dostupnost (mobilního) připojení k internetu odkudkoliv
- standardizace ISO, ITU, průmysl: JPEG, MPEG, DVD, OSI aj.

- obecné pro informace, nejen multimediální obsah
- 1** získání a tvorba obsahu = **authoring**
 - digitalizace a zpracování médií, prvotní reprezentace
- 2** reprezentace obsahu pro uložení a přenos
 - **komprese**: ztrátová i bezztrátová
- 3** distribuce obsahu
 - po (paketových) sítích uživatelům do PC, TV, mobilů aj.
 - streamování, zabezpečení (DRM)

Obrázek: Komponenty multimediálního systému





- text, obraz, video, audio – dnes, je možné vytvořit, uložit a šířit, co v budoucnu??

Vlastnosti

- digitální forma – jednoduchá reprezentace a manipulace, ale zkreslení
- objemnost – detaily, ale problémy při vyhledávání a přenosu
- interaktivita – výběr obsahu pomocí voleb a odkazů, přímé ovládání
- real-time a synchronizace – aktuálnost, ale problémy při prezentaci a přenosu

- **text**: posloupnost znaků, formátovaný a provázaný odkazy (hypertext), (stále) nejpoužívanější, nejjednodušší
- (rastrový) **obraz**: 2D pole obrazových bodů (pixelů), atributy velikost (šířka, výška), poměr stran (aspect ratio, šířka/výška, např. 4:3, 16:9), hloubka (bitů/pixel)
- (vektorová) **grafika**: množina 2D a 3D vektorových elementů, atributy barva (výplně), tloušťka čáry, atd.
- **video/animace**: posloupnost obrazů se stejnými (prostorovými) atributy, další časový atribut rychlost (frame rate, obrazů/s), u rastrového obrazu skenovací formát (skenování = převod posloupnosti 2D obrazů na posloupnost 1D vzorků), např. prokládaný, progresivní), dimenze? (obrazy v obrazu)
- **audio**: posloupnost úrovní zvuku (vzorků), atributy rychlost vzorkování (sampling, vzorků/s, určuje frekvenční rozsah), velikost vzorku, dimenze (počet kanálů, např. mono, stereo, surround)

- v závislosti na použití
 - statický vs. dynamický – obsah v čase stejný nebo proměnný
 - real-time? – „živý“ přenos?
 - lineární vs. nelineární – souvislá nebo libovolná interakce (pomocí odkazů, hypertext, **hypermedia**)
 - interakce uživatel-PC vs. uživatel-uživatel
 - distribuce k jednomu nebo více konzumentům (bez omezení)
 - ...

Získání obsahu

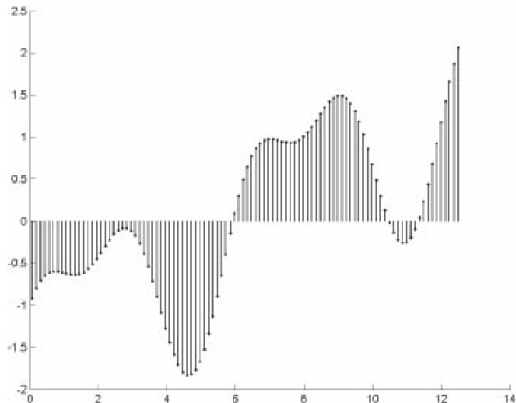
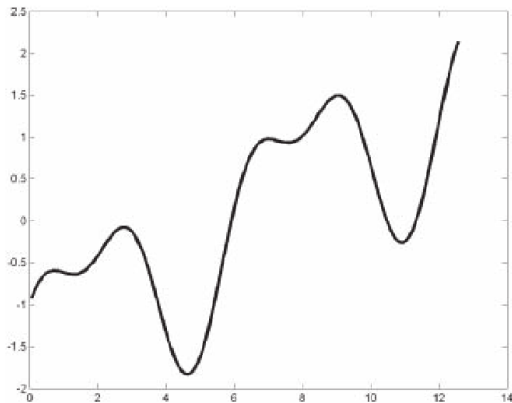
- text, grafika, animace: přímo
 - audio: jednorozměrný signál v čase
 - obraz: dvojrozměrný signál v prostoru
 - video: trojrozměrný signál v prostoru (2 dimenze) a čase
- převod **spojitého analogového signálu** ze záznamového zařízení na digitální data – oblast **zpracování signálu**
- určuje kvalitu a kvantitu digitalizovaného obsahu

- **analogový signál** = spojitá funkce domény (času, prostoru nebo obojího) reprezentující úroveň signálu, např. amplituda v čase
- **digitální signál** = posloupnost diskrétních hodnot úrovní signálu vyjádřených v binární (digitální) formě získaných ve specifických bodech domény (času, prostoru nebo obojího), např. velikost amplitudy v pravidelných stejně velkých intervalech času

Obrázek: Analogový a digitální signál

Výhody digitálního signálu (a obráceně nevýhody analogového)

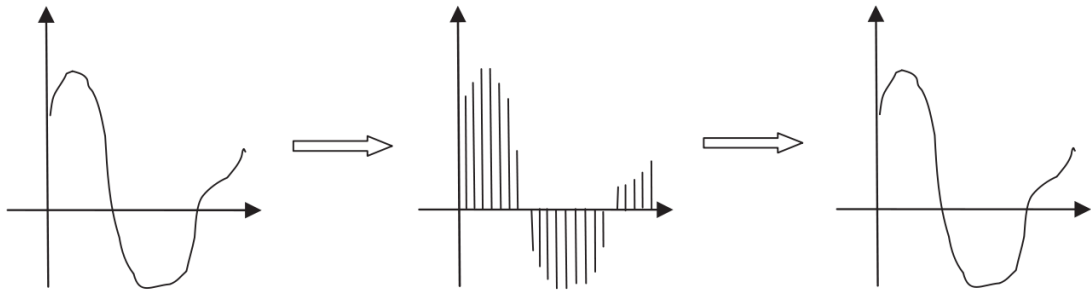
- jednotná reprezentace pro uložení a přenos – v digitální formě na záznamová média, v počítačové síti; ale obvykle objemná. . .
- jednoduchá editace a kombinace (včetně interaktivity) – přístup k „jednotkám“ informace, např. pixel; ale některé úpravy mohou být (výpočetně) náročnější. . .
- efektivní uložení, přenos, zabezpečení – komprese, šifrování; ale může být degradace informace. . .
- vyšší spolehlivost – možnost uložení bez (další) degradace informace, médiiem, prostředím, časem, např. tzv. „duchy“ u analogového videa; ale nároky na úložný systém



- = Analog to Digital Conversion, ADC
- = **vzorkování (sampling)** a **kvantizace**
 - obrácený převod (Digital to Analog Conversion, DAC) = **interpolace** – pro prezentaci obsahu pomocí (analogového) prezentačního zařízení
- cílem je signál co nejpodobnější původnímu analogovému signálu před digitalizací (zachování signálových charakteristik)

Obrázek: Převod analogového signálu na digitální a zpět

Převod analogového na digitální signál





- = **diskretizace** spojitého signálu na posloupnost diskrétních hodnot (**vzorků**)
 - jednorozměrný analogový signál (v doméně t): $x(t)$
- **navzorkovaný signál**: $\mathbf{x}_s(\mathbf{n}) = \mathbf{x}(\mathbf{nT})$, $n = 1, 2, \dots$, T je vzorkovací perioda, $f = 1/T$ je **vzorkovací frekvence**
 - perioda/frekvence – kritický parametr, podvzorkování vede k artefaktům (aliasům) při prezentaci, nadvzorkování k redundantní reprezentaci
 - přibližné hodnoty – získané měřením (a filtrováním) signálu a průměrováním hodnot



- Nyquist, 1920s, Shannon, 1950
- popisuje vztah mezi (digitální) reprezentací signálu a vzorkovací frekvencí
- = pro plnou rekonstrukci ze vzorků musí být analogový signál vzorkován s frekvencí rovnou alespoň dvojnásobku maximální frekvence v signálu, tzv. **Nyquist vzorkovací frekvencí**:

Theorem (Sampling theorem)

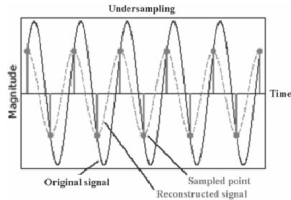
- 1** Signál $f(t)$ s maximální frekvencí ωF je plně určen vzorky $f_s(nT)$, jestliže $\frac{2\pi}{T} > 2\omega F$.
- 2** Spojitý signál $f(t)$ lze zrekonstruovat ze vzorků $f_s(nT)$ pomocí konvoluce s filtrem $r(t) = \text{sinc}\left(\frac{\omega F(t-nT)}{2\pi}\right)$.

- při podvzorkování nejsou zachyceny vyšší frekvence, artefakty „rozostření“ = **aliasing** (= ztráta informace při digitalizaci), efekty např. plochost a tupost zvuku, rozostření obrazu nebo tzv. moiré efekt u opakujících se vzorů v obrazu, jiná rychlost a směr pohybujících se částí videa apod.

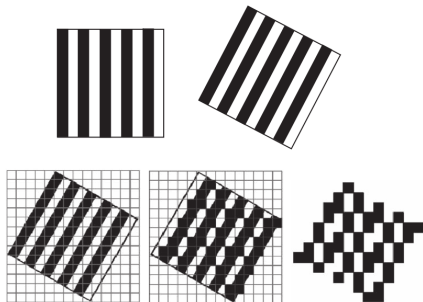
Obrázek: Efekty aliasingu

- potřeba znát **frekvenční rozsah** signálu, nebo určit podle požadovaného typu signálu, např. hudba nebo řeč

Vzorkovací věta (Sampling theorem)



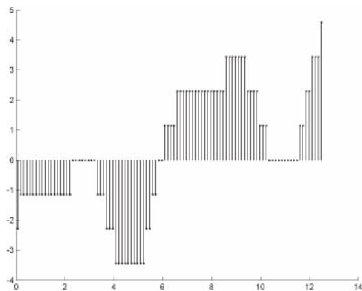
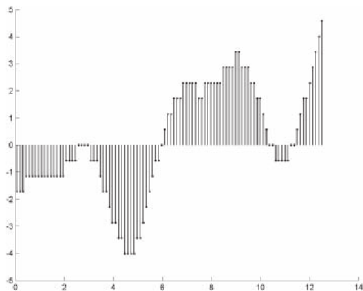
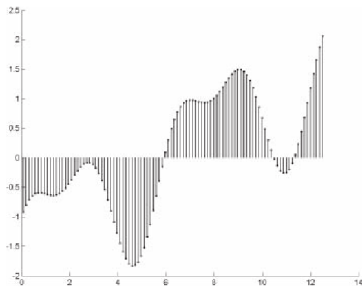
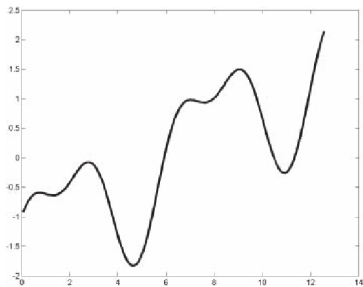
Vzorkovací věta (Sampling theorem)



- = vyjádření navzorkovaného signálu jako posloupnosti binárních (digitálních) hodnot
- **digitální signál**: $\mathbf{x}_q(\mathbf{n}) = \mathbf{Q}(\mathbf{x}_s(\mathbf{n}))$, $n = 1, 2, \dots$, kde Q je zaokrouhlovací funkce mapující vzorky na nejbližší binární hodnotu s uložením do b bitů
- počet b bitů pro hodnotu udává počet $N = 2^b$ různých binárních hodnot vzorků (a tedy **kvantizačních úrovní** analogového signálu z rozsahu R) a přesnost binární reprezentace (kvantizační krok $\delta = R/2^b$) – kritický parametr

Obrázek: Kvantizace s různým počtem úrovní

- zaokrouhlování na konečný počet úrovní = rozdíly analogových signálů = **kvantizační chyby** – max. polovina kvantizačního kroku





6 bits



5 bits



4 bits



3 bits



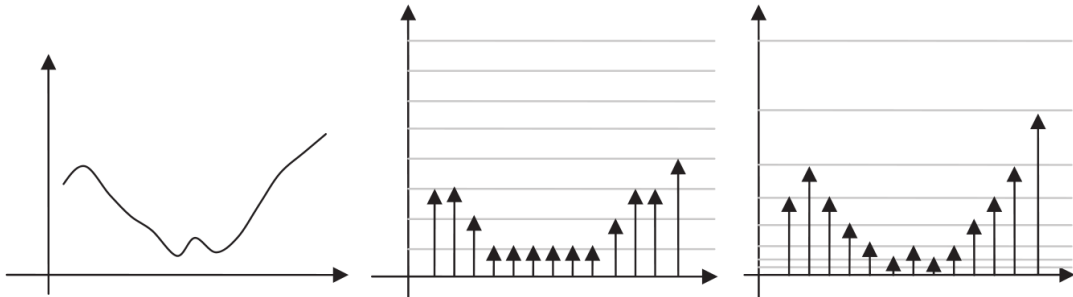
2 bits



1 bit

- *kolik bitů pro vzorek?* – záleží na hodnotovém rozsahu signálu, jeho typu, použití apod., např. hudba nebo řeč
- **kvantizační schéma** – rozložení kvantizačních úrovní v hodnotovém rozsahu signálu, např. uniformní, lineární, logaritmické (např. u dynamického rozsahu audia), určuje distribuci chyby

Obrázek: Kvantizační schémata



= počet bitů pro digitální signál za sekundu

- závisí na periodě/frekvenci vzorkování a počtu bitů pro vzorek:

$$\begin{aligned} \text{bit rate} &= \frac{\text{bitů}}{s} = \frac{\text{počet vzorků}}{s} \times \frac{\text{bitů}}{\text{vzorek}} \\ &= \text{vzorkovací frekvence} \times \text{bitů na vzorek} \end{aligned}$$

- ideální hodnota = kompromis mezi uložením maxima informace (bez redundance) a minimální velikostí dat (s akceptovatelnou degradací informace)

Tabulka: Příklady bit rate u různých signálů

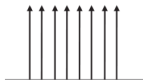
Signal	Sampling rate	Quantization	Bit rate
Speech	8 KHz	8 bits per sample	64 Kbps
Audio CD	44.1 KHz	16 bits per sample	706 Kbps (mono) 1.4 Mbps (stereo)
Teleconferencing	16 KHz	16 bits per sample	256 Kbps
AM Radio	11 KHz	8 bits per sample	88 Kbps
FM Radio	22 KHz	16 bits per sample	352 Kbps (mono) 704 Kbps (stereo)
NTSC TV image frame	Width – 486 Height – 720	16 bits per sample	5.6 Mbits per frame
HDTV (1080i)	Width – 1920 Height – 1080	12 bits per pixel on average	24.88 Mbits per frame

- typy signálu: spojité, hladké, symetrické, periodické atd.
 - (lineární) invariantní v doméně = (lineární) transformace, kde výstupní hodnoty v daném bodě domény závisí pouze na vstupních hodnotách ve stejném bodě domény, ne na bodě nebo změně hodnot domény
 - **konvoluce**: $f * g = \int_{-\infty}^{\infty} f(\tau) \cdot g(t - \tau) d\tau$
- vzorkování = konvoluce vstupního signálu a tzv. impulse response funkce (fungující jako filtr) – **doménový pohled**
- **impulse reponse funkce** – např. (Diracova) Delta ($\delta(x) = 0$ pro $x \neq 0$, $\lim_{x \rightarrow 0} \delta(x) = \infty$, $\int_{-\infty}^{\infty} \delta(x) dx = 1$, jednotka konvoluce), comb, box, step, sinc ($= \frac{\sin(\pi x)}{\pi x}$)

Obrázek: Grafy impulse response funkcí



Delta function



Comb function

$$\Delta_T(t) \stackrel{\text{def}}{=} \sum_{k=-\infty}^{\infty} \delta(t - kT)$$

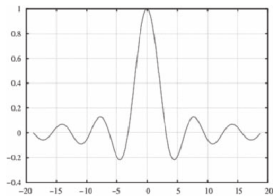


Step function $H(x) = \int_{-\infty}^x \delta(t) dt$



Box function

$$\text{rect}(t) = \Gamma(t) = \begin{cases} 0 & \text{if } |t| > \frac{1}{2} \\ \frac{1}{2} & \text{if } |t| = \frac{1}{2} \\ 1 & \text{if } |t| < \frac{1}{2} \end{cases}$$



Sinc function $\text{Sa}(x) = \text{sinc}(x) = \frac{\sin(\pi x)}{\pi x}$



- **frekvenční pohled** – ve **frekvenční doméně** (doméně frekvence změn hodnot): součin transformovaného vstupního signálu a tzv. transfer funkce = transformovaná impulse response funkce
- nahrazení konvoluce součinem, transformace = **Fourierova transformace**

– Fourier, 1807, Dirichlet

= reprezentace periodického spojitého signálu $f(t)$ jako součtu vážených sum sinusovek a kosinusovek (Fourierova řada):

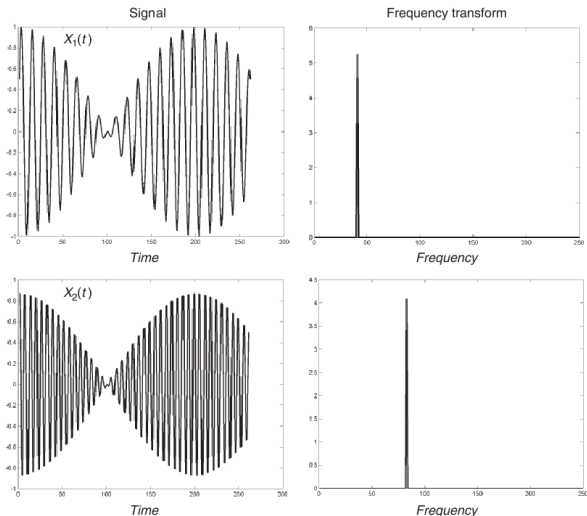
$$f(t) = \sum_{i=0}^{i=\infty} A_i \times \sin(i\omega t) + \sum_{j=0}^{j=\infty} B_j \times \cos(j\omega t)$$

- váhy = **frekvenční koeficienty** A_i, B_j = tzv. spektrální komponenty – určují amplitudy základních frekvencí (násobky $\omega = 2\pi/T$) v signálu a **plně signál reprezentují ve frekvenční doméně** jako funkci základních frekvencí
- běžnější reprezentace pomocí pouze kosinusovek (nebo analogicky sinusovek) = **kosinová** (sinová) **transformace**:

$$f(t) = \sum_{i=0}^{i=\infty} B_i \times \cos(i\omega t), B_i = \int_{-\infty}^{\infty} f(t) \times \cos(i\omega t) dt$$



Obrázek: Fourierova (kosinová) transformace analogového signálu



- pro digitální signál s diskrétními hodnotami - **diskrétní Fourierova** nebo **kosinová transformace** (**DFT** nebo **DCT**):

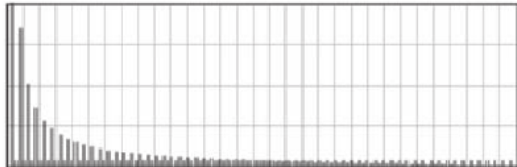
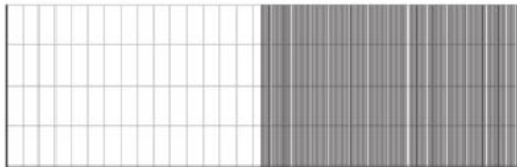
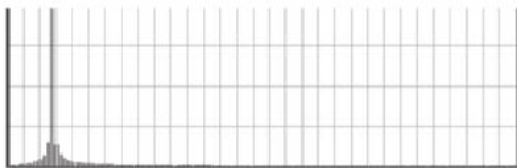
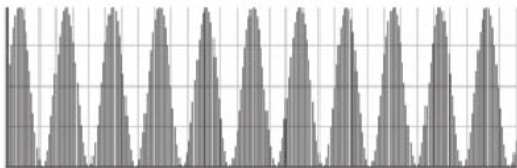
$$f_s(n) = \sum_{i=0}^{i=N-1} A_i \times \sin\left(\frac{i(2n+1)\pi}{2N}\right) + \sum_{j=0}^{j=N-1} B_j \times \cos\left(\frac{j(2n+1)\pi}{2N}\right), n = 0, \dots, N-1$$

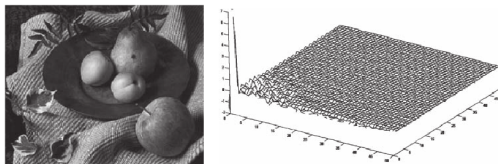
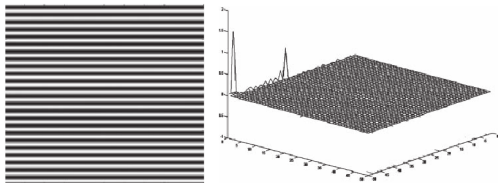
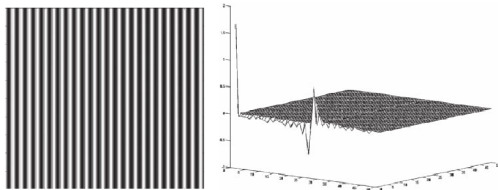
$$f_s(n) = \sum_{i=0}^{i=N-1} B_i \times \cos\left(\frac{i(2n+1)\pi}{2N}\right), B_i = \sum_{n=0}^{n=N-1} f_s(n) \times \cos\left(\frac{i(2n+1)\pi}{2N}\right)$$

Obrázek: Diskrétní Fourierova (kosinová) transformace digitálního signálu

- lze zobecnit pro lib. počet dimenzí signálu

Obrázek: Fourierova (kosinová) transformace signálu dvojrozměrného signálu (obrazu)





- **analogový filter**: (hardwarové) odfiltrování nechtěných frekvencí, např. šumu, ekvalizace, analogového signálu před digitalizací, pro snížení potřebné bit rate
- **digitální filter**: (numerické) obecné úpravy (např. i různé „efekty“) digitálního signálu, pomocí **digitálních signálových procesorů (DSP)**

Výhody digitálních filtrů

- obecné úpravy signálu – včetně adaptace filtru podle signálu
- snadná tvorba, změna, kombinace – programovatelnost
- stabilita (vzhledem k času a prostředí) – software
- zvládnutí i velmi nízkých frekvencí – narušitel od analogových

Kategorizace

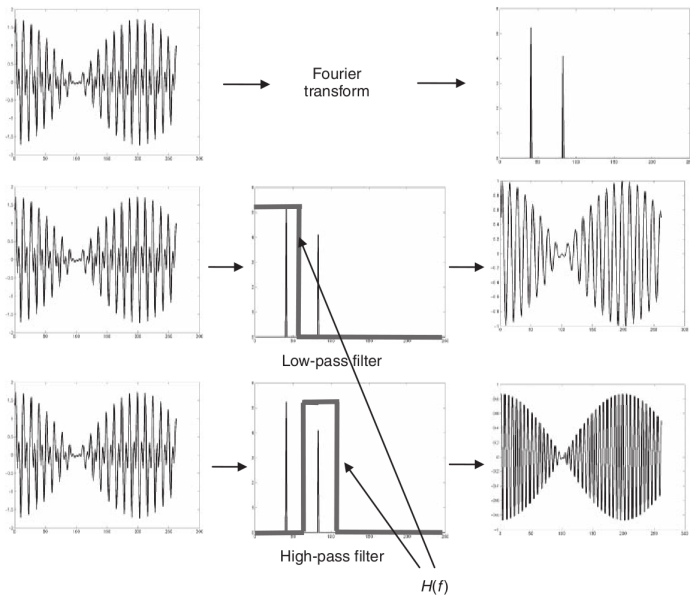
- **low-pass** - odstraňují vyšší frekvence, pro předcházení/odstranění aliasingu (tzv. **antialiasing**) při podvzorkování, např. často u obrázků

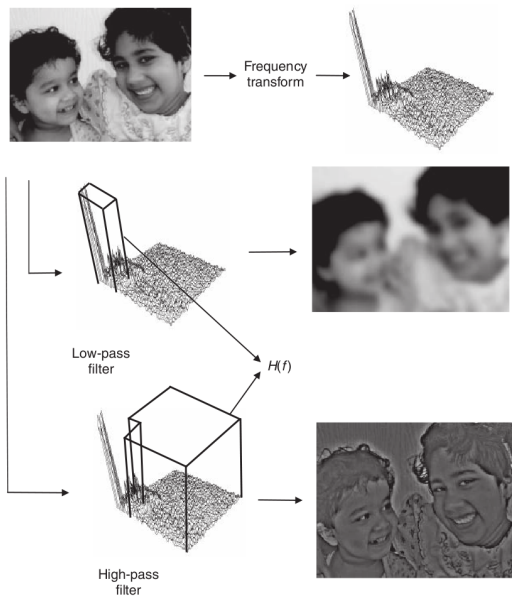
Obrázek: Ilustrace antialiasingu při podvzorkování

- **high-pass** - odstraňují nižší frekvence, pro zaostření signálu (tzv. **sharpening**)
 - **band-pass** - odstraňují frekvence mimo definované pásmo, pro zvýraznění frekvencí v signálu
- = efekty ve frekvenční doméně → studie a použití skrze frekvenční doménu

Obrázek: Efekty filtrů



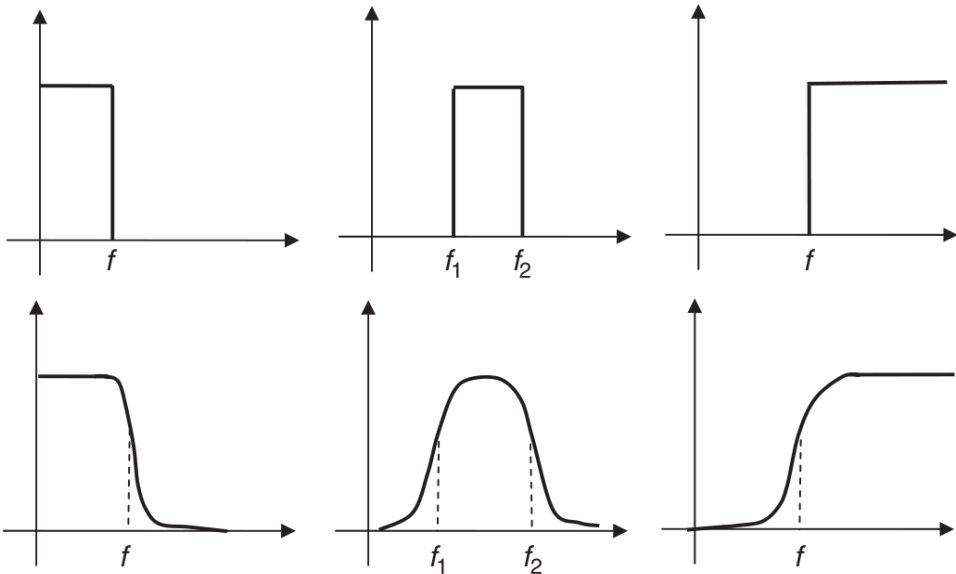




- hranice mezi zachovanými a odstraněnými frekvencemi nebývá ostrá (složitost až nemožnost realizace filtru) – aproximace, tzv. **filter roll-off** = další chyby

Obrázek: Frekvenční charakteristiky filtrů (ideální a reálné)

→ vzorkovací frekvence bývá „o něco“ vyšší než Nyquistova, např. 44 kHz místo 40 kHz u audia (CD kvality) s max. frekvencí 20 kHz





Reprezentace a formáty

Reprezentace

- = 2D pole obrazových bodů (**pixelů**) – atribut **rozlišení** v počtu pixelů na jednotku délky
= **velikost pixelu**
- atributy **šířka** a **výška** v počtu pixelů a **hloubka** v bitech na pixel
 - **aspect ratio** = poměr šířky a výšky, např. 3:2 (fotografie), 4:3 (TV), 16:9 (širokoúhlý, HD), 47:20 (anamorphic, kino), změna mění velikost pixelu (tzv. pixel aspect ratio, typicky čtvercové 1:1) – vnímáno jako roztažení/stlačení obrazu → převzorkování
 - přechody mezi aspect ratio: **pan/scan** = oříznutí bočních okrajů, např. z 16:9 na 4:3, nebo horních okrajů, např. z 4:3 na 16:9, **letterbox** = přidání prázdné (černé) plochy nad a pod, např. z 16:9 na 4:3, nebo na strany, např. z 4:3 na 16:9

Obrázek: Ilustrace změny aspect ratio



- hloubka závisí na reprezentaci barev (barevného prostoru) – rozdělena do **barevných kanálů**, např. pro odstíny šedé 1 kanál 8 b = hloubka 8 b, pro barevnou RGB 3 kanály R (červená), G (zelená), B (modrá) po 3, 3, 2 b = hloubka 8 b (8-bit color), 5, 5/6, 5 b = hloubka 15/16 b (tzv. high color), 8, 8, 8 b = hloubka 24 b (tzv. true color) nebo více bitů na kanál (tzv. deep color)
- **alfa kanál** – míra průhlednosti pixelu při skládání obrazů (z vrstev), 0 = neprůhledný, např. 8 b (true color) nebo 1 b (high color), proměnné hodnoty např. na hranicích složených obrazů pro potlačení aliasingu

Obrázek: Využití alfa kanálu při skládání obrazů

- tzv. **halftone obraz** – technika simulace barvy nebo plynulého přechodu mezi barvami pomocí pravidelného rozmístění bodů jiných barev a proměnné velikosti místo pixelů, velikost v počtu bodech na jednotku délky, používané při tisku, barevné kanály C (azurová), M (fialová), Y (žlutá), K (černá)

Obrázek: Halftone obraz



Background

+

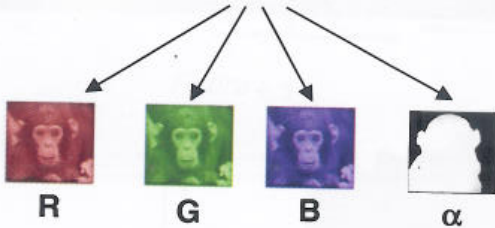


Foreground

=



Final



$$Final[i][j] = Foreground[i][j] * \alpha [i][j] + Background[i][j] * (1 - \alpha [i][j])$$



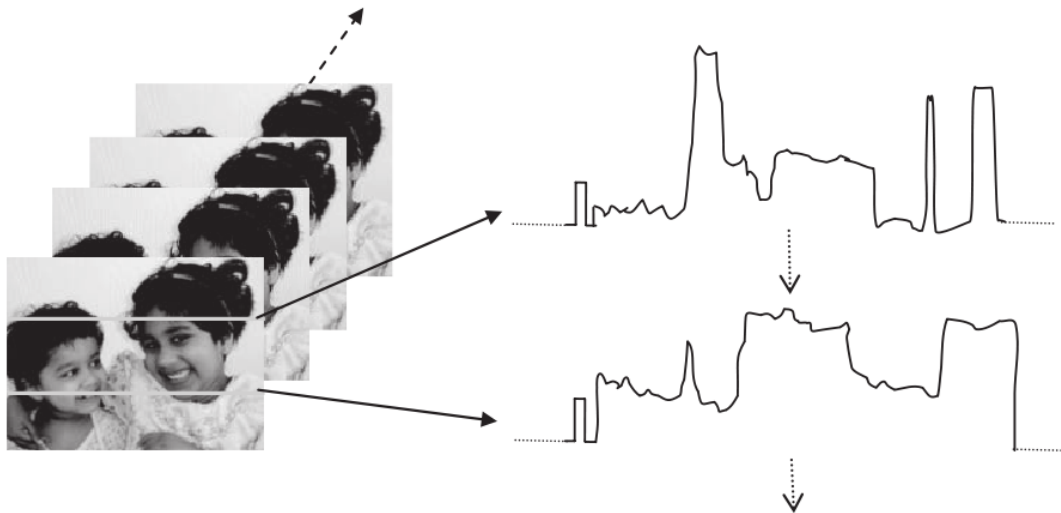
Formáty

- nekomprimované (**raw**) např. Tagged Image File Format (TIFF, i komprimovaný), Windows bitmap (BMP, hloubka až 24 b), Adobe Photoshop (PSD, více barevných modelů, vrstvy), proprietární např. z fotoaparátu
- komprimované např. JPEG (ne alfa), Portable Network Graphics (PNG, hloubka až 48 b, alfa), Graphics Interchange Format (GIF, 256 barev z palety, alfa, LZW komprese)

Reprezentace

- = posloupnost obrazů, tzv. **snímků (frame)**, se stejnými atributy, zobrazovaných rychle po sobě
- **frame rate** = rychlost zobrazování obrazů v obrazech/s (frames per second, fps), např. 24 (film), 25/30 (TV, PAL/NTSC), při nedostatečné vnímáno „blikání“ (flickering)
- z historických důvodů jsou standardy digitálního videa těsně svázány se standardy videa analogového – přenos a zobrazování z analogového signálu na digitálních zařízeních nebo digitálního na analogových
- převod na jednorozměrný signál jako funkci hodnot bodů (pixelů) řádků obrazů v čase, tzv. **skenování (scanning)** – kvůli distribuci (přenosu) analogové TV a zobrazování na analogových CRT obrazovkách

Obrázek: Analogový video signál

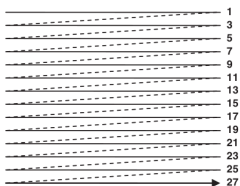


- **skenovací formát** – body na řádku zleva doprava, řádky shora dolů, např. prokládaný, progresivní
- **prokládané (interlaced)** skenování: snímek skenován (a zobrazován) po dvou částech, tzv. **půlsnícímích** = liché a sudé řádky – pohybové artefakty, pokud jsou půlsnímky složeny a zobrazeny zároveň, progresivně (např. na digitálním zařízení)

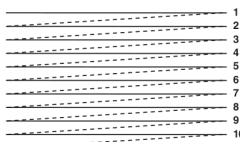
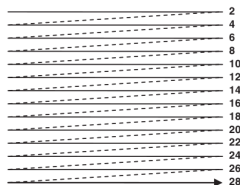
Obrázek: Prokládané skenování

- **progresivní** skenování: snímek skenován (a zobrazován) celý, tzn. všechny řádky

Obrázek: Progresivní skenování



Upper field

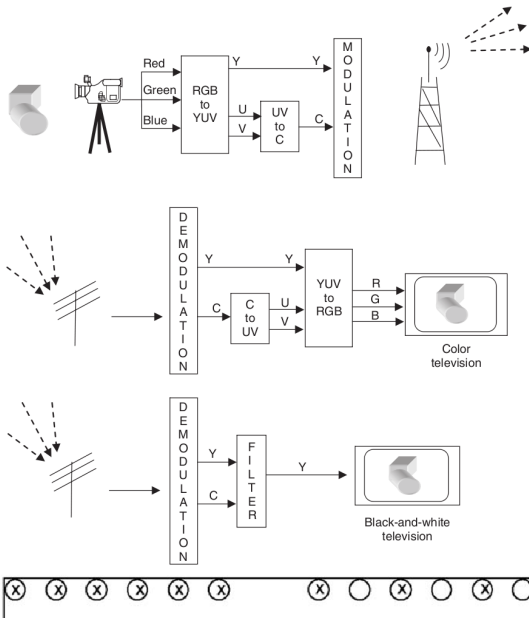


- kvůli přenosu TV signálu (synchronizaci a omezenému pásmu): prokládané skenování (synchronizační impulsy, vertikální pro další snímek 50/60-krát/s) a převod barevného prostoru RGB na YUV s podvzorkováním

Obrázek: Přenos TV signálu

- **barevný prostor YUV**: jasová Y a barevné U, V (C) složky, stejné hloubky jako u RGB
- **YUV podvzorkování** = podvzorkování barevných U, V složek pro redukci pásma, redukce barevného rozlišení méně ovlivňuje vnímání obrazu než redukce jasového rozlišení
 - u analogového signálu: polovina pásma pro jasovou složku
 - u digitálního signálu: schémata uložení pro ne každý pixel, např. 4:2:2 pro každý druhý, 4:1:1 každý čtvrtý na řádku, 4:2:0 průměr pro oblast 2×2 pixelů, 4:4:4 bez podvzorkování

Obrázek: Schémata YUV podvzorkování



– typy signálu:

- **kompozitní (baseband, RCA)** – kombinace jasové a barevných složek do jednoho signálu, pro redukci pásma, problém interference při slabé úrovni signálu (fluktuace jasu a barev, falešné barvy), na analogových zařízeních (koaxiální vstup/výstup)
- **S-video (Super, Y/C)** a **komponentní** – samostatné signály pro jasovou a barevné složky, na digitálních zařízeních

Formáty

- analogové: PAL, SECAM, NTSC

Obrázek: Analogové video formáty

- digitální (**Digital Video, DV**): komprimované CIF (Common Interchange Format), SIF, CCIR 601, HDTV (High Definition TV), Digital cinema
 - kompatibilita s analogovou TV – podpora prokládání
- oproti analogové TV vysoká rozlišení – **HDTV**: 720p = 1280×720 progresivní, 1080i/p = 1920×1080 prokládané/progresivní
- komprimované např. ITU H.26x , ISO MPEG x, HEVC, Theora, Google VPx, Windows Media Video (WMV)

Obrázek: Digitální video formáty

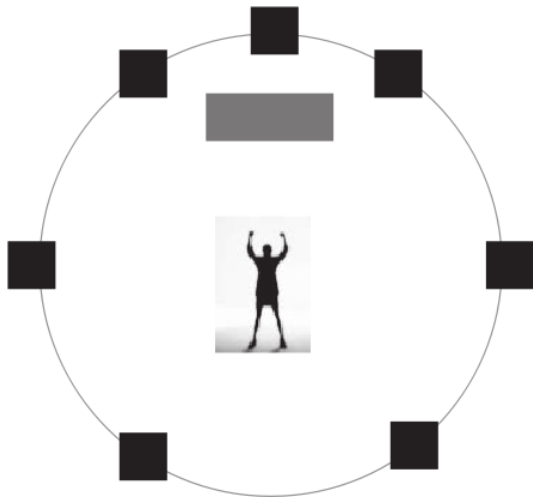
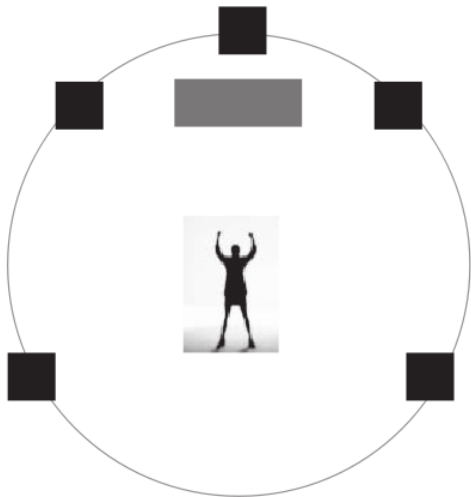
Property	NTSC	PAL	SECAM
Frame rate	30	25	25
Number of scan lines	525	625	625
Number of active lines	480	576	576
Aspect ratio	4:3	4:3	4:3
Color model	YIQ	YUV	YDbDr
Primary area of usage	North America (USA and Canada), Japan	Asia	France and Russia

Format name	Lines per frame	Pixels per line	Frames per second	Support for interlaced format	Subsampling scheme	Image aspect ratio
CIF	288	352		N	4:2:0	4:3
QCIF	144	176		N	4:2:0	4:3
SQCIF	96	128		N	4:2:0	4:3
4CIF	576	704		N	4:2:0	4:3
SIF-525	240	352	30	N	4:2:0	4:3
SIF-625	288	352	25	N	4:2:0	4:3
CCIR 601 NTSC (DV, DVB, DTV)	480	720	29.97	Y	4:2:2	4:3
CCIR 601 PAL/SECAM	576	720	25	Y	4:2:0	4:3
EDTV (576p)	480/576	720	29.97	N	4:2:0	4:3/16:9
HDTV (720p)	720	1280	59.94	N	4:2:0	16:9
HDTV (1080i)	1080	1920	29.97	Y	4:2:0	16:9
HDTV (1080p)	1080	1920	29.97	N	4:2:0	16:9
Digital cinema (2K)	1080	2048	24	N	4:4:4	47:20
Digital cinema (4K)	2160	4096	24	N	4:4:4	47:20

Reprezentace

- = posloupnost úrovní analogového zvukového signálu (**vzorků**)
- digitalizace (vzorkování + kvantizace) = **pulsní kódová modulace (PCM)**
- **frekvence vzorkování** v počtu vzorků/s (Hz), **velikost vzorku** (počet kvantizačních úrovní) v bitech, rozložení kvantizačních úrovní
- **dimenze (počet kanálů)**: mono (1), stereo (2, binaural, 1920), vícekanálové (> 2, stereophonic, 1940)
- **surround sound** = vícekanálový prostorový zvuk, posluchač obklopen reproduktory, např. 5.1 = přední středový (dialog), levý a pravý (soundtrack), boční/zadní levý a pravý (efekty), subwoofer (efekty na nízkých frekvencích, LFE, kdekoliv), 7.1 = zadní levý a pravý navíc

Obrázek: Rozmístění reproduktorů pro surround sound



- **spatial audio** („virtual surround sound“) = iluze vícekanálového prostorového zvuku s méně kanály (typicky dvěma), psychoakustické efekty pro simulaci lokalizace zvuku (posunutí kanálů vůči sobě využívající tzv. HRTF funkcí, dozvuků apod.)

Formáty

- nekomprimované např. WAV (Audio CD), G.711 (A/ μ -law, telefon, logaritmická kvantizace), G.729 (IP telefonie)
- komprimované např. MPEG 1 Layer III (MP3, psychoakustika pro kompresi), MPEG 4 Advanced Audio Codec (AAC), Dolby Digital (AC-3), Vorbis, Windows Media Audio (WMA), DTS, Speex (řeč), Opus, GSM 06.10 (AMR, mobily, komprese predikcí), Free Lossless Audio Codec (FLAC, bezztrátová)
- spolu s videem (a titulky apod.) uloženo pomocí tzv. **kontejnerů**, např. Audio Video Interleave (AVI), MPEG (PS, TS), MP4, Ogg, Matroska, WebM, QuickTime, FLV, 3GP (mobily)

Reprezentace

- = množina 2D a 3D **geometrických objektů** (body, čáry, křivky, polygony, plochy) popsaných **vektory**
- atributy délka, typ, tloušťka, barva, zakřivení atd. čáry, barva výplně apod.
 - neomezené rozlišení, ale pro zobrazení **rasterizování** = převod na rastrový obraz

Animace

- = posloupnost transformované grafiky zobrazované rychle po sobě
- **transformace** posunu, rotace, změna měřítka, zkosení (afinní) aj. – reprezentace vektorů a matic transformací pomocí homogenních souřadnic

Obrázek: Transformace pomocí homogenních souřadnic

Translations:

$$\begin{bmatrix} x_i + t_x \\ y_i + t_y \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix}$$

Rotations:

$$\begin{bmatrix} x_i \cos\alpha - y_i \sin\alpha \\ x_i \sin\alpha + y_i \cos\alpha \\ 1 \end{bmatrix} = \begin{bmatrix} \cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix}$$

Scaling:

$$\begin{bmatrix} s_x x_i \\ s_y y_i \\ 1 \end{bmatrix} = \begin{bmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \\ 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix}.$$

Formáty

- = vektorové – popis objektů, textové, binární nebo kombinované
- kombinované s rastrovým obrazem (tzv. **metafile**) – platformově nezávislé, např. Computer Graphics Metafile (CGM), WebCGM, Scalable Vector Graphics (SVG), (Encapsulated) Postscript ((E)PS, nekomprimovaný), Adobe Portable Document Format (PDF, komprimovaný)

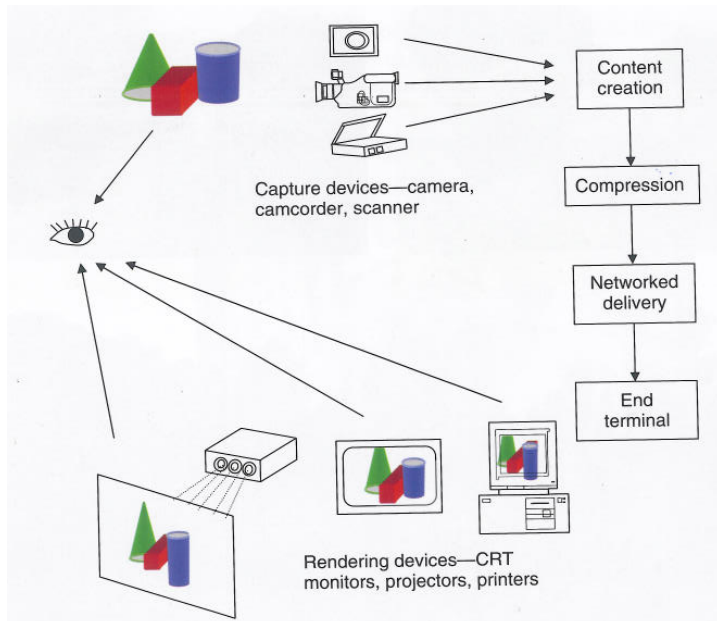


Barvy

- barvy jsou součástí obrazu, videa, grafiky
- = barvy zobrazené pomocí zobrazovacích zařízení (TV, počítače, projektory, tiskárny) by měly „**vypadat stejně**“ (= **být vnímány stejně**) jako barvy vzoru zaznamenaného pomocí záznamových zařízení (fotoaparáty, skenery, kamery), nezávisle na zařízeních

Obrázek: Ilustrace problému barev

- *co je barva?*: fyzikální veličina vs. smyslové vnímání okem interpretované mozkiem, **teorie barev** (17. st, Newton, separace světla na barvy duhy pomocí hranolu a složení zpět, později frekvence elektromagnetického záření)

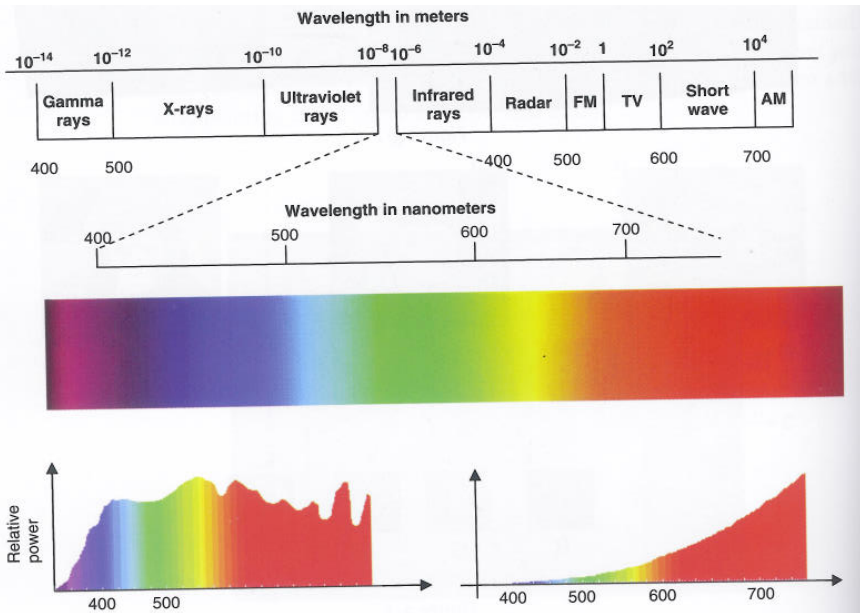


- = **kombinace frekvencí**, s různými amplitudami (energiemi), z viditelného spektra elektromagnetického vlnění (**světla**) s vlnovou délkou od 400 do 700 nm, < 400 = ultrafialové a > 700 = infračervené neviditelné
- různé zdroje světla mají různá **spektrální rozložení** (**spektrum**, „barva světla“), např. bílé (denní), žluté (žárovkové), červené (laser)

Obrázek: Viditelné spektrum světla a spektra zdrojů světla

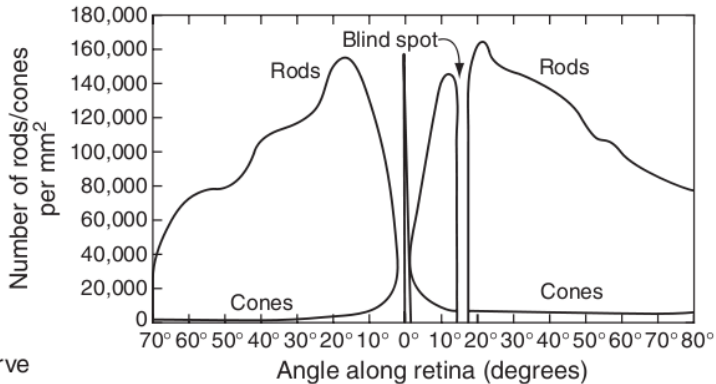
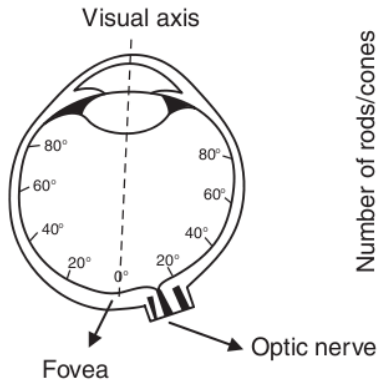
- **barva světla** = vnímaná kombinace frekvencí ze spektra světla
- interakce světla s materiály vzoru a obrazu – odrazy, absorpce, lomy → odražené světlo (kombinace frekvencí ze spektra) vnímané jako **barva materiálu**

Barva jako fyzikální veličina



- Young-Helmholtz, 1802, 1850, Maxwell
- = **trojité selektivní**: oko má tři druhy barevných receptorů (**trichromacita**), pro **červenou (R)**, **zelenou (G)** a **modrou (B) primární barvu** (absorpci spektra) – potvrzeno 1965:
- receptory oka (na sítnici): tyčinky pro jas (citlivější, větší počet, uniforměji rozložené) a čípky pro barvu (nejvíce u ohniska) – tři druhy, „modré“ (mimo ohnisko) nejcitlivější

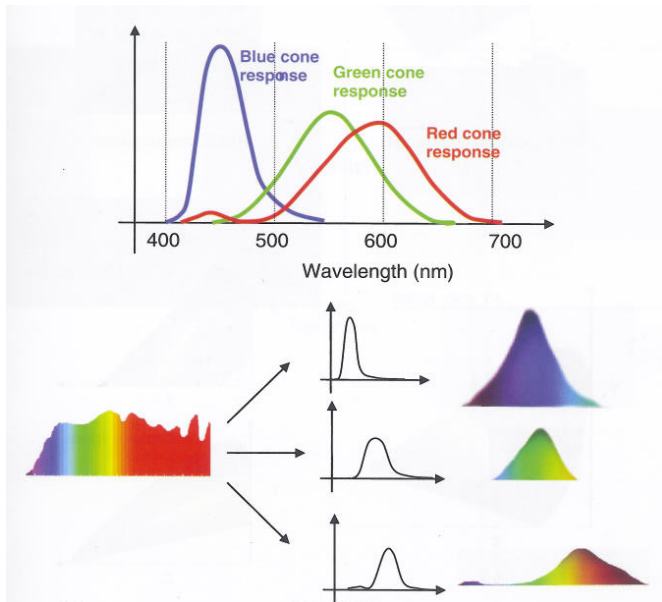
Obrázek: Sítnice oka a distribuce tyčinek a čípků



- citlivosti receptorů (tzv. **color-matching funkce**): funkce $s_i(\lambda)$, $i = 1, 2, 3$

Obrázek: Citlivost barevných receptorů oka

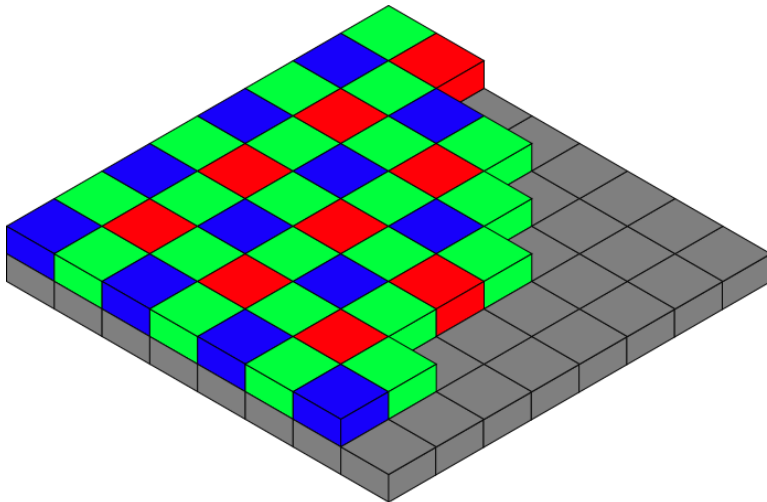
- vnímaná intenzita spektra: $c_i = \int_{\lambda=400}^{\lambda=700} s_i(\lambda) f(\lambda) d\lambda$, $f(\lambda)$ je spektrum světla, nebo digitalizované $c_i = \sum_{j=400}^{j=700} s_i(\lambda_j) f(\lambda_j)$, vektorově $c = S^T f$, kde S je matice se sloupci s_i
 - vnímaná barva = (c_1, c_2, c_3) – tzv. **tristimulus vektor**
 - **řešení problému barev**: pro stejné vnímání barev ze dvou spekter f a g , tj. $c_f = S^T f = S^T g = c_g$, nemusí být nutně $f = g$
- ⇒ jestliže záznamové zařízení zaznamenává f a zobrazovací zařízení zobrazuje g , g nemusí být rovno f , stačí takové, aby bylo $S^T f = S^T g \rightarrow$ **kalibrace zařízení**



- zaznamenaná barva (pixelu, digitalizovaná): $a = M^T f$, kde M je matice se sloupci m_i , charakterizující **odevzu barevných filtrů** pro primární barvy, více pro zelenou pro vyšší citlivost (řešené jako tzv. Bayer filtr)

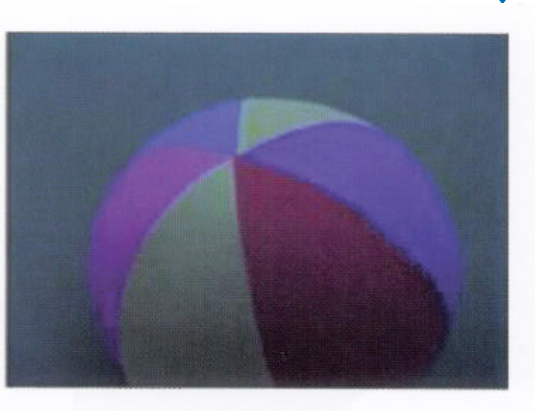
Obrázek: Bayer filtr

- zobrazené spektrum (pixelu): $g = vp$, kde p je vektor frekvencí **zobrazovaných primárních barev** a v je vektor zobrazované barvy
- stejné vnímání zaznamenané a zobrazované barvy: $S^T f = S^T g = S^T ap$,
 $M^T f = a = (S^T p)^{-1} S^T f$ – zobrazované primární barvy p a záznamové barevné filtry m_i musí korespondovat
- vztah není jednoznačně definovaný → CIE definované **standardní primární barvy**
 $p = (700, 546.1, 435.8)nm \Rightarrow$ výpočet m_i pro neznámou S – člověk ji „provádí“
- postupem **porovnávání vybraných skutečných a zobrazených barev** ze spektra f a úprav a lze vypočítat filtry m_i pro automatické úpravy a pro stejné vnímání ostatních zobrazovaných barev



- **gamma** CRT zobrazovacího zařízení: faktor (exponenciálního) zesílení (jasu) spektra zobrazované barvy – **gamma korekce**, $g = \left(\frac{v}{v_{max}}\right)^\gamma p$, např. standardní 2.2, LCD (a Plasmy) mají vyhledávací tabulku hodnot gamma pro primární barvy
- **zbývající problém**: různé světelné podmínky („barva“ světla) při záznamu a při zobrazování
- **stálost barev** = mozek automaticky kompenzuje „barvu“ světla (!) – např. bílý papír je vnímám bílý za různých světel, např. fotoaparát nebo kameru je potřeba nastavit \Rightarrow vnímání barev není lokální

Obrázek: Stálost barev



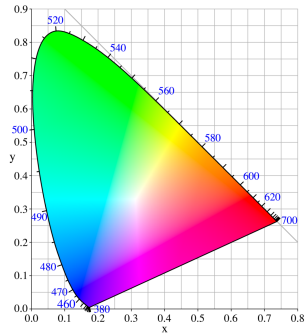
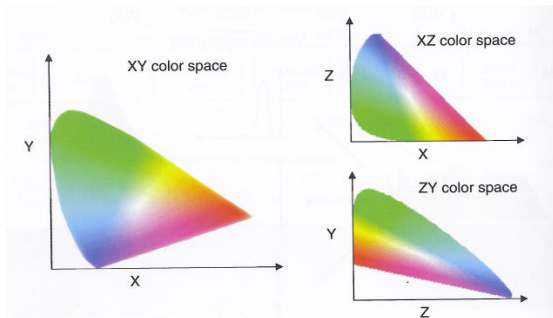
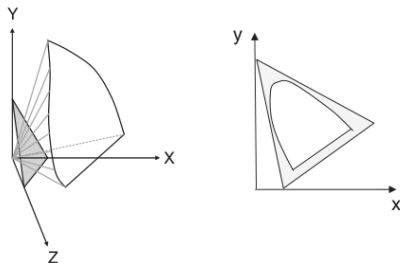


- = modely pro reprezentaci barev na základě složek (souřadnic) = **hodnot primárních barev**, typicky trojrozměrné
- **vektor barvy** = lineární kombinace vektorů hodnot složek
- oblast všech reprezentovatelných barev = **barevný gamut**

- základní, schopný reprezentovat **všechny barvy viditelného spektra** = **gamut vidění**
- **teoretický**: používaný pro studium jiných prostorů, jejich barevných rozsahů (gamutů), dominantních frekvencí barev, mixování a komplementárních barev apod.
- osy = projekce **tristimulus vektoru** na základě citlivosti receptorů oka do „barevných“ složek, osa Y pro jasovou složku barvy, osy nerealizovatelné
- odvozené prostory, např. nejpoužívanější **CIE xyY (xy)**: projekce prostoru XYZ na rovinu $X + Y + Z = 1$: $(x, y, z) = \left(\frac{X}{X+Y+Z}, \frac{Y}{X+Y+Z}, \frac{Z}{X+Y+Z}\right)$, osy x a y pro barevné složky barvy, tzv. **barevné (chromacity) diagramy**

Obrázek: Barevné prostory CIE XYZ a xyY

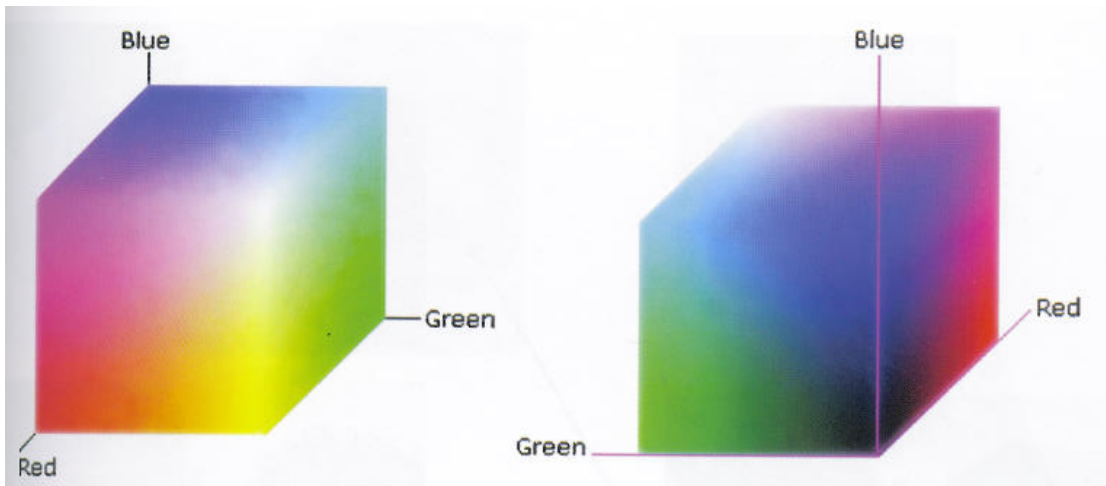
Barevný prostor CIE XYZ

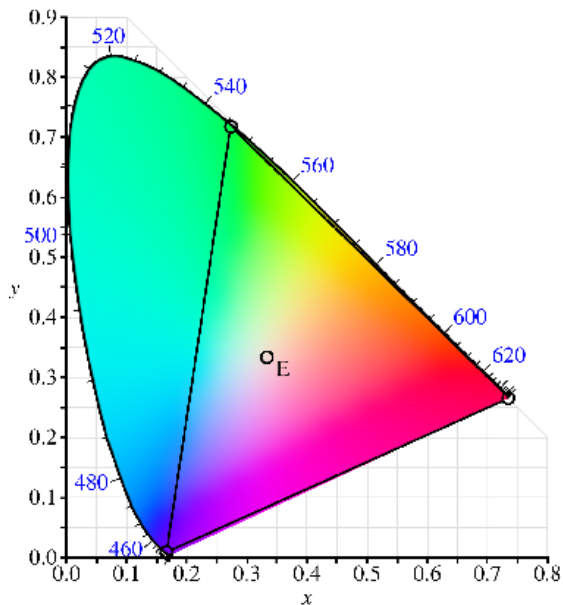


- osy = **primární barvy červená R 645.16 nm, zelená G 526.32 nm, modrá B 444.44 nm**
- normalizovaný do jednotkové „**RGB kostky**“
- **aditivní**: vektor barvy je součtem vektorů hodnot primárních barev, $(0, 0, 0) =$ černá, $(1, 1, 1) =$ bílá
- používáný záznamovými a zobrazovacími zařízeními – závislý na zařízení (nemají všechna stejná primární barvy)
- nereprezentuje všechny barvy viditelného spektra, zejména zelené

Obrázek: Barevný prostor RGB

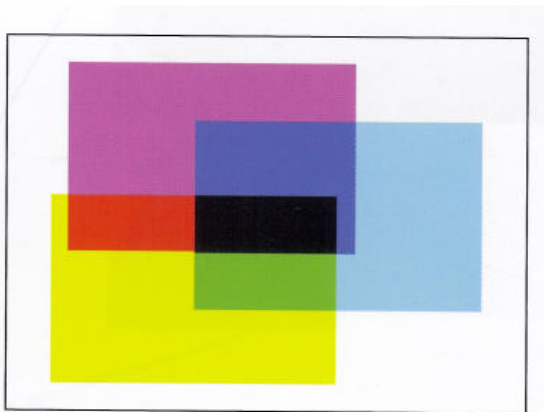
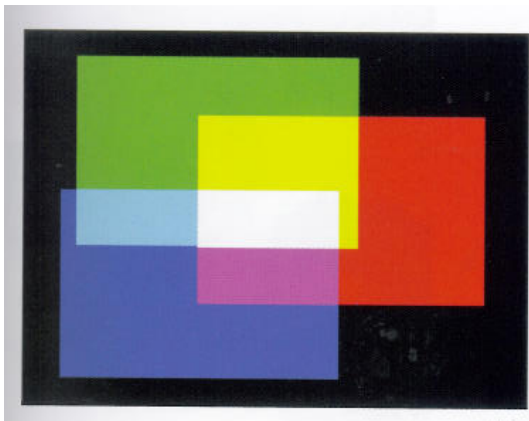
Barevný prostor RGB





- = komplementární k prostoru RGB – primární barvy **azurová C**, **fialová M**, **žlutá Y**
 - **subtraktivní** – vzhledem k prostoru RGB
 - používaný tiskovými zařízeními – tisk primárních barev (kanálů) postupně přes sebe
 - černá (1, 1, 1) – často používaná, neekonomická reprezentace, v praxi nikdy černá
- **CMYK** – navíc **černá K primární barva**, převod z CMY: $C_{CMY} = (C, M, Y)$, jestliže $\min(C, M, Y) = 1$, pak $C_{CMYK} = (0, 0, 0, 1)$, jinak $K = \min(C, M, Y)$ a $C_{CMYK} = (\frac{C-K}{1-K}, \frac{M-K}{1-K}, \frac{Y-K}{1-K}, K)$

Obrázek: Aditivní a subtraktivní slučování primárních barev v prostorech RGB a CMY





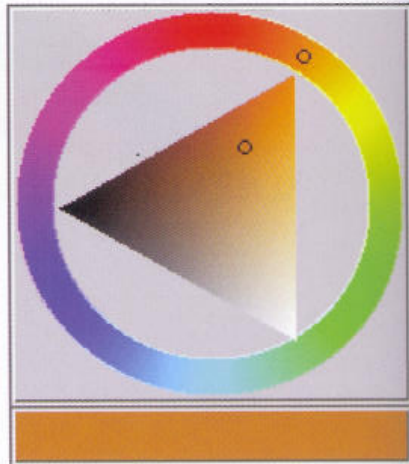
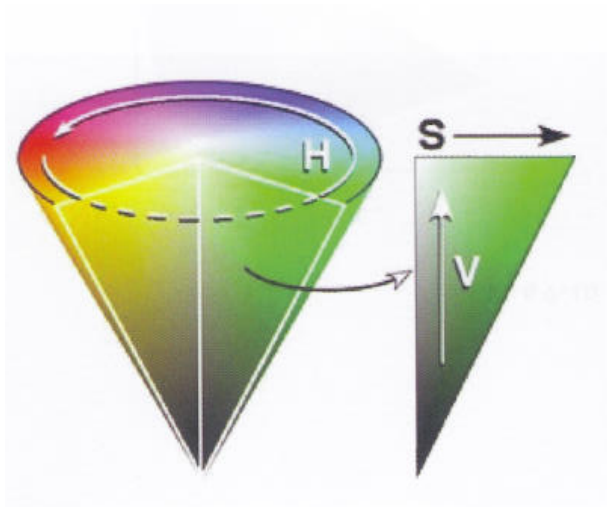
- osa Y pro **jasovou složku** barvy, osy U/I/Cb, V/Q/Cr pro **barevné složky** – šetření pásma s podvzorkováním barevných složek (redukce barevného rozlišení méně ovlivňuje vnímání obrazu než redukce jasového rozlišení)
- používaný u analogové TV (YUV v PAL a SECAM, YIQ v NTSC) a kompresních formátů JPEG a MPEG (YCbCr)
- převod z RGB:

$$T_{RGBtoYUV} = \begin{pmatrix} 0.299 & 0.587 & 0.114 \\ -0.147 & -0.289 & 0.436 \\ 0.615 & -0.515 & 0.312 \end{pmatrix}$$

- **nelineární**, použití v (digitálním) zpracování obrazu pro výběr a modifikaci barev – lépe zachycuje **lidské chápání barev** (sousední odstíny)
- **hue** ... **odstín barvy** (dominantní frekvence spektra)
- **saturation** ... **nasycení barvy** od bílé k plné barvě
- **brightness (lightness, value)** ... **světlost barvy** od černé k bílé

Obrázek: Barevný prostor HSV

Obrázek: Převod z barevného prostoru RGB do HSV

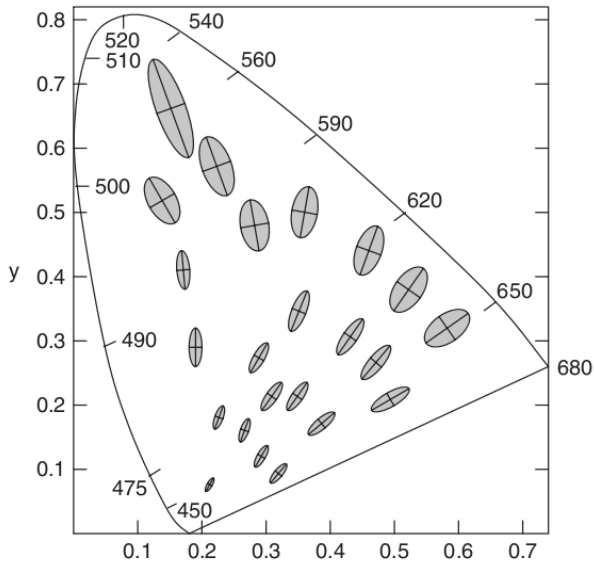


```
if (R > G) then
    Max = R; Min = G; position = 0;
else
    Max = G; Min = R; position = 1;
end
if (Max < B) then
    Max = B; position = 2
end
if (Min > B) then
    Min = B;
end
V = Max;
if (Max != 0) then
    S = (Max - Min)/Max ;
else
    S = 0;
end
if (S != 0) then
    if (position = 0) then
        H = (1+G-B)/(Max-Min);
    else if (position = 1) then
        H = (3+B-R)/(Max-Min);
    else
        H = (5+R-G)/(Max-Min) ;
    end
end
end
```


- **uniformní** = rozdíl souřadnic odráží vnímaný **rozdíl mezi barvami** – třídy vnímáním nerozlišitelných barev jsou zobrazeny jako vyplněné kruhy, narozdíl od různě natočených elips v CIE xy prostoru

Obrázek: Nerozlišitelné barvy v CIE xy prostoru (tzv. MacAdamovy elipsy, zvětšené)

- LUV neodráží rozdíl jasů, LAB ano
- pro zkoumání rozdílů barev tak, jak jsou vnímány člověkem





Authoring

- = **vytvoření (*interaktivního*) multimediálního obsahu** – získáním, úpravami a složením více typů médií dohromady
- vyžaduje znalost typů médií, manipulace s nimi (včetně programování/skriptování), umělecké cítění aj.
- náročné na čas: množství, velikost médií, požadavky na layout/strukturu (prostorové, časové), způsoby interakce aj.
- **authoring nástroje** – pro úpravu konkrétních typů médií nebo pro složení různých typů, přidání interaktivity, přizpůsobení pro distribuci a zobrazení (formát, velikost, layout/struktura – profily, např. MPEG4)
- **pasivní obsah** – uživatelská akce nemá efekt na (lineární) průběh, např. film/TV (video), hudba/rádio, kniha
- **aktivní (interaktivní) obsah** – (nelineární) průběh zvolený uživatelem, např. hra, DVD/Blue-ray disk, video-on-demand (streamovací) aplikace, fotoalbum, hypertextový dokument(s odkazy, např. webový) aj.

- získání (capturing), úpravy (post-processing) a uložení v různých formátech (formatting) pro jednotlivé typy médií – **intramedia (media) authoring**
- prostorová a časová organizace prvků různých typů médií = **layout** – **intermedia (multimedia) authoring**
- specifikace způsobů **interakce** mezi prvky = vytvoření (nelineárního) průběhu s možností volby – **intermedia (multimedia) authoring**
- ideálně podpora různých platform a způsobů interakce, obsah nezávislý na platformě, v praxi pro konkrétní platformu(y)

- nástroje pro jednotlivé typy médií, např. video-, audio-**editory**, editory obrázků, grafiky (animace), textový editor apod.
- získání a úpravy médií:
 - obraz: kreslení, ořezání, transformace, kompozice, filtrace, (prostorové) efekty, ...
 - video/animace: obraz/grafika + střih, změna rychlosti, skenování, (časoprostorové) efekty, ...
 - audio: generování zvuku, střih, změna rychlosti, kompozice filtrace, (časové) efekty, ...
 - grafika: tvorba objektů, ořezání, transformace (i barev, např. osvětlení), kompozice, **renderování** = převod z vektorové do rastrové grafiky (obrazu), ...
- nastavení parametrů formátu pro uložení, např. komprese

- **specifické nástroje** pro specifické formáty a použití, např. audio-video, DVD/Blue-ray, MPEG4, herní enginy
- kategorizace na základě prostorové a časové organizace prvků médií a specifikace interakce:
 - **prostorová organizace (layout)**: rozmístění prvků, kompozice v prostoru, např. vložení grafiky do videa – šablony, graf scény
 - **časová organizace**: rozmístění prvků, kompozice v čase, např. synchronizace videa a audia, dynamická kompozice – časové rozvrhy
 - **specifikace interakce**: umístění uživatelských ovládacích prvků pro události a specifikace jejich akce (navigace, např. „přehrávač“), programovatelné složené akce, např. interaktivní dynamické prostorové a časové organizace – multimediální aplikace, DVD/Blue-ray disk, interaktivní TV, hry, . . . , webová stránka

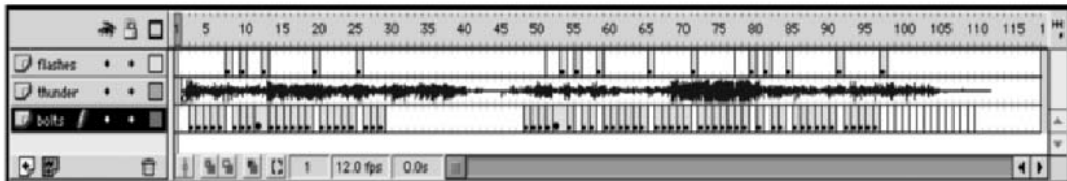
- **časová osa (timeline)** – osnovy, kanály, „přehrávač“, pro úpravy video a audio médií a časovou organizaci („režie“)

Obrázek: Timeline

- **skriptování** – skriptovací jazyk pro programování akcí „za hranicemi UI“ (dynamické organizace, interaktivita), automatizace, např. JavaScript, SMIL (streaming), VRML (3D grafika), XMT (MPEG4)

Obrázek: SMIL skript

- další, např. flow control (diagram akcí mezi prvky médií), karty (prostorová organizace, např. u prezentací) aj.



```
<smil>
  <head>
    <layout>
      <region id="Region1" top="5" left="20" />
      <region id="Region2" top="100" left="20" />
    </layout>
  </head>
  <body>
    <image region="Region1"
      src="http://www.mysite.org/logo.jpg"
      dur="10s" />
    <video region="Region2"
      src="http://www.mysite.org/myvideo.avi"
      clipBegin="10" clipEnd="20">
    </video>
  </body>
</smil>
```

- jednoduchá, komfortní interakce pro jednotlivé typy médií a multimédií
- vizuální/grafická média ⇒ **grafické UI** (z PC desktopu), např. tlačítka (pro „**přehrávač**“), ikony, dialogy (výběr TV kanálu/streamu)
- prostorová omezení zařízení (nejen displeje), jiné (dotykové) ovládání: např. mobilu, tabletu → upravené UI, responzivita
- spolupráce (přenosných) propojených a synchronizovaných zařízení sdílejících obsah



- **nezávislost na zařízení** = stejně nebo podobně koncipovaný obsah na různých zařízeních, „**univerzální přístup**“ (sít, HW/SW platforma, lokace, jazyk apod.) – univerzální formáty definující možnosti zařízení, např. HTML, více různě velkých forem obsahu pro distribuci → **profily** (MPEG4)
- **distribuovaný authoring** = současná spolupráce více lidí na obsahu, použití upravených systémů na správu verzí z vývoje SW (využívající mediální prostoročasové sémantiky dat) – nároky na síť (velký objem, rychlá odezva), řešení konfliktů (zamykání objektů), náročná organizace (např. real-time), např. WebDAV pro web
- přidružené služby – např. **(automatická) správa obsahu** = prohlížení, vyhledávání, výběr, vytváření přizpůsobeného, aktualizace, sdílení – databáze, indexace, metadata, archivy, různé distribuce, např. Google News



Kompresa

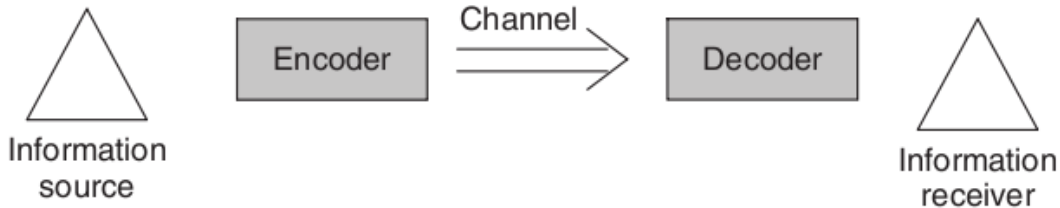


- velké objemy dat, potřeba uložit a distribuovat po různých sítích
- dříve omezené kapacity a rychlosti, dnes větší objemy – např. video HDTV 1080p 30 fps 12 bpp (YUV 4:2:0) = 712 Mb/s, obraz 10 Mpx 24bpp (RGB) ~ 30 MB
- komprese = **redukce dat při zachování (přijatelné úrovně) obsažené informace** – využití uspořádání a vzorů (a redundancí na přijatelnou úroveň) v multimediálních datech

- Claude Shannon, 1940s
- = **efektivita a spolehlivost přenosu informace ze zdroje k cíli daným (chybovým) komunikačním kanálem** – **minimální objem dat** (pomocí komprese) a zabezpečení správnosti dat (pomocí přidání kontrolních dat pro detekci a opravy chyb)
- **enkoder** a **dekoder**, produkované informační **symboly** = pravděpodobnostní proměnné, bez sémantiky

Obrázek: Shannonův komunikační model

- data = posloupnost bitů s hodnotami 0 a 1 (digitální binární forma)



- **abeceda** = neprázdná konečná množina symbolů, $S = \{s_1, \dots, s_n\}$, např. 256 hodnot v 8 b, 26 znaků abecedy aj.
- **kódování (kód)** = zobrazení symbolů na posloupnosti bitů (kódová slova pro symboly), $C = \{\langle s_1, c_1 \rangle, \dots, \langle s_n, c_n \rangle\}$, např. ASCII tabulka
- **zpráva** = konečná posloupnost symbolů z abecedy, např. $s_1 s_2 s_1 s_4$, reprezentována proudem bitů
- **informace (informační obsah)** zprávy \sim počet a uspořádání symbolů ve zprávě – důležitá frekvence výskytů symbolů
- **pravděpodobnost (výskytu) symbolu**: $p_i = \frac{f_i}{|M|}$, kde f_i je počet (frekvence) výskytů symbolu s_i ve zprávě M délky $|M|$ – použita pro optimální kódování symbolů/zprávy s méně bity pro pravděpodobnější symboly – kódy **proměnné délky** l_i

Obrázek: Kódování symbolů a zprávy

- **jednoznačně dekódovatelný kód** = v proudu bitů je možné jednoznačně rozpoznat kódová slova symbolů, např. první 2 kódy z předchozího obrázku
- **prefixový kód** = žádné kódové slovo není prefixem jiného
- **Kraftova-McMillanova věta** – prefixové kódy jsou dostačující

Symbol	Code	Number of occurrences (probability)	Bits used by each symbol
s_1	00	70 (0.7)	$70 \times 2 = 140$
s_2	01	5 (0.05)	$5 \times 2 = 10$
s_3	10	20 (0.2)	$20 \times 2 = 40$
s_4	11	5 (0.05)	$5 \times 2 = 10$

200

Symbol	Code	Number of occurrences (probability)	Bits used by each symbol
s_1	1	70 (0.7)	$70 \times 1 = 70$
s_2	001	5 (0.05)	$5 \times 3 = 15$
s_3	01	20 (0.2)	$20 \times 2 = 40$
s_4	000	5 (0.05)	$5 \times 3 = 15$

140

Symbol	Code	Number of occurrences (probability)	Bits used by each symbol
s_1	0	70 (0.7)	$70 \times 1 = 70$
s_2	01	5 (0.05)	$5 \times 2 = 10$
s_3	1	20 (0.2)	$20 \times 1 = 20$
s_4	10	5 (0.05)	$5 \times 2 = 10$

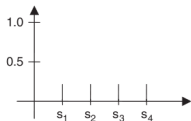
110

- délka zprávy v počtu symbolů: $|M| = \sum_{i=1}^n f_i$
- délka zprávy v bitech: $l_M = \sum_{i=1}^n f_i l_i$
- **průměrná délka symbolu** v bitech: $\lambda_M = \frac{l_M}{|M|} = \sum_{i=1}^n p_i l_i$
- **cíl komprese = nalézt prefixový kód takový, že λ_M je minimální** – tj. když p_i roste, l_i musí klesat
- **entropie**: $E_M = -\sum_{i=1}^n p_i \log_2 p_i$, teoreticky minimální λ_M (Shannon), slouží k **vyjádření míry informace ve zprávě M** , nejvyšší při stejné pravděpodobnosti symbolů – např. při náhodném výskytu symbolů
 - $\log_2 \frac{1}{p_i} =$ vlastní informace symbolu s_i , reprezentuje l_i
 - teoretický limit komprese

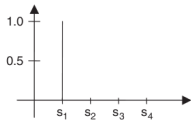
Obrázek: Entropie

- **efektivita** (kódu): $\frac{E_M}{\lambda_M}$, míra „účinnosti“ komprese, nejlepší = 1

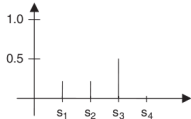
$$P_i = \{0.25, 0.25, 0.25, 0.25\}$$
$$H = -(4 \times 0.25 \times \log_2 0.25)$$
$$H = 2$$



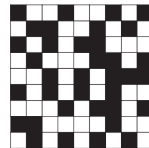
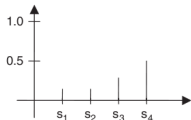
$$P_i = \{1.0, 0.0, 0.0, 0.0\}$$
$$H = -(1 \times \log_2 1)$$
$$H = 0$$



$$P_i = \{0.25, 0.25, 0.5, 0.0\}$$
$$H = -(2 \times 0.25 \times \log_2 0.25 + 0.5 \times \log_2 0.5)$$
$$H = 1.5$$



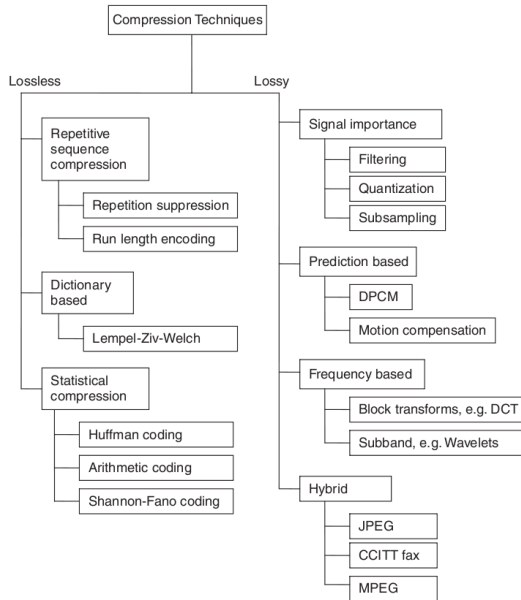
$$P_i = \{0.125, 0.125, 0.25, 0.5\}$$
$$H = -(2 \times 0.125 \times \log_2 0.125 + 0.25 \times \log_2 0.25 + 0.5 \times \log_2 0.5)$$
$$H = 1.75$$



- **bezeztrátová (lossless)** = po dekompresi stejná data, bez ztráty informace
- **ztrátová (lossy)** = po dekompresi jiná data (podobný signál), změna informace, **zkreslení dat (signálu)**
- používané samostatně i v kombinaci v metodách komprese multimediálních dat

Obrázek: Taxonomie kompresních metod

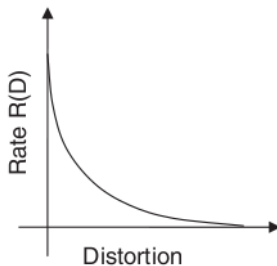
- **compression (bit) rate** = komprimovaná velikost dat v bitech na symbol/vzorek/pixel/sekundu, použití u přenosu dat
 - **proměnná** u bezztrátové komprese (proměnná délka kódových slov symbolů), udržována **konstantní** u ztrátové komprese (proměnným zkreslením dat) – výhodnější pro úpravy a (real-time) přenos
- **compression ratio (kompresní poměr)** = poměr velikosti (rate) původních a komprimovaných dat, použití u uložení dat



- = vztah mezi compression (bit) rate a mírou zkreslení (změny informace) u ztrátové komprese – nepřímá úměra, použití pro hodnocení efektivity metod
- **míra zkreslení** $f(y - \hat{y})$: typicky měření rozdílů mezi původními (y) a dekomprimovanými (\hat{y}) daty
 - např. součet rozdílů hodnot pixelů, populární střední kvadratická chyba (**mean square error**) nebo poměr (špičkového) signálu k šumu (**(peak) signal-to-noise ratio, (P)SNR**)
 - čím vyšší (a vyšší kompresní poměr), tím menší bit rate a naopak → komprese = kompromis

Obrázek: Rate distortion a míry zkreslení

- nemusí odpovídat **vnímanému zkreslení**, např. posunutí obrazu o pixel



$$\sigma_d^2 = \frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2$$

$$SNR = \frac{\sigma_{\hat{y}}^2}{\sigma_d^2}, \quad SNR(db) = 10 \times \log\left(\frac{\sigma_{\hat{y}}^2}{\sigma_d^2}\right)$$

Where $\sigma_{\hat{y}}$ is the averaged squared value of the output

$$PSNR(db) = 10 \times \log\left(\frac{\sigma_{peak}^2}{\sigma_d^2}\right)$$

Where σ_{peak} is the squared value of the peak output



- = odstranění **statistické redundance** v datech
- kratší kódová slova pro častější (pravděpodobnější) symboly a naopak
- **statistické metody** – použití **pravděpodobnostních modelů** = (odhadů) pravděpodobnosti výskytů symbolů dat spočítaných před (**semi-adaptivní**) nebo během (**adaptivní**) komprese nebo empiricky zjištěných (**statický**)
 - semi-adaptivní model je potřeba uložit s komprimovanými daty
- **slovníkové metody** – použití **slovníku** = seznamu dříve se v datech vyskytujících řetězců symbolů vytvořeného typicky během komprese

Run Length Encoding (RLE)

- = kódování posloupností opakujících se symbolů dvojicemi počet opakování a symbol (nebo novými symboly)
- detaily: kódování počtu opakování, nekódování posloupností kratších než dvojice, aj.
- použití ve standardech pro kompresi obrazu/grafiky a audia

- = kódování dříve se vyskytujících řetězců symbolů, vzorů, ukládaných do **slovníku**, pozicemi ve slovníku
- vytváření slovníku při kompresi i dekompresi
- (explicitně) nevyužívají pravděpodobností výskytu symbolů
- Lempel, Ziv, 1970s: **metody LZ***

LZW

- uložení samostatných symbolů do slovníku
- komprese = opakování kódování nejdelšího řetězce na vstupu, který je ve slovníku, a uložení spojení řetězce a dalšího symbolu na vstupu do slovníku
- dekomprese = opakování uložení spojení předchozího dekódovaného řetězce a prvního znaku aktuálně, popř. předchozího, dekódovaného řetězce do slovníku a dekódování kódu

Obrázek: LZW komprese

- detaily: omezená velikost slovníku (velký = lepší komprese, ale delší kódy a doba vyhledávání, v praxi 4096), vyprázdnění slovníku
- použití např. v obrazovém formátu GIF

x x y y x y x y x x y y y x y x x y x x y y x

0 0 1 1 3 6 3 4 7 5 8 0

Iteration number	Current length symbol(s)	Dictionary code used	Next symbol in string	New dictionary code generated	New dictionary index generated
				x	0
				y	1
1	x	0	x	xx	2
2	x	0	y	xy	3
3	y	1	y	yy	4
4	y	1	x	yx	5
5	xy	3	x	xyx	6
6	xyx	6	x	xyxx	7
7	xy	3	y	xyy	8
8	yy	4	x	yyx	9
9	xyxx	7	y	xyxy	10
10	yx	5	x	yxx	11
11	xyy	8	x	xyyx	12
12	x	0	END		

Index	Entry	Index	Entry
0	x	7	xyxx
1	y	8	xyy
2	xx	9	yyx
3	xy	10	xyxy
4	yy	11	yxx
5	yx	12	xyyx
6	xyx		

- = vytvoření optimálního prefixového kódu s kratšími kódovými slovy pro pravděpodobnější symboly a naopak
- Huffman, 1952
- (Huffmanovy) kódy symbolů = posloupnosti označení 0 a 1 hran na cestách binárním stromem od kořene k listům označených symboly
- konstrukce **(Huffmanova) stromu** – pravděpodobnější symboly blíže kořenu:
 - 1 listy (setříděně) označeny pravděpodobnostmi výskytu symbolů
 - 2 opakování vytvoření rodiče pro 2 uzly označené minimálními pravděpodobnostmi a označení rodiče součtem pravděpodobností a hran k rodiči 0 a 1

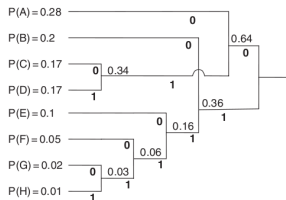
Obrázek: Huffmanovo kódování

Symbol	Probability	Binary code	Code length
A	0.28	000	3
B	0.2	001	3
C	0.17	010	3
D	0.17	011	3
E	0.1	100	3
F	0.05	101	3
G	0.02	110	3
H	0.01	111	3

Average symbol length = 3

Entropy = 2.574

Efficiency = 0.858



Symbol	Probability	Huffman code	Code length
A	0.28	00	1
B	0.2	10	3
C	0.17	010	3
D	0.17	011	3
E	0.1	110	3
F	0.05	1110	4
G	0.02	11110	5
H	0.01	11111	5

Average symbol length = 2.63

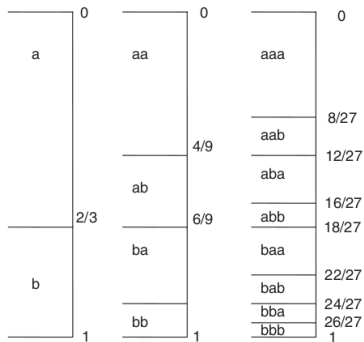
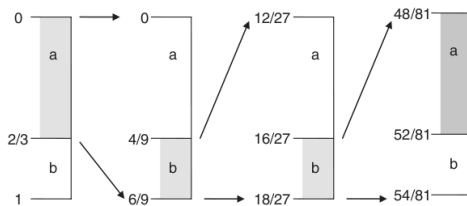
Entropy = 2.574

Efficiency = 0.9792

- **kódování řetězce symbolů** jako celku místo jednotlivých symbolů (kdy je potřeba celý počet bitů pro kódové slovo symbolu)
- = kódování řetězce symbolů do dvojice minimální binární reprezentace racionálního čísla v **podintervalu intervalu** $[0, 1)$ a délky řetězce
- konstrukce podintervalu:
 - 1 interval nastaven na $[0, 1)$
 - 2 opakování rozkladu intervalu na podintervaly s délkami proporčně odpovídajícími pravděpodobnostem výskytu pevně uspořádaných symbolů a nastavení intervalu na podinterval pro symbol na vstupu

Obrázek: Aritmetické kódování, $[0.10010111, 0.10100100]$, $0.101 = \frac{50}{80}$, 101

- detaily: zvětšení a posunutí malého podintervalu kolem 0.5, celočíselná aritmetika, rozdělení dat na bloky a kódování bloků



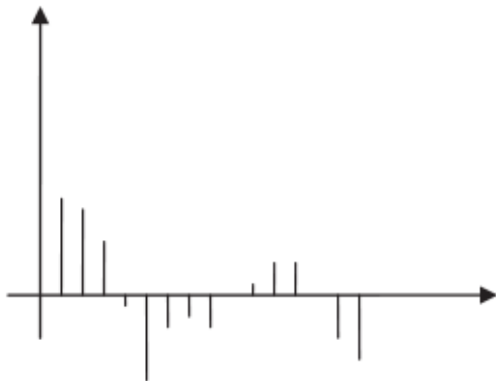
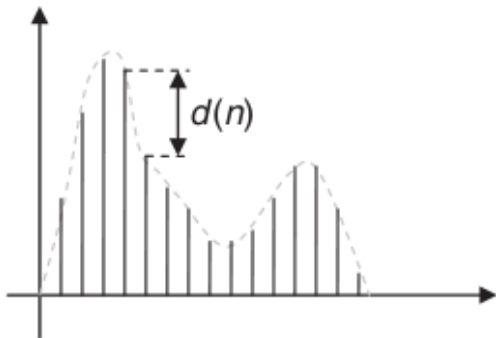


- = **změna („ztráta“)** informace a **zkreslení** (při vnímání) dekomprimovaných dat pro dosažení vyšší komprese, za teoretický limit bezztrátové
- určitá míra zkreslení je lidskými smysly nerozpoznatelná a dále pak (v závislosti na aplikaci) akceptovatelná
- kvantizace je ztrátová sama o sobě

- signál o nízké frekvenci se mění (v čase) pomalu → vzorky mají blízké hodnoty ⇒ mohou být **predikovány z předchozích hodnot** – metody založené na predikci (**prediction-based**)
- nejjednodušší predikce = přímo předchozí hodnota
- = kódování **chyby predikce (prediction error)** = rozdílu mezi predikovanou a skutečnou hodnotou – má menší rozsah než signál samotný a tedy menší entropii = méně bitů pro reprezentaci – kvantizaci

Obrázek: DPCM

Diferenční PCM (DPCM)



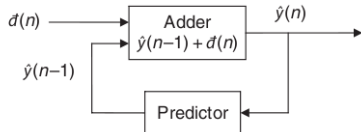
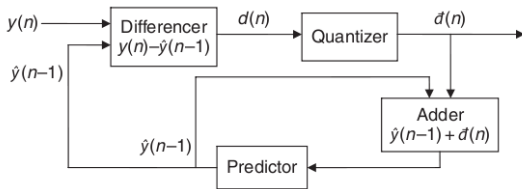
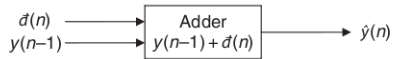
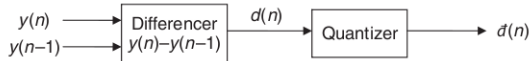
- **open loop případ** = na výstupu enkoderu kvantizovaný rozdíl mezi z předchozích hodnot predikovanou a skutečnou hodnotou signálu – **problém**: dekoder nemá k dispozici přesné předchozí hodnoty, ale pouze zkrslené, o kvantizační chybu, která se ovšem akumuluje

Obrázek: Open loop případ

- **closed loop případ** = na výstupu enkoderu kvantizovaný rozdíl mezi z **dekódovaných** (zkrslených) předchozích hodnot predikovanou a skutečnou hodnotou signálu – redukce akumulace kvantizační chyby, dekoder jako součást enkoderu

Obrázek: Closed loop případ

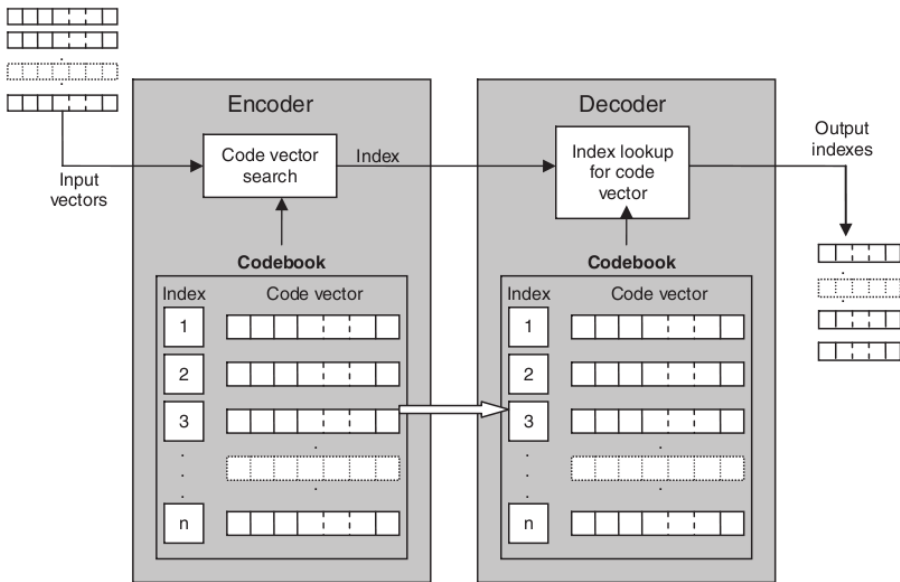
Diferenční PCM (DPCM)



- skalární kvantizace = na jednotlivých vzorcích – jednoduchá, rychlá, ale nezohledňuje korelace mezi (sousedními) vzorky
- = kvantizace **skupin (vektorů) vzorků** – zohlednění korelací → lepší komprese
 - vytvoření **slovníku vektorů**: statistická analýza signálu nebo empiricky (trénováním), nejlepší reprezentanti s maximálním tolerovaným zkreslením, otázka velikosti slovníku (větší = menší zkreslení, ale více místa, např. pro obraz 128–2048) a vektorů (větší = více korelací)
 - kódování vektorů na vstupu pozicemi vektorů ve slovníku s minimálním zkreslením (měřeným např. střední kvadratickou chybou)

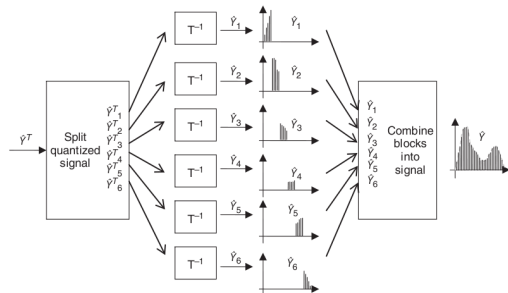
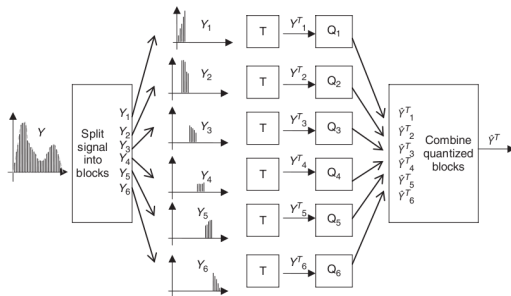
Obrázek: Vektorová kvantizace

- např. barevné obrazy – pixel má barvu reprezentovanou hodnotami v barevných kanálech = vektor, výběr barev = slovník (paleta barev) – tzv. **indexované barvy**



- = (invertibilní) transformace signálu z domény vzorků do jiné domény s nižší entropií signálu a kvantizace (ve vhodném pořadí prvků domény)
- **frekvenční** – do frekvenční domény, reprezentace koeficienty základních frekvencí, např. diskrétní **Fourierova/kosinová**, Hadamard, Lapped Orthogonal aj.
- **statistické** – do (méně nebo) nekorelovaných dimenzí, např. pomocí **analýzy hlavních komponent (PCA)**, např. Karhunen-Loeve
- **podpásmové (subband)** – do vícespektrální frekvenční domény
- transformace **bloků signálu**, např. 8x8 pixelů u JPEG komprese, a nezávislé kvantizace (s různými rozsahy) – pro kontrolu bit rate v závislosti na entropii

Obrázek: Blokově transformační kódování



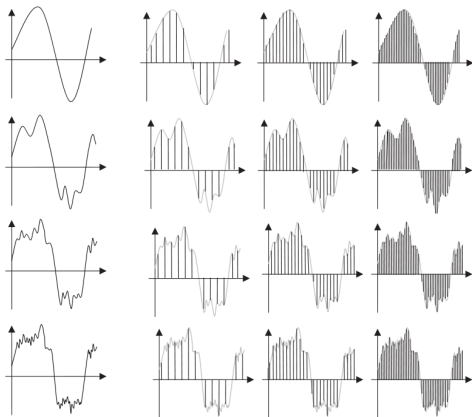
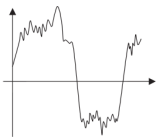
- **postupné jemnější aproximace/rozlišení** (spektrálního) popisu **signálu** rozděleného do **(pod)pásem**, s přidáváním vyšších frekvencí (tj. detailů) v rámci aproximace

Obrázek: Pásmový popis signálu

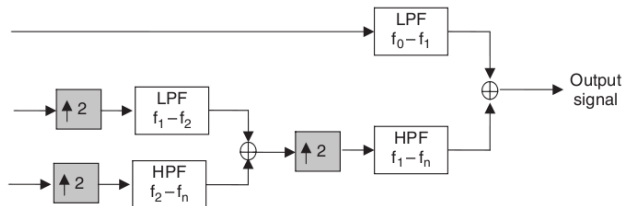
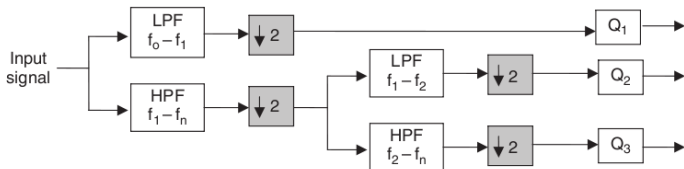
- = nezávislé (transformační) kódování signálů v pásmech v závislosti na významu pásma (ve vnímání) a výšce frekvencí pásma (určuje frekvenci vzorkování), např. u audia
- rozložení signálu do pásem – fitrace, např. band-pass pomocí **filtrových bank**
- **waveletová (víceúrovňová pásmová)** transformace/komprese = rozložení signálu do nízkofrekvenčního a vysokofrekvenčního pásma, poloviční podvzorkování, rekurzivní opakování na vysokofrekvenční, poté (transformační) kódování signálů (frekvenčních koeficientů) v pásmech

Obrázek: Waveletová transformace, LPF \leftrightarrow HPF

Podpásmové (subband) kódování

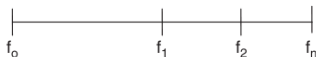


Podpásmové (subband) kódování



$\downarrow 2$ Downsampling by 2

$\uparrow 2$ Upsampling by 2





- = kombinace ztrátových a bezztrátových metod (v tomto pořadí!) – **ztrátové** (kvantizace) **snižují entropii**
- např. JPEG komprese: DCT bloků obrazu a DPCM kvantizovaných frekvenčních koeficientů, pak RLE a Huffmanovo kódování



- **rychlost a (časová/prostorová) složitost** kódování – real-time (horší) vs. offline (lepší komprese, rychlost obráceně), vliv množství signálu dostupného před kompresí (lepší komprese, vyšší složitost)
- **omezení bit rate** – variabilní pro uložení, konstantní pro úpravy a přenos → spolupráce ztrátové a bezztrátové části pro minimální zkreslení při omezené bit rate (**rate control**, rate distortion teorie – aktivní výzkum)
- **symetrické vs. asymetrické** komprese – enkoder a dekoder zhruba stejně vs. různé složitě/rychlé, u multimédií přirozeně asymetrické se složitějším (a lepším) enkoderem (offline) a rychlejším dekoderem (real-time), opačně např. u zálohování
- **adaptivní vs. neadaptivní** komprese – adaptace kompresních technik (modelu dat) podle předchozích dat, neadaptivní s empiricky zjištěným modelem dat, semi-adaptivní s analytickým statistickým průchodem před kompresí (**víceprůchodové komprese**)

Kompresa obrazu

- obrazová média (obraz, video, grafika po rasterizaci) všudypřítomná, nárůstající na rozlišení a objemu → **potřeba komprese**
- např. 4288×3216 24bpp (foto) ~ 40 MB \rightsquigarrow 4 MB (JPEG)
- např. formáty (ISO) JPEG pro foto a video, GIF/PNG pro obrázky na internetu
- software pro práci s komprimovaným obrazem
- bezetrátové (GIF, PNG) i ztrátové, typicky hybridní kompresní metody (JPEG)

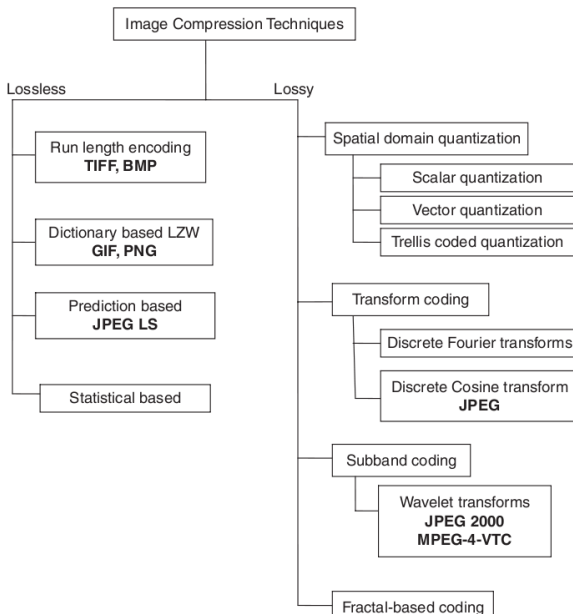


- **irelevance**: vizuální (parametry obrazu jsou **nad možnosti zobrazení nebo vnímání** za daných podmínek), aplikační (některé části obrazu nepotřebné)
- ztrátová komprese – neznatelné nebo přijatelné zkreslení, redukce entropie
- **redundance**: statistická, **barvy (sousedních) pixelů jsou (lokálně) korelované** – objekty, oblasti = „podobné“ pixely, postupné změny
 - **prostorová** = „podobné“ pixely v lokálních oblastech
 - **spektrální** = málo dominantních frekvencí
- bezztrátová komprese



- bezztrátové: využití prostorových redundancí
- ztrátové/hybridní: kvantizace v prostorové nebo frekvenční doméně (po transformaci) a bezztrátové statistické kódování

Obrázek: Taxonomie metod komprese obrazu



- omezená úroveň komprese (entropie)
- využití pro **zachování** (maximální, původní) **kvality**, např. u specifických fotografií (lékařských, astronomických apod., ztráty jsou nepříjemné) nebo ve filmovém průmyslu
- převod 2D pole pixelů na posloupnost hodnot pomocí (progresivního) **skenování**
- RLE: formáty např. **BMP**, **TIFF** (rozšiřitelný se specifikací komprese), RLA, PICT
- slovníkové (LZW): např. **GIF**, **PNG**, komprese **indexovaných barev** místo barev pixelů (po předchozím výběru optimální palety barev dané velikosti = vektorové kvantizaci, např. 8b)
- založené na predikci (DPCM): predikce barev sousedních pixelů, statistická komprese rozdílů, např. **JPEG LS** (lossless mode) – predikce ze sousedního levého, horního nebo/a levého horního pixelu → chybový obraz s menší entropií

Obrázek: Predikce v JPEG LS (lossless mode)

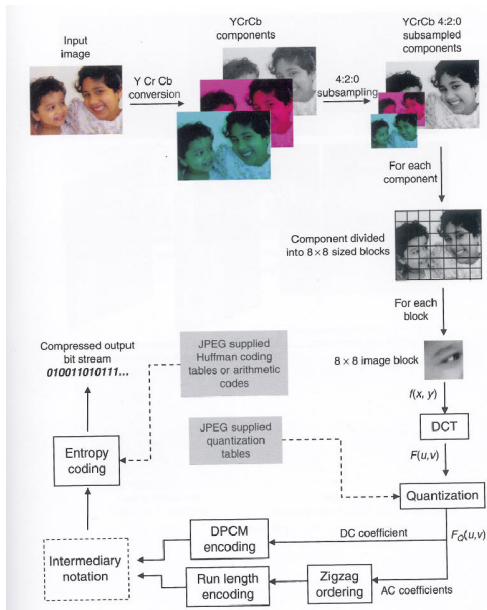
	C	B		
	A	X		

Prediction index	Prediction
0	No prediction
1	A
2	B
3	C
4	$A + B - C$
5	$A + ((B - C)/2)$
6	$B + ((A - C)/2)$
7	$(A + B)/2$

- ISO standard, 1993
- široce rozšířený, volné implementace a použití
- použití **diskrétní kosinové transformace (DCT)** – popis ve frekvenční doméně, hardwarová implementace

Obrázek: JPEG komprese (baseline mode)

- 1 transformace barev do **barevného prostoru YCbCr** (podobný YUV)
- 2 **podvzorkování** ve schématu 4:2:0
- 3 každý barevný kanál zpracováván samostatně
- 4 **rozklad** obrazu do **bloků 8×8 pixelů** (velikost zvolena experimentálně jako optimální pro využití prostorových a frekvenčních korelací) – příp. rozšíření obrazu o nulové okraje
- 5 každý blok zpracován samostatně, zpracování bloků v pořadí daném (progresivním) skenováním – v **baseline mode**



- 6 **DCT**: výpočet frekvenčních koeficientů $F(u, v)$ ze vzorků $f(x, y)$, první, největší, koeficient na pozici $\langle 0, 0 \rangle$ (nejnižší frekvence) = **DC koeficient**, ostatní = **AC koeficienty**

$$f(x, y) = \frac{1}{4} \left[\sum_{u=0}^{u=7} \sum_{v=0}^{v=7} C(u)C(v) \times F(u, v) \times \cos\left(\frac{(2x+1)u\pi}{16}\right) \cos\left(\frac{(2y+1)v\pi}{16}\right) \right]$$

$$F(u, v) = \frac{1}{4} C(u)C(v) \left[\sum_{x=0}^{x=7} \sum_{y=0}^{y=7} f(x, y) \times \cos\left(\frac{(2x+1)u\pi}{16}\right) \cos\left(\frac{(2y+1)v\pi}{16}\right) \right]$$

$$C(u), C(v) = \frac{1}{\sqrt{2}} \text{ pro } u, v = 0, \text{ jinak } 1$$

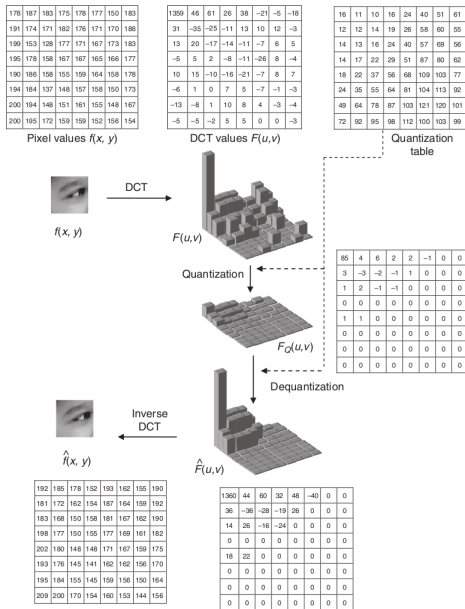
- 7 kvantizace** koeficientů $F(u, v)$ na $F_Q(u, v)$: s využitím **kvantizační tabulky** Q (vyplněné experimentálně, kdy **nízké frekvence** jsou v obrazech **dominantní** a vnímání ztráty na nich je citlivější \rightarrow větší rozsah) – **kvantizační faktory** $Q(u, v)$ udávají kvantizační rozsah (menší hodnota = větší rozsah), **nastavení nižší „kvality“** komprese u software **zvětšuje faktory**, skoro všechny pro vyšší frekvence vysoké \rightarrow ztráta původních hodnot koeficientů po dekvantizaci

$$F_Q(u, v) = \left\lceil \frac{F(u, v)}{Q(u, v)} \right\rceil$$

Obrázek: DCT a kvantizace bloku

- 8** DC koeficient první, cikcak seřazení AC koeficientů – ke konci řady řetězce 0 = nižší entropie

Obrázek: Cikcak seřazení AC koeficientů



- 9 **DPCM** DC koeficientů bloků – dočasná reprezentace jako dvojice velikost (v bitech) kódu rozdílu a rozdíl koeficientů pro sousední bloky
- 10 (varianta) **RLE** AC koeficientů bloku – kódovány pouze nenulové koeficienty, dočasná reprezentace jako dvojice dvojice počet předchozích nulových koeficientů v řadě a velikost (v bitech) kódu hodnoty koeficientu, a tato hodnota

Obrázek: Dočasná a kódová reprezentace koeficientů a bloku

- 11 kódování dvojic bloku: prvního prvku dvojice **Huffmanovým kódováním** (se statickým statisticky zjištěným modelem), druhého **celočíselným kódováním** s proměnlivou velikostí kódů – neprefixové, kódy kratší než u prefixového, jednoznačně dekódovatelné se znalostí velikosti kódů

Obrázek: Ilustrace JPEG komprese

DC coefficient representation:

symbol -1 symbol -2
 <SIZE> <AMPLITUDE>

AC coefficient representation:

symbol -1 symbol -2
 <RUNLENGTH, SIZE> <AMPLITUDE>

Intermediary stream

<2><3> <0,3><4> <0,2><3> <0,1><1> <0,2><-3> <0,3><6>
 <0,2><2> <0,2><-2> <0,2><2> <1,1><1> <1,1><-1> <0,1><-1>
 <0,2><2> <0,1><-1> <0,1><1> <0,1><-1> <1,1><1> EOB

Intermediary symbol	Binary representation of first symbol (prefixed Huffman Codes)	Binary representation of second symbol (non-prefixed variable integer codes)
<2> <3>	011	11
<0,3> <4>	100	100
<0,2> <3>	01	11
<0,1> <1>	00	1
<0,2> <-3>	01	00
<0,3> <6>	100	110
<0,2> <2>	01	10
<0,2> <-2>	01	01
<0,2> <2>	01	10
<1,1> <1>	11	1
<1,1> <-1>	11	0
<0,1> <-1>	00	0
<0,2> <2>	01	10
<0,1> <-1>	00	0
<0,1> <1>	00	1
<0,1> <-1>	00	0
<1,1> <1>	11	1
EOB	1010	

Binary Stream:

0111110010001110010100100110011001010110111100000110000010001111010



Original—24 bits per pixel



Compressed—2 bits per pixel



Compressed—0.5 bits per pixel



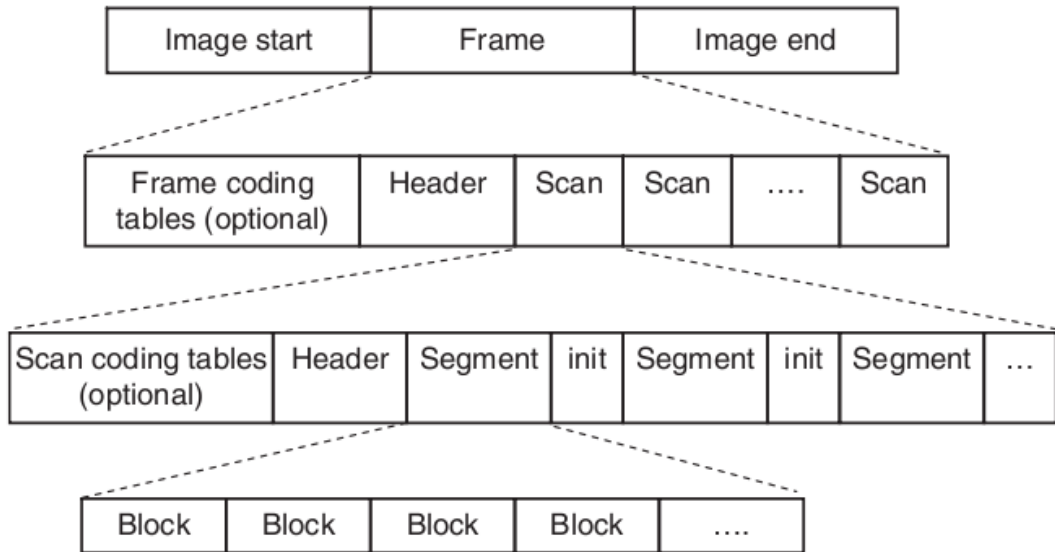
Compressed—0.15 bits per pixel

12 složení kódů bloků do bitového proudu – podle JPEG standardu

Obrázek: JPEG bitový proud

Nevýhody

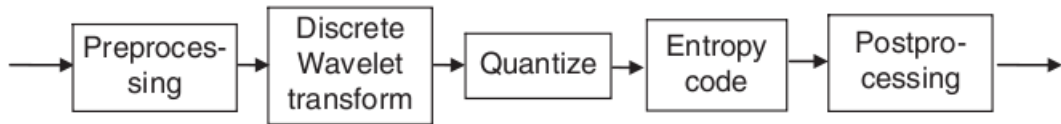
- pro narůstající rozlišení obrazu a potřebu úprav komprimovaného obrazu je dnes potřeba lepších metod
- vysoké **zkreslení na nízkých bit rate** (velmi nízké na středních a vyšších), **nemožnost náhodného přístupu** k částem obrazu (závislé kódování bloků), komprese max. 64kB bloků dat obrazu, kolem 40 různých **módů** (většinou aplikačně specifických, neimplementovaných dekodery), neodolnost proti chybám přenosu, neoptimální pro počítačově generované obrazy (vytvořený pro foto)



- ISO standard
- pro různé typy obrazů a různé modely přenosu a zobrazování
- použití **diskrétní waveletové transformace (DWT)** – lepší popis ve frekvenční doméně než DCT

Obrázek: Obr. JPEG 2000 komprese

- 1** (volitelně) **dlaždicování**: rozklad obrazu na čtvercové stejně velké bloky, s příp. rozšířením obrazu o nulové okraje, bloky zpracováván samostatně
- 2** transformace barev do **barevného prostoru YCbCr** (podobný YUV), každý barevný kanál zpracováván samostatně
- 3** **posunutí úrovní**: odečtení poloviny rozsahu hodnot od hodnot pixelů pro rozsah s 0 uprostřed, kvůli DWT



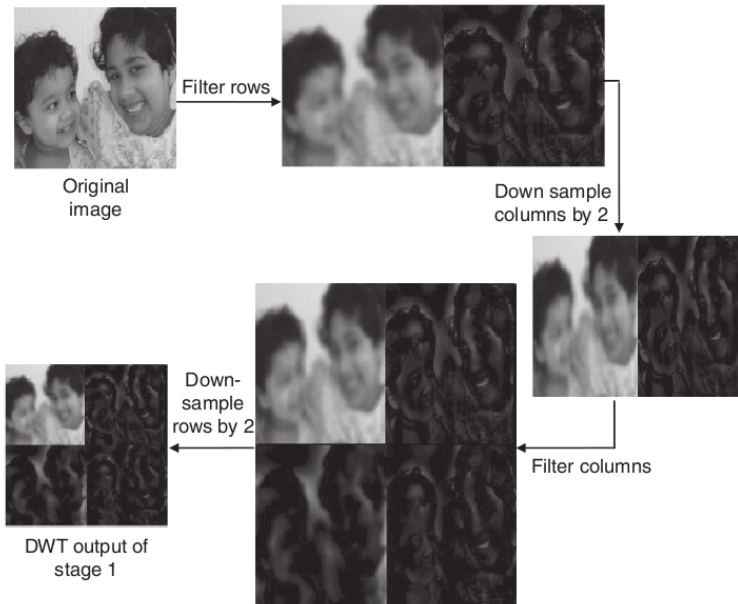
- 4 DWT: jednorozměrná**, krok rekuze nejdříve po řádcích, pak po sloupcích → 4 **kvadranty pásem LL, LH, HL a HH**, rekurzivní opakování na kvadrant LL, až 32-krát (v praxi 4–8-krát)

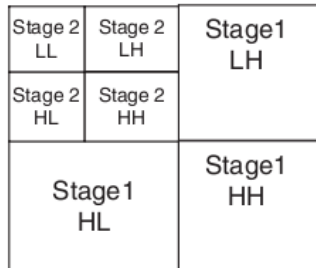
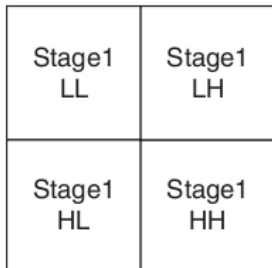
Obrázek: Obr. DWT

- 5 kvantizace** frekvenčních koeficientů v kvadrantech a **přeskládání** hodnot do paketů s hodnotami pro dané rozlišení od nejnižšího
- 6** pakety hierarchicky uspořádány do bloků a bloky komprimovány **statistickým kódováním**

Výhody (oproti JPEG)

- DWT poskytuje **lepší kompresní poměr** než DCT
- komprese a dekomprese (vybraných částí) obrazu v **různých kvalitách** (hierarchická organizace, region of interest, ROI, coding), obrazové **operace** (geometrické transformace, filtrace) na komprimované reprezentaci

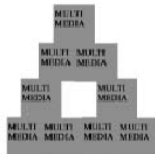
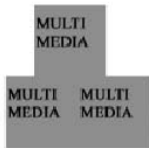
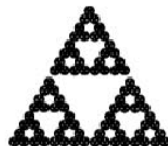
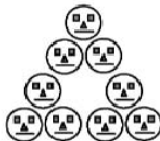
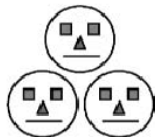




- = vyhledávání **podobných částí obrazu a transformací** (predikcí) vzorových na podobné a kódování vzorových částí a transformací
- vyhledávání výpočetně náročné
- použití **fraktálů** = sobě podobné geometrické objekty, kde části vypadají stejně jako celek – vytvořen a plně určen **rekurzivní afinní transformací** libovolného vzoru (tzv. seed), např. Sierpinskyho transformace (trojúhelník)

Obrázek: Obr. Fraktál definovaný Sierpinskyho transformací

- ideálně nalezení transformace, kde obraz = fraktál – „inverzní transformace“, v praxi (pro běžné obrazy) kombinace **více transformací pro části obrazu**



- = **rozklad** obrazu na čtvercové stejně velké bloky a pro každý blok **vyhledávání** vzorového bloku (v obrazu) a jeho **transformace** na blok s minimálním zkreslením

Obrázek: Obr. Fraktálová komprese

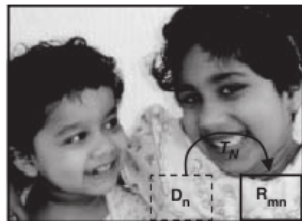
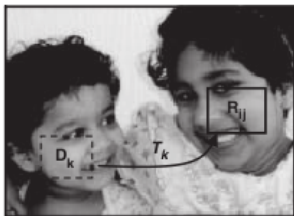
- experimentální stav (vyhledávání transformací), formát **FIF**), výsledky srovnatelné s DCT a DWT až mírně lepší



Construct
range blocks



For each range block R_{ij} , compute the domain block D_k and the affine transform T_k

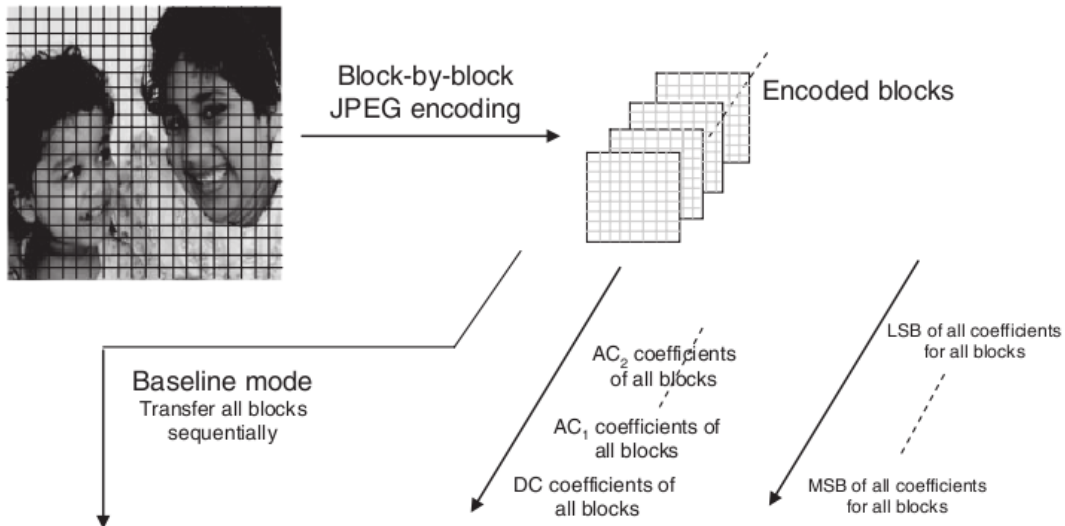


- problém: uložení a zobrazení (dekomprese) obrazu v plné kvalitě sekvenčně po blocích
 - pro celý obraz potřeba všechny
- **uložení a zobrazení obrazu postupně po úrovních kvality od nejhorší** (nejhrubší obraz) po plnou = **progresivní**
- JPEG: **progresivní módy** s více skeny bloků, např.
 - **spektrální selekce** – v 1. skenu jen DC koeficienty všech bloků, ve 2. skenu první AC koeficienty atd.
 - **postupná bitová aproximace** – v 1. skenu nejvýznamnější bity všech koeficientů všech bloků, ve 2. skenu druhé nejvýznamnější bity atd.

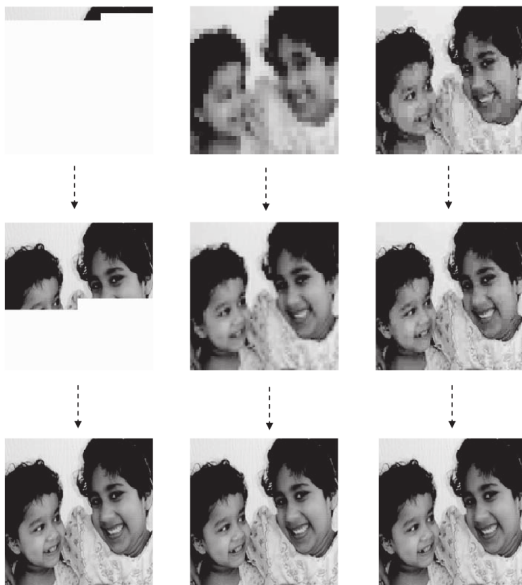
Obrázek: Obr. Baseline a progresivní uložení v JPEG

- JPEG 2000: samo o sobě (**hierarchická organizace**), na dané úrovni rekurze (rozlišení) nejdříve nízkofrekvenční pásma

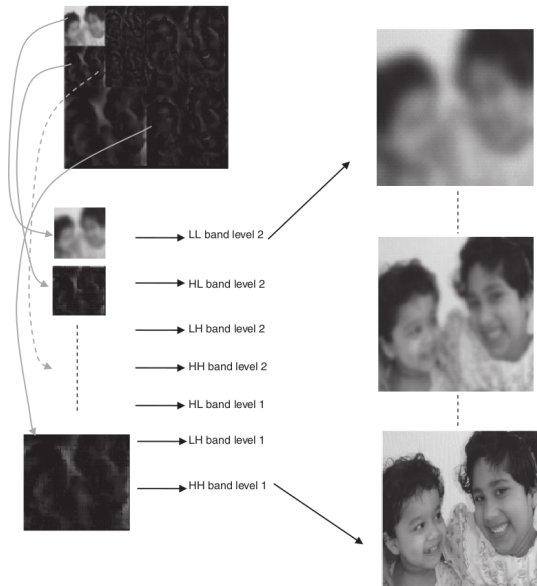
Obrázek: Obr. Progresivní uložení v JPEG 2000



Progresivní uložení komprimovaného obrazu



Progresivní uložení komprimovaného obrazu





Kompresa videa



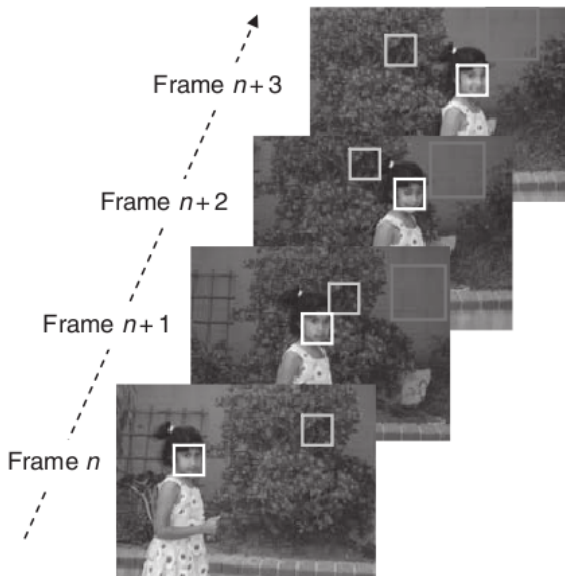
- video významné podobně jako obraz – zábavní průmysl, nárůstající na rozlišení a objemu, real-time distribuce → **(nezbytná) potřeba komprese**
- např. 720×576 , 12bpp (DVD), 25fps prokládaně ~ 120 Mbps → 9.8 Mbps max, běžně 2 Mbps (MPEG-2), 1920×1080 , 12bpp (HDTV), 25fps prokládaně ~ 590 Mbps → 8 Mbps (MPEG-2), 4 Mbps (MPEG-4)
- např. standardy (ISO) MPEG-x, (ITU) H.26x
- metody založené na technikách **predikce a kompenzace pohybu**

- u obrazu prostorová (a spektrální) redundance – oblasti podobných pixelů (málo frekvencí)
- video = sekvence obrazů → **kompresi jednotlivých snímků**, tzv. **intra-frame** – např. formáty **M-JPEG (Motion JPEG) (2000)**, u aplikací s nízkou frame rate (1–2 fps), např. instant messaging, videokonference, kamery pro ostrahu, nízká komprese (pro video)
- = **(lokální) korelace barev pixelů v čase/mezi snímky** – pixely zůstávají podobné, např. pozadí, postupné změny

Obrázek: Obr. Časová redundance, podobné snímky

- uložení pouze **změn mezi snímky**, tzv. **rozdílových snímků** – menší entropie, podobné DPCM

Obrázek: Obr. Rozdílový snímek





Frame n



Frame $n+1$



Frame $(n+1) - \text{Frame } n$



Frame n



Frame $n+1$



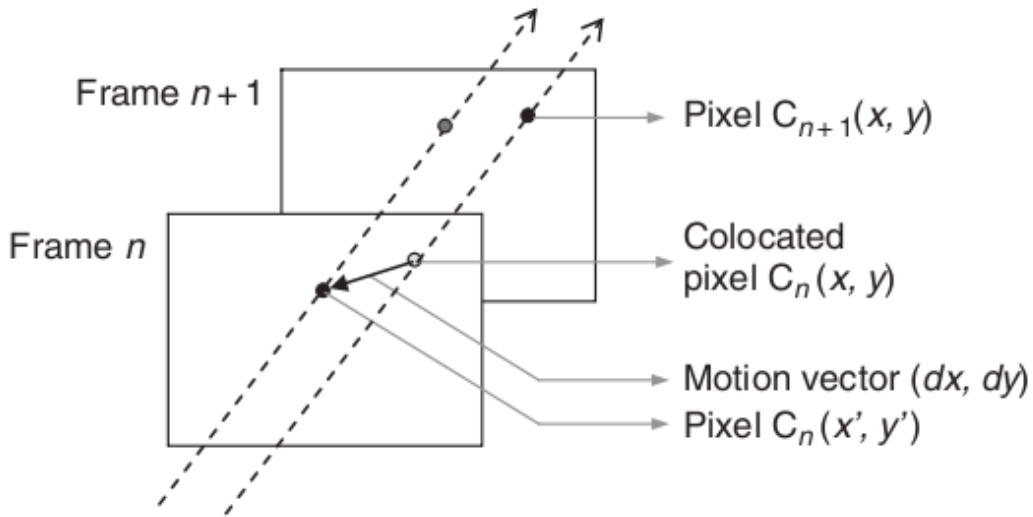
Frame $(n+1) - \text{Frame } n$

- další **časová redundance**: **pohyb (lokálních) oblastí pixelů** – objektů, při pohybu kamery

Obrázek: Obr. Časová redundance, pohyb oblastí

- entropie rozdílového snímku nepřímo závislá na množství a rychlosti pohybů – plynulý, malý vzhledem k frame rate \Rightarrow predikovatelný \rightarrow **predikce pohybů oblastí**
 - barva pixelu na pozici (x, y) ve snímku n : $C_n(x, y)$
 - $C_{n+1}(\mathbf{x}, \mathbf{y}) = C_n(\mathbf{x} - \mathbf{dx}, \mathbf{y} - \mathbf{dy}) + e(\mathbf{x}, \mathbf{y})$ – posunutí pixelu o **vektor pohybu** (dx, dy) (v důsledku pohybu objektu nebo kamery) plus chyba $e(x, y)$ (kvantizační chyba, šum, změna světelných podmínek apod.)
- \rightarrow predikce $C_{n+1}(x, y)$ jako $C_n(x - dx, y - dy)$ a opravení o $e(x, y)$

Obrázek: Obr. Predikce pohybu pixelu



- stejný pohyb (lokálních) oblastí pixelů
- = rozklad snímku na tzv. **makrobloky** (čtvercové, stejně velké) a **predikce makrobloků** ze stejně velkých (libovolně umístěných) oblastí předchozího, **referenčního, snímku**
- ne vždy možná – když pro makroblok není oblast
- predikce (cílových) snímků z referenčních snímků = **kompensace pohybu (motion compensation)**

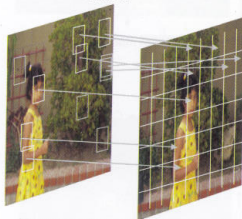
Obrázek: Obr. Predikce snímků



Frame n



Frame $n+1$



Macroblock prediction of frame $n+1$ from regions in frame n



Reconstructed frame $n+1$ by macroblock prediction



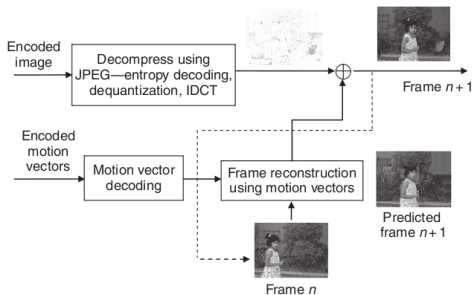
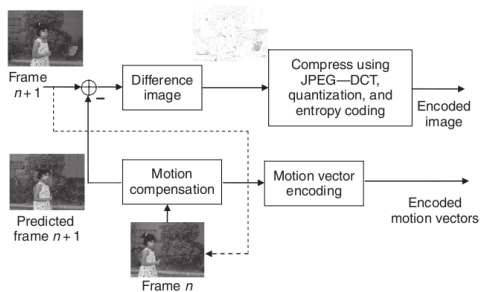
Frame difference
frame $n+1$ - frame n



Frame difference
Reconstructed frame $n+1$ - frame $n+1$

- 1 první snímek přímo komprimován **obrazovou kompresí**, např. JPEG – tzv. **intra mód**
- 2 pro další snímky – tzv. **inter mód** – **nalezení vektorů pohybu** (a tedy oblastí v referenčním snímku) pro makrobloky
- 3 **vytvoření predikovaného** (pohybově kompenzovaného) **snímku** = zkopírování oblastí z referenčního snímku daných vektory pohybu do makrobloků
- 4 **odvození chybového snímku** = rozdíl mezi skutečným a predikovaným snímkem
- 5 **kompresa chybového snímku** obrazovou kompresí, např. JPEG, (ztrátové) a **vektorů pohybu** pomocí statistického kódování (Huffmanova nebo aritmetického, bezztrátové, se statickým empirickým modelem)

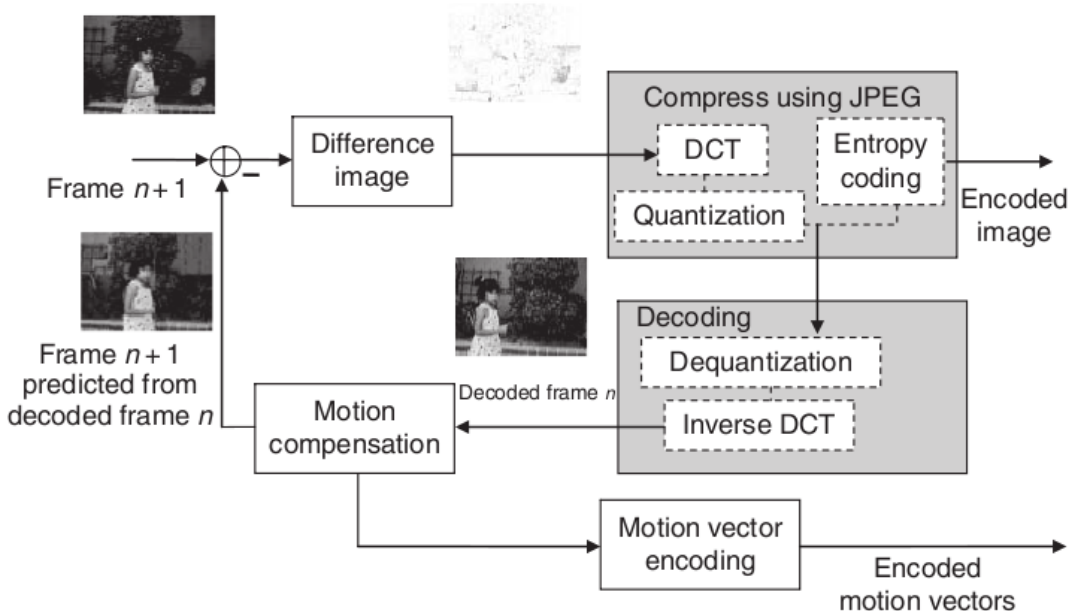
Obrázek: Obr. Kompresa (open loop případ) a dekomprese



- open loop vs. closed loop případ – jako DPCM pro snímky jako hodnoty signálu

Obrázek: Obr. Closed loop případ komprese

Obrázek: Obr. Pseudokódy komprese a dekomprese




```
referenceFrame = NULL;
While input frames are available
{
    currentFrame = getNextFrame();
    if (intramode)
    {
        codedFrame = codeFrame (currentFrame); //Using JPEG pipeline
        dumpToStream (codedFrame);
        reconstructedFrame = decodeFrame (codedFrame);
        referenceFrame = reconstructedFrame;
    }
    else if (intermode)
    {
        for each macroblock in image
        {
            // motion vector computed for each macroblock
            mVs[i] = computeMotionVector(macroblock, referenceFrame);
        }

        // Create a motion compensated frame by copying blocks from reference
        // frame using computed motion vectors
        predictedFrame = compensateFrame(mVs, referenceFrame);

        // Compute the error Image
        errorFrame = currentFrame - predictedFrame;

        codedFrame = codeFrame(errorFrame); //Using JPEG pipeline
        codedMvs = codeMotionVectors (mVs); // entropy code motion vectors
        dumpToStream (codedMvs);
        dumpToStream (codedFrame);
    }
}
```

```
referenceFrame = NULL;
While encoded data is available
{
    currentCodedFrame = getNextCodedFrame();
    currentCodedMvs = getNextCodedMotionVectors();

    if (intramode)
    {
        frame = decodeFrame (currentCodedFrame); // Using JPEG pipeline
        referenceFrame = frame;
        sendToDisplay (frame);
    }
    else if (intermode)
    {
        errorFrame = decodeFrame (currentCodedFrame);
        // Using JPEG pipeline
        mVs = decodeMotionVectors (currentCodedMvs);
        predictedFrame = compensateFrame (mv, referenceFrame);
        frame = predictedFrame + errorFrame;
        sendToDisplay (frame)
    }
}
```

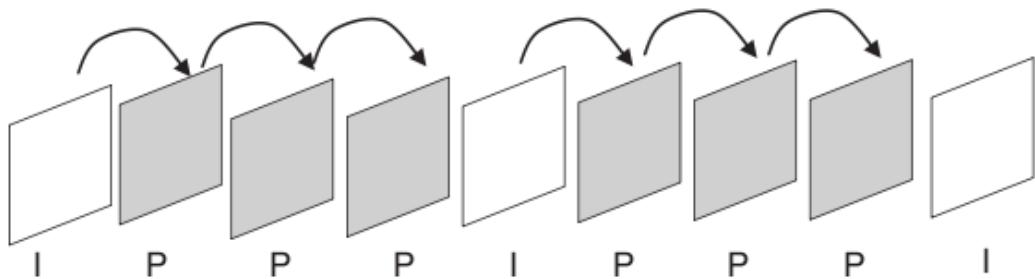
Snímky I

- = **komprimované** přímo, samostatně **bez predikce**, obrazovou kompresí, např. JPEG, v intra módu \Rightarrow snadná komprese, ale větší bit rate než predikované snímky
- (typicky) kódované **periodicky** v posloupnosti snímků, **referenční snímky** pro predikci
 - když by chybový snímek byl příliš obsáhlý (měl velkou entropii = velký bit rate, např. při změně scény) nebo by jeho výpočet trval příliš dlouho (real-time přenos) \rightarrow možnost detekce

Snímky P

- = **prediktivně (P) komprimované z (přímo) předchozího snímku I nebo P** \Rightarrow náročná komprese, ale nižší bit rate než snímky I
- chybový snímek má nižší entropii než snímek I \rightarrow hrubější kvantizace (např. vyšší kvantizační faktory u JPEG, typicky čtyřnásobné)
- vytváří dopřednou závislost mezi snímky

Obrázek: Obr. Snímky P

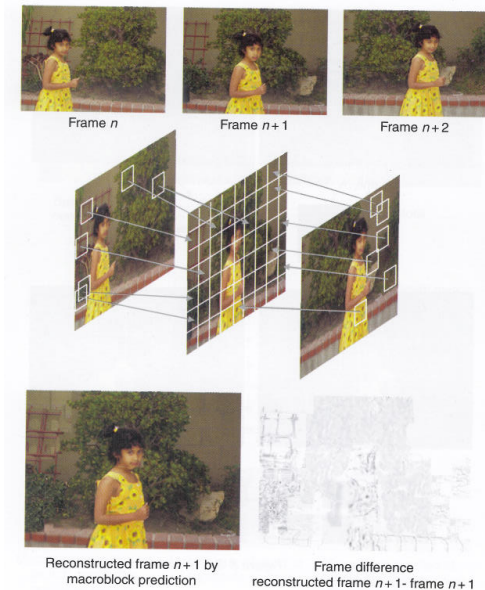


Snímky B

- = **dvousměrně** (bidirectionally, B) **prediktivně komprimované z předchozího a následujícího snímku I nebo P** \Rightarrow náročnější komprese, ale nižší bit rate než snímky P
- někdy je pro kompenzaci pohybu lepší následující než předchozí snímek – např. objevení se objektů při pohybu

Obrázek: Obr. Predikce snímků B

- vyhledání a **výběr z vektorů pohybu** (a tedy oblastí odpovídajících makroblokům) pro oba referenční snímky **nebo použití obou vektorů** (a interpolace oblastí jako predikce makrobloku)



Snímky B

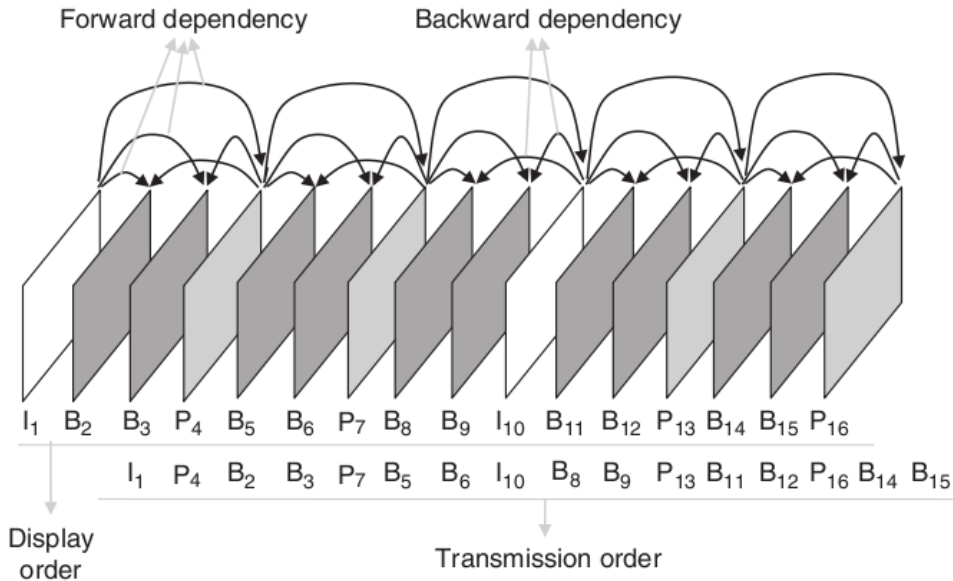
- vytváří dopřednou i **zpětnou závislost mezi snímky** → potřeba je při uložení/distribuci kvůli možnosti dekomprese **přeskládat** do tzv. **přenosového pořadí** (snímky B před referenčními snímky I nebo P) – může způsobovat prodlevy při zobrazení (v zobrazovacím pořadí), **bufferování snímků**

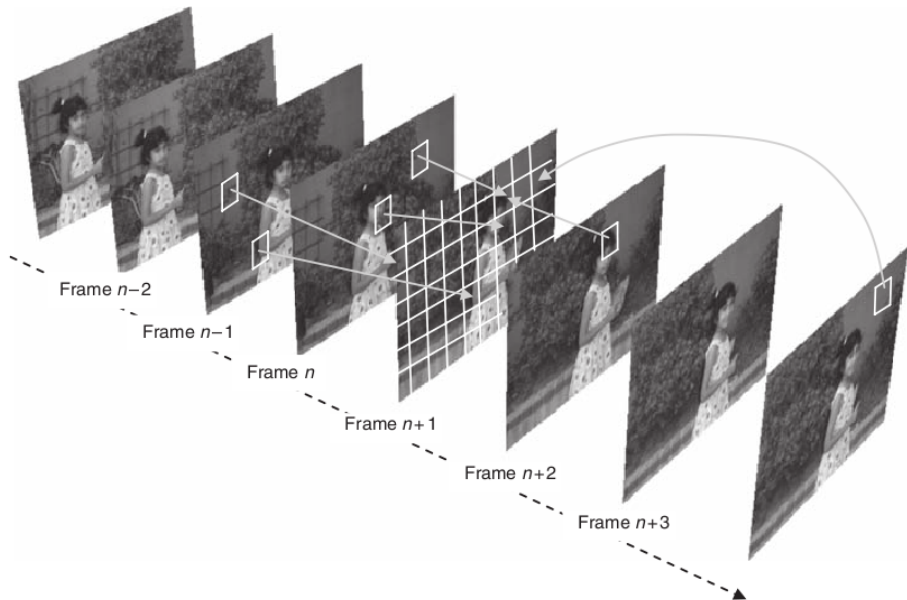
Obrázek: Obr. Snímky B

Multiframe predikce

- = **predikce z více předchozích i následujících snímků** ⇒ lepší, náročnější, ale menší bit rate

Obrázek: Obr. Multiframe predikce





- = posloupnost kódovaných snímků daná frekvencí a použitím typů snímků, **opakující se skupiny snímků**, tzv. **group of pictures (GOP)**
- první snímek I, pak několik (včetně 0) snímků P a mezi nimi snímky B
- reprezentována počtem snímků a vzdáleností (v počtu snímků + 1) mezi nejbližšími I/P snímky nebo počty snímků P mezi I a B mezi P

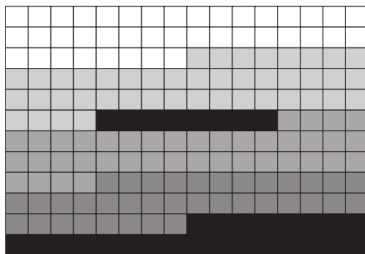
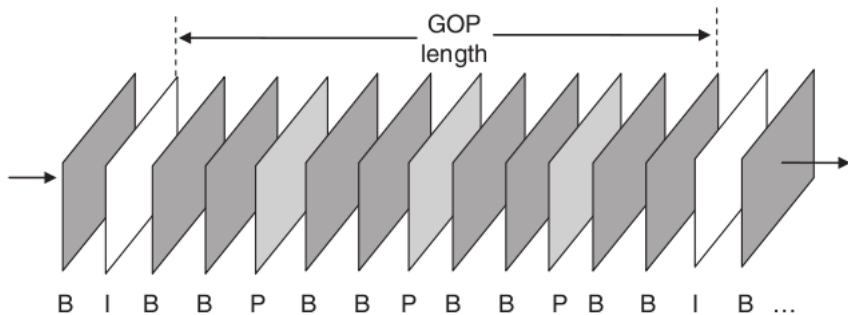
Obrázek: Obr. GOP (MPEG-1)

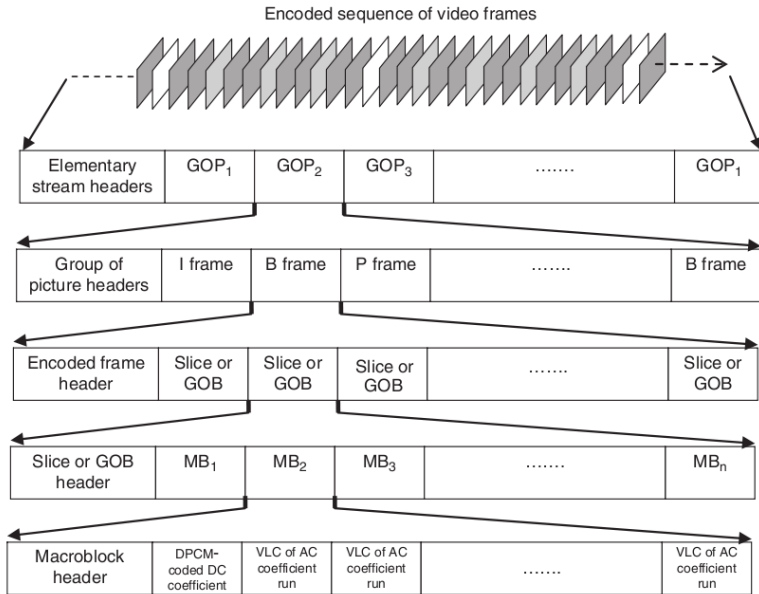
- podporuje **náhodný přístup** (např. „přetáčení“) – k snímkům I
- **slice/GOB** (group of macroblocks) = skupina makrobloků ze snímku, pro variabilní kódování částí snímků

Obrázek: Obr. Slices/GOBs (MEPG-1)

- standardově závislé

Obrázek: Obr. Bitový proud



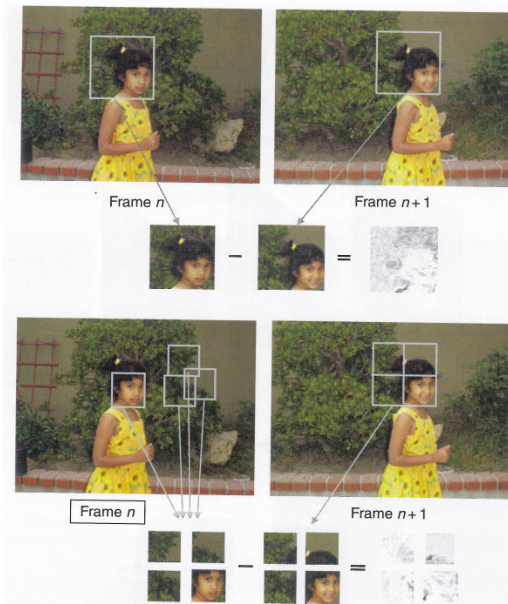


- = prohledávání plochy v referenčním snímku kolem korespondující (kosituované) pozice makrobloku v cílovém snímku na nejpodobnější oblast (**referenční makroblok**) a určení vektoru pohybu
- barvy pixelů makrobloku: $C_{n+1}(x, y)$, kde $x, y \in [0, m]$ a m je velikost makrobloku

Velikost makrobloků

- = kompromis mezi menšími (větší počet makrobloků a vektorů pohybu, ale vektory menší = lepší chybový snímek) a většími (méně vektorů, ale horší chybový snímek – potenciálně různé oblasti pro části makrobloku)
- ve standardech **16 × 16 pixelů** nebo variabilní, pak např. pro snímek 1920×1080 (HDTV) je 8100 makrobloků

Obrázek: Obr. Velikost makrobloku



Míry nepodobnosti/rozdílu oblasti a makrobloku

- mean absolute difference (MAD):

$$MAD(dx, dy) = \frac{\sum_{x=1}^m \sum_{y=1}^m |C_{n+1}(x, y) - C_n(x + dx, y + dy)|}{m^2}$$

- **sum of absolute difference (SAD)**, používanější): $MAD \cdot m^2$
- mean square difference (MSD): $MAD^2 \cdot m^2$
- **pel difference** classification, kde t je zvolený práh shody a $\|p\| = 1$ pro pravdivé p , jinak 0:

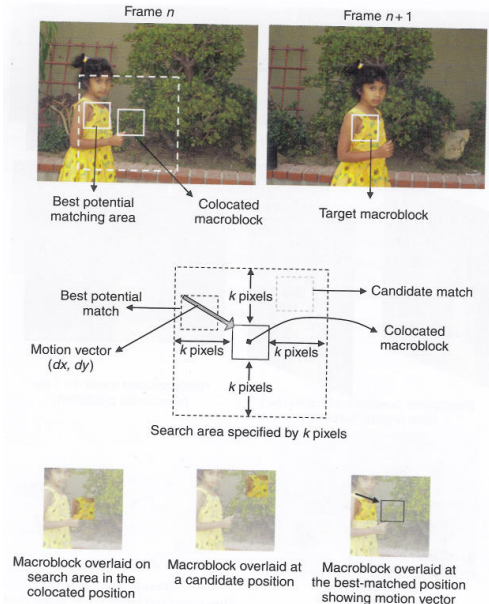
$$PEL(dx, dy) = \sum_{x=1}^m \sum_{y=1}^m \mathbb{1}_{\|C_{n+1}(x, y) - C_n(x + dx, y + dy)\| \leq t}$$

- pro $dx, dy \in [-k, k]$

- **prohledávaná plocha** = čtverec s hranami k pixelů od hran makrobloku, k je parametr vyhledávání, typicky 0 až 32, volený podle rychlosti pohybu objektů/oblastí mezi snímky

Obrázek: Obr. Vyhledávání vektoru pohybu

- nalezení oblasti/vektoru – kritický, **centrální problém**: lepší vektory pohybu (= podobnější oblast) = lepší predikce makrobloku = lepší chybový makroblok (menší entropie) = menší bit rate, ale delší doba vyhledávání (která nemusí být dostupná, např. u real-time komprese/přenosu), typicky 60–80 % celkového času komprese



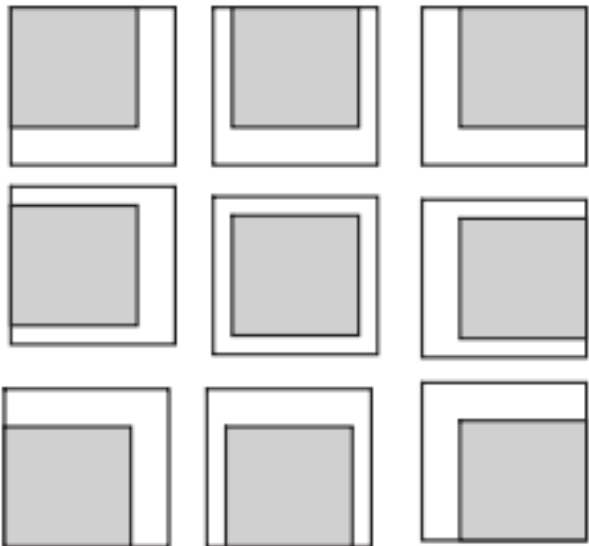
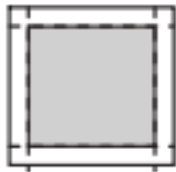
- složitost porovnání oblasti a makrobloku (výpočet míry rozdílu): $O(m^2)$
- složitost v počtu porovnání při hledání nejpodobnější oblasti/nejlepšího vektoru pohybu
- algoritmus jako další parametr vyhledávání, vedle k

Sekvenční/Hrubá síla

= porovnání na **všech** $(2k + 1)^2$ **pozicích**

Obrázek: Obr. Sekvenční vyhledávání pro $k = 1$

- např. pro snímek 1920×1080 (HDTV) s makrobloky 16×16 a $k = 32$ je přes 34 mil. porovnání
- nepoužitelné pro real-time přenos, ale záruka nejlepších vektorů = nejlepší komprese



Logaritmické/Binární (Fast motion estimation, FME)

- rozdíl oblasti a makrobloku se často monotónně zmenšuje k nejpodobnější oblasti, ve směru pohybu
- = opakované porovnání na 9 (v 1. iteraci) a 8 pozicích s $dx, dy \in \{-\frac{k}{2}, 0, \frac{k}{2}\}$ bez $(0, 0)$ (od 2. iterace), výběr nejlepší a zmenšení k na polovinu = **zpřesňování vektoru pohybu**: $8 \log_2 k + 1$ porovnání

Obrázek: Obr. Logaritmické vyhledávání pro $k = 16$

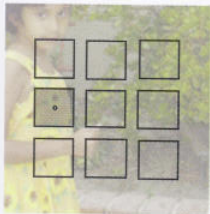
- např. pro snímek 1920×1080 (HDTV) s makrobloky 16×16 a $k = 32$ je přes 330 tis. porovnání



Search area around the colocation



Macroblock of frame $n+1$



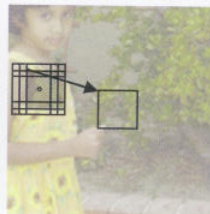
Iteration 1



Iteration 2



Iteration 3



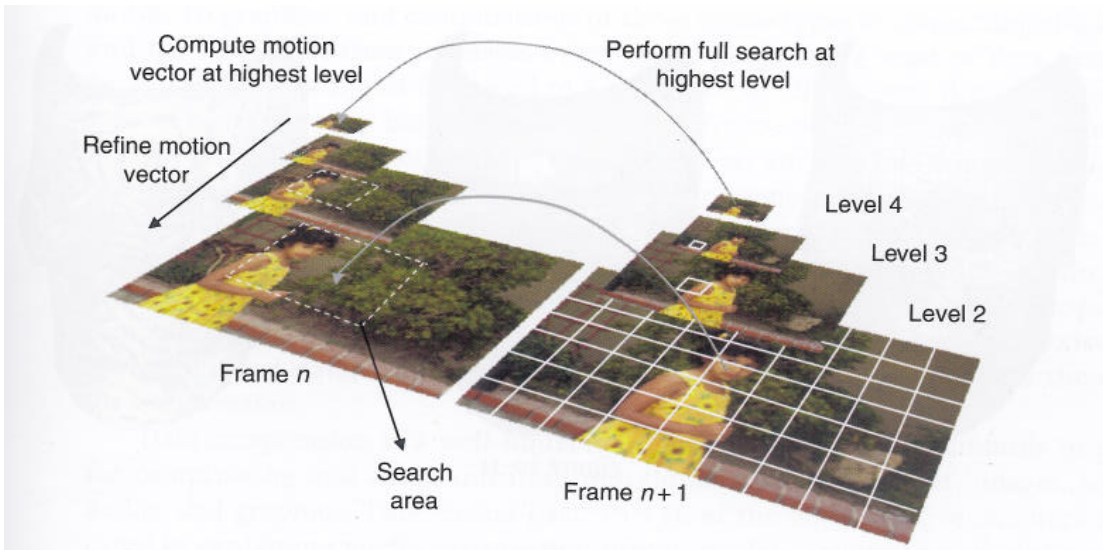
Iteration 4 showing motion vector

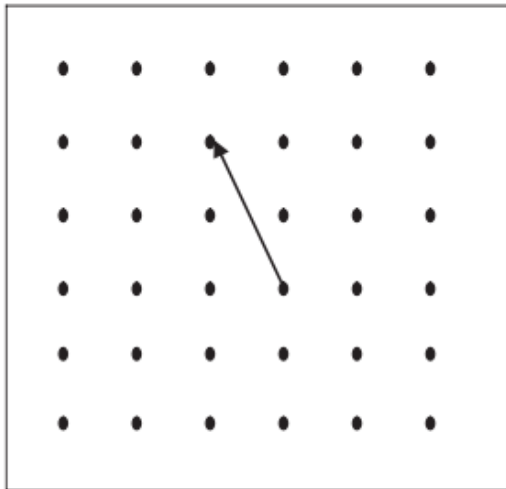
Hierarchické

- vytvoření **hierarchie verzí snímků** pro cílový i referenční snímek **se zmenšujícím se rozlišením** na polovinu (podvzorkování a filtrace) – zmenšuje se i velikost makrobloků/oblastí a k
- = sekvenční vyhledání na nejvyšší úrovni (s nejnižším rozlišením a nejmenšími makrobloky/oblastmi a k), pak logaritmické na nižších úrovních = zpřesňování vektoru pohybu

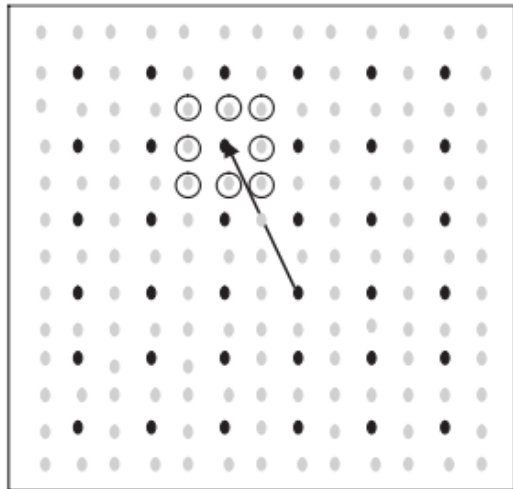
Obrázek: Obr. Hierarchické vyhledávání

- např. pro snímek 1920×1080 (HDTV) s makrobloky 16×16 , $k = 32$ a čtyřmi úrovněmi je přes 850 tis. porovnání





Level $k + 1$



Level k

- od 1984 (H.120), pro aplikace dané doby
- některé standardizované – ISO MPEG-x, ITU H.26x
- standardy definují komprimovaný bitový proud (**pro dekódování**), ne konkrétní algoritmy kodeku (implementace)
- patentované, s poplatky za užití (MPEG-x/H.26x) i bez (Theora, VPx), svobodné (AV1) – otázka pro web

H.261 (kodek FFmpeg)

- 1988, \sim px64 – pro ISDN (rychlost v násobcích 64 kbps)
- až 352×288 (CIF), jen progresivní, snímky jen I (JPEG) a P

MPEG-1 (FFmpeg)

- 1991, i audio, použití ve **Video CD (VCD)** – srovnatelné s VHS
- $352 \times 240/288$ (SIF), jen progresivní, až 1.5 Mbps
- **half pixel** kompenzace pohybu (**HPEL**, zvýšení přesnosti, lepší predikce), použití slices/GOBs u predikovaných snímků

H.262/MPEG-2 (FFmpeg, x262)

- 1994, i audio, pro **digitální TV** (SIF/**SDTV**, pak **HDTV**), použití v **DVD**
- **kontejnerové formáty program stream (PS)** pro uložení na disková média, a **transport stream (TS)** pro přenos chybovým kanálem se synchronizací
- i prokládané – **predikce pulsů** z pulsů/snímků, až 20 Mbps (2 SDTV, 15–18 HDTV)
- half pixel kompenzace pohybu, **škálovatelné kódování** – možnost dekódovat různá rozlišení, base (nižší) a enhancement (vyšší frekvence) **vrstva videa**

H.263/MPEG-4 (Advanced Simple Profile) (FFmpeg, DivX, Xvid)

- 1999, i audio a 2D/3D grafika – **media objekty** = více proudů (videa, audia, grafiky)
- i prokládané, rozsah bit rate od 64 kbps (streaming, wireless) do 4 Mbps
- **quarter pixel (QPEL)** a globální kompenzace pohybu, popisy rozmístění objektů, **roviny videa** (video object planes, VOPs) – libovolný neobdélníkový tvar snímků, různé bit rate, skládání rovin v přehrávači (např. klíčování), synchronizovaná **textová a metadatová stopa** (na úrovni snímků)
- nekompatibilní varianta MS MPEG-4v3 (kodek DivX)

H.264/MPEG-4 AVC (Advanced Video Coding) (x264, FFmpeg, OpenH264, DivX, Quicktime)

- 2004/2003, pro **digitální TV (HDTV)**, **Blu-ray** (a HD-DVD), internet streaming (Youtube, Netflix aj.), web atd.
- **menší bloky** u JPEG až do 4×4 pixely, v intra módu slices/GOBs bloků s **prostorovou predikcí** bloků z předchozích a DCT rozdílů, tzv. **deblocking filter** – prevence blokových artefaktů DCT na hranicích bloků, v enkoderu i dekoderu
- **variabilní velikost makrobloků**: hierarchický rozklad na 8×16 , 16×8 a 8×8 a poslední na 8×4 , 4×8 a 4×4 – lepší predikce (na hranicích pohybujících se objektů)
- multiframe predikce, kontextově adaptivní binární aritmetické kódování (CABAC, kontextově adaptivní model) vektorů pohybu a chybových snímků
- varianty MPEG-4 Web Video Coding, Scalable Video Coding, XAVC

DODELAT

H.265/MPEG-H HEVC (High Efficiency Video Coding) (x265)



Theora (libtheora, FFmpeg)

- Xiph.org

VP6–9 (libvpx, FFmpeg)

- Google (On2), použití v Adobe Flash

AV1 (aomenc)



Další

- Windows Media Video (WMV), Apple ProRes, Apple Sorenson, Cinepak (~ Apple Compact Video), RealVideo, Flash, Dirac

- **variabilní (VBR)** – „pomalé“ a „rychlé“ pasáže videa, čím více pohybu na sousedních snímcích, tím vyšší bit rate (větší, nepřesnější, vektory pohybu, obsáhlejší chybové snímky)
- **konstantní (CBR)** – ideálně, pro přenos, zvláště real-time – omezená rychlost (šířka pásma) sítě → **řízení bit rate**

Řízení

- **pevná GOP a bit rate pro GOP**
 - pevné bit rate pro snímky – empiricky snímek I = 3× snímek P = 6 – 15× snímek B ⇒ horší kvalita (variabilní složitost obsahu)
 - **variabilní bitrate pro snímky** – cache snímků, adaptivní kódování podle množství pohybu, méně = více bitů pro snímky I, více = více bitů pro snímky P, méně bitů pro snímky B, (proprietární) heuristiky
- **variabilní bit rate pro GOP** – omezené rychlostí a velikostí bufferu dekodéru

- předzpracování videa – low-pass filtry pro **odstranění šumu a chyb** (při digitalizaci z analogového signálu)

Parametry

- **rozlíšení, frame rate** – standardizované (PAL/NTSC, DVD apod.)
- **bit rate, CBR/VBR, počet průchodů** – více pro CBR, první pro statistickou analýzu pro variabilní alokaci bit rate, adaptace kompresních modelů atd.
- struktura GOP: počty snímků P mezi I – u „rychlého“ videa menší (horší predikce), B mezi P – kompromis mezi bit rate a dobou komprese, lepší odolost proti chybám přenosu při více snímcích I a P
- obraz: kvantizační tabulky (J/MPEG, ITU) a faktor („JPEG kvalita“)
- kompenzace pohybu: velikost prohledávané plochy a vyhledávací algoritmus, half/quarter pixel vyhledávání
- přeskokování snímků – při omezené rychlosti přenosu
- **informace pro dekoder (DSI)** – vkládat parametry komprese průběžně, nejen do záhlaví, např. u streamu
- ...



Komprese audia

- Audio CD nekomprimované, ale šíření audia po sítích s omezenou rychlostí (streaming, telefonní sítě) a médiích s omezenou kapacitou (přehrávače), narůstající požadavek kvality ⇒ **potřeba komprese**
- např. 44.1 kHz, 16 bps, stereo (CD) ~ 1.3 Mbps \rightarrow 160 – 230 kbps (MP3), 192 kHz, 16 bps, 5.1 surround ~ 18 Mbps \rightarrow 256 – 2048 kbps (AC-3)
- např. standardy (ISO) MPEG, (ITU) G.7xx, Dolby AC-x aj.
- audio: **širokopásmové** = hudba a **úzkopásmové** = mluvené slovo (řeč)
- zvuk reprezentován jako 1D signál, ale **méně redundance** a lidský zvukový systém citlivější na šum a zkreslení než u obrazu a videa, vnímání závisí na intenzitě a frekvenci zvuku
- různé **odlišné metody**: predikce (DPCM) a bezztrátové (odstranění statistické redundance), využití psychoakustických modelů, modelů zdroje zvuku, sémantické modely (strukturované audio)



- = **relativní** jednotka pro měření amplitudy (intenzity) signálu
- = $10 \log \frac{I}{I_0}$, kde I (I_0) je intenzita (referenčního) signálu v W/cm^2
- referenční signál: experimentálně zjištěný sotva slyšitelný zvuk s $I_0 = 10^{-6} W/cm^2$
- např. konverzace 40 dB, doprava 70–90 dB, práh bolesti 120–130 dB

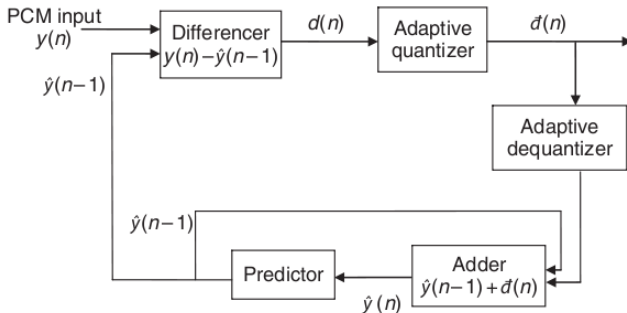
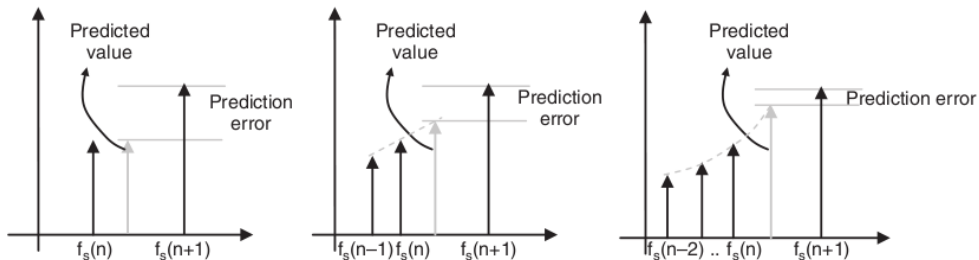
(A)DPCM

- predikovaná hodnota = předchozí hodnota nebo aproximace z **více předchozích hodnot** – s (predikčními) koeficienty, menší chyba

Obrázek: Obr. DPCM z více předchozích hodnot

- ADPCM (Adaptive DPCM): adaptace počtu bitů pro kvantizaci chyby (rozdílu)
 - řídicí bity navíc zvyšují entropii → dva módy: pro nízké a vysoké frekvence

Obrázek: Obr. ADPCM



Delta modulace

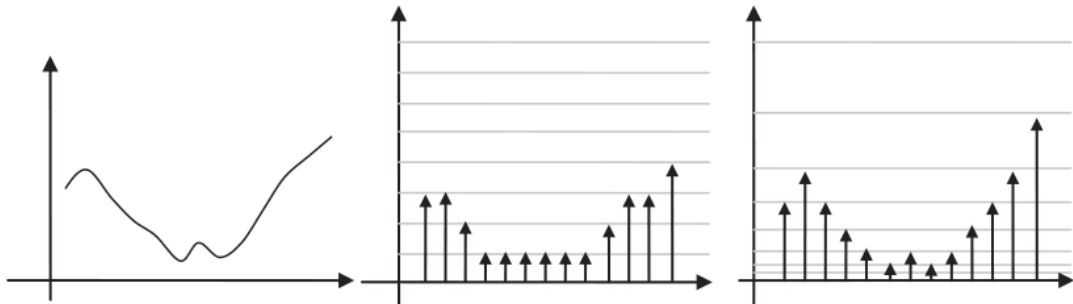
- jednoduchá, jako DPCM, ale chyba (rozdíl) kódována jedním bitem – 0 při zvýšení nebo 1 při snížení hodnoty o **konstantu (delta)**
- delta je pevná, určená na základě vzorkovací frekvence
- pro pomalu se měnící signál (řeč, do 8 kHz), jinak velké chyby, kontrolovaná bit rate

A/ μ -law

- **logaritmický kvantizační rozsah (companding)** – empiricky častěji nízké intenzity signálu, např. u řeči → rozdíl kvantizačních úrovní přímo úměrný intenzitě, častější nízká intenzita = menší kvantizační chyba

Obrázek: Obr. Uniformní a logaritmická kvantizace

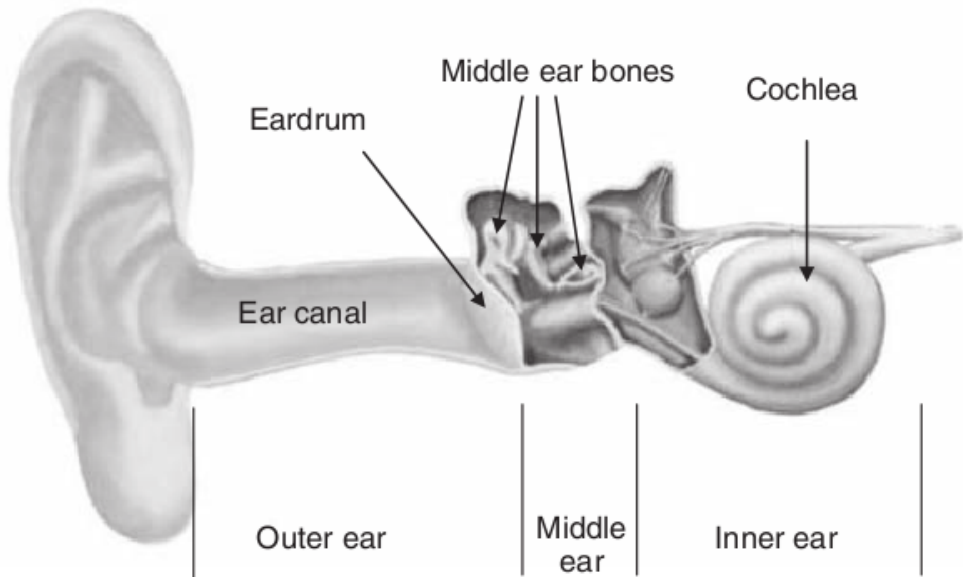
= ITU standardy pro **kompresi řeči**: až 60 dB, ucho citlivější na nízké intenzity, 13/14 bps uniformně → 8 bps logaritmicky



- = **psychologie vnímání zvuku** lidským zvukovým systémem → využití omezení pro filtraci neslyšitelných zvuků a kvantizaci
- ucho: **zvuk** jako tlakové vibrace vzduchu je přes bubínek převeden na pohyb kůstek a přes membránu na tlakové vibrace tekutiny hlemýždě snímané jako změny frekvence vlásenkovými receptory – nelineární (tvar hlemýždě, distribuce receptorů aj.), uvažované (**zjednodušeně!**) jako **1D signál**

Obrázek: Obr. Struktura ucha

- **frekvenční omezení:** slyšitelný rozsah asi 20 Hz až 20 kHz → filtrace, dynamický rozsah asi 120 dB
- **časová omezení:** různé zvuky audio signálu v čase jsou vnímané odděleně, pokud jsou aspoň asi 30 ms od sebe, jinak je vnímání jako u frekvencí

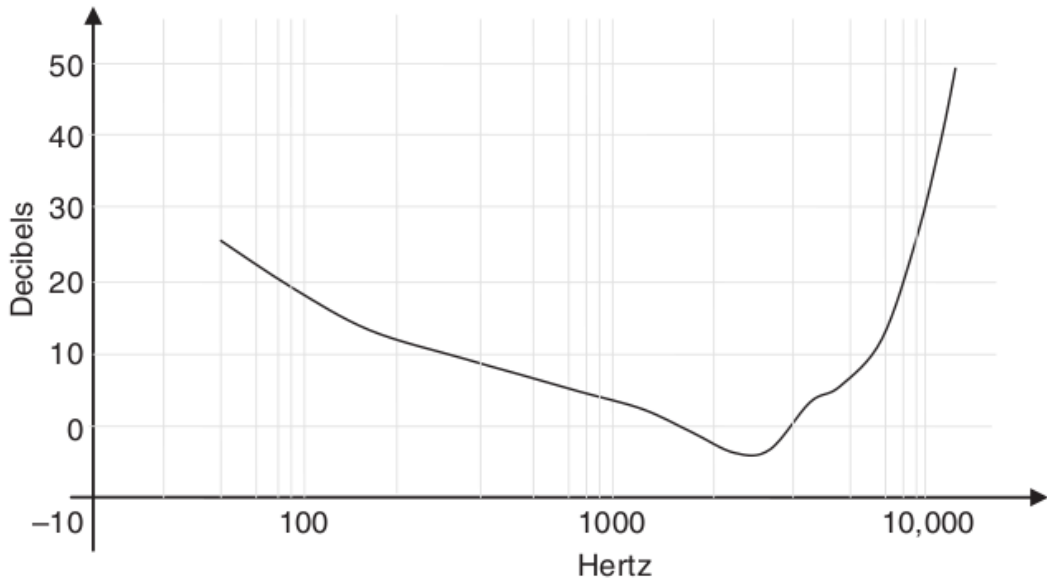


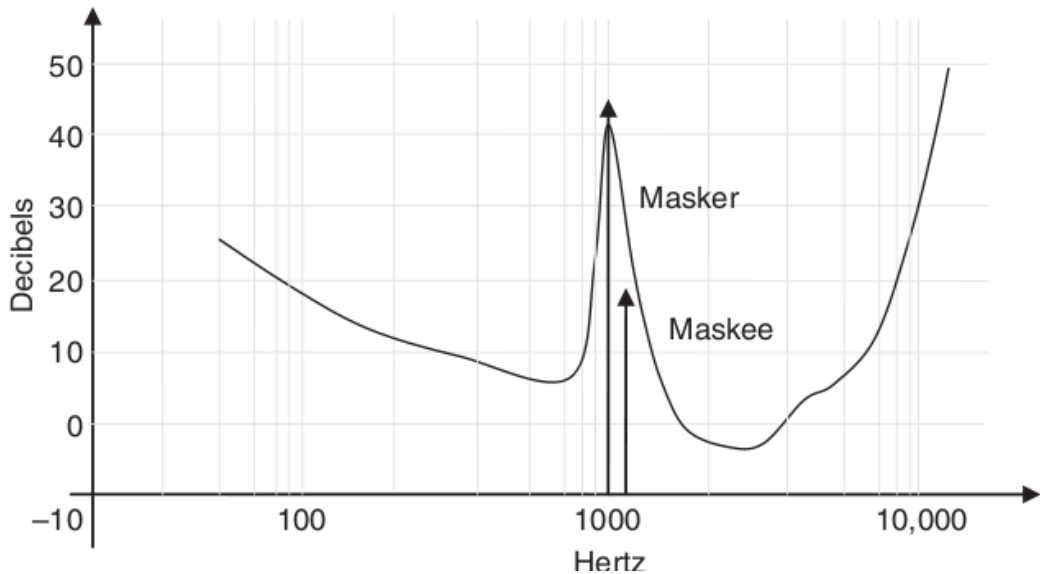
- každá frekvence (tón) má samostatně jiný **práh slyšitelnosti**

Obrázek: Obr. Práh slyšitelnosti frekvencí

- = frekvence přítomná (s nějakou intenzitou) v signálu maskuje jiné frekvence, které tak nejsou vnímány = **frekvenční maskování**
- = přítomnost frekvence (nad prahem slyšitelnosti) zvyšuje práh slyšitelnosti pro jiné frekvence – změny zjištěny experimentálně
- maskované frekvence = sousední menší (do menší vzdálenosti) a větší (do větší vzdálenosti) s menší intenzitou, rozsah maskovaných frekvencí a jejich intenzit závisí (přímo, nelineárně) na frekvenci i intenzitě maskující frekvence

Obrázek: Obr. Změna prahu slyšitelnosti v přítomnosti frekvence





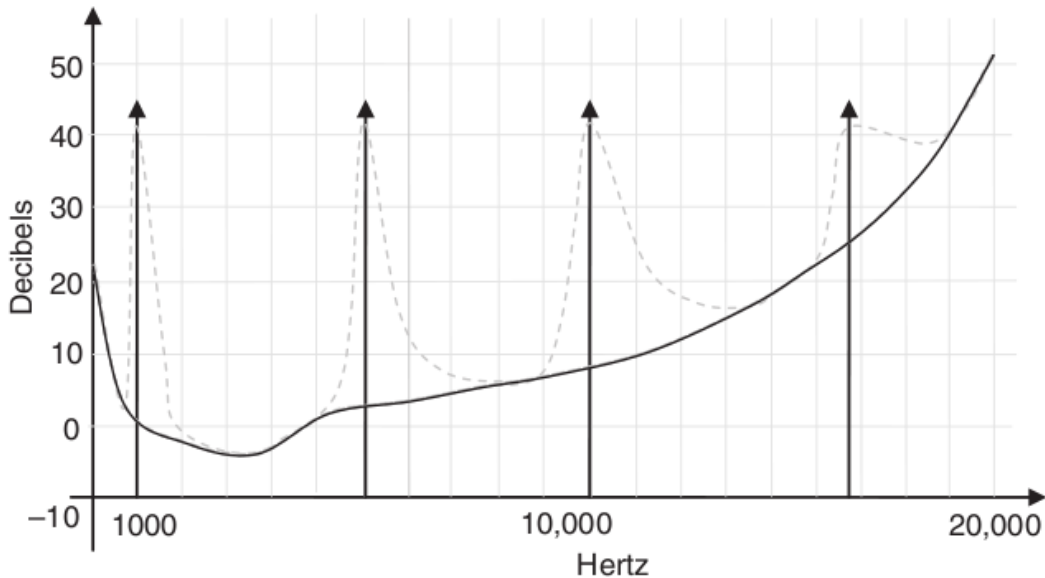
- rozsah maskovaných frekvencí je pro každou maskující frekvenci jiný

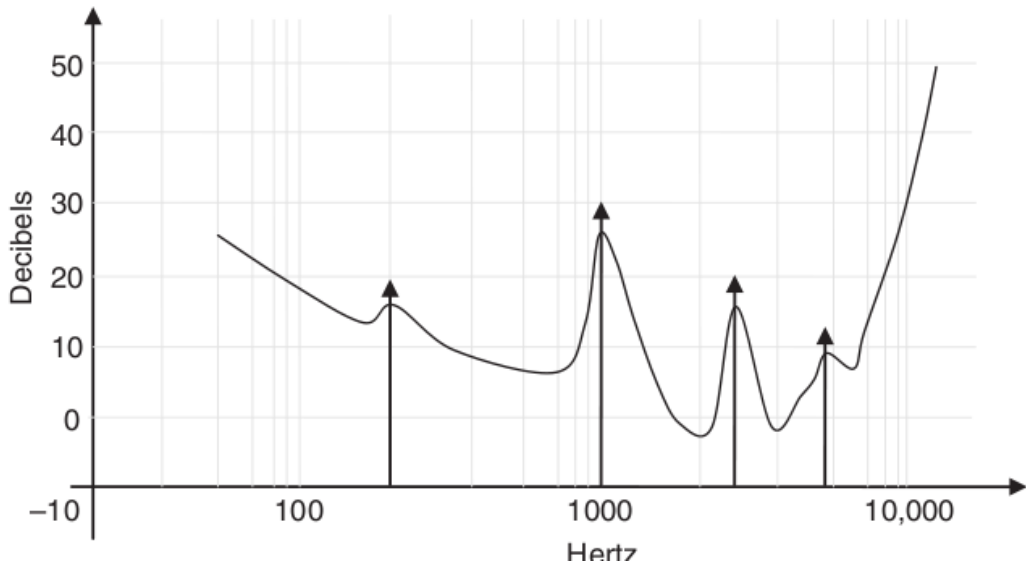
Obrázek: Obr. Změna prahu slyšitelnosti pro různé přítomné frekvence

- rozsah maskovaných frekvencí pro danou maskující frekvenci je jiný i v případě přítomnosti dalších maskujících frekvencí a práh slyšitelnosti frekvencí se celkově zvyšuje

Obrázek: Obr. Změna prahu slyšitelnosti v přítomnosti více frekvencí

- ⇒ lidský zvukový systém vytváří a filtruje **pásma** (nerozliší frekvence v jednom pásmu) – šířky pásem 100 Hz pro frekvence pod 500 Hz, pro zvyšující se frekvence se lineárně zvyšuje až do 6 kHz, zjištěno experimentálně



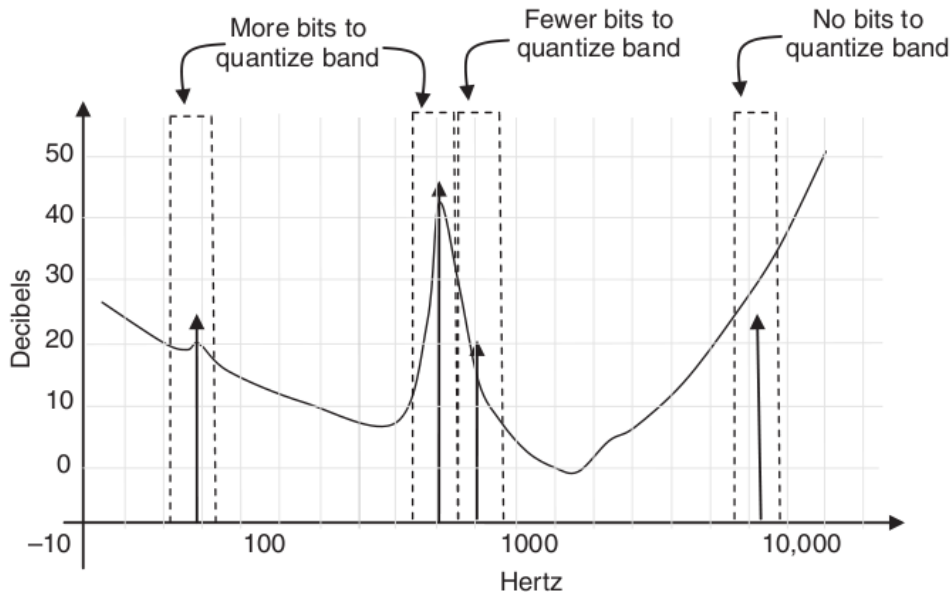




- **temporální maskování** = maskující frekvence maskuje maskované frekvence po nějakou dobu i po svém vymizení, s postupným zmenšováním rozsahu maskování, doba závisí na rozdílu a hodnotě intenzit (přímo) a vzdálenosti frekvencí (nepřímo)
- **psychoakustický model** = model maskovaných frekvencí v závislosti na přítomnosti (a vymizení) maskujících frekvencí

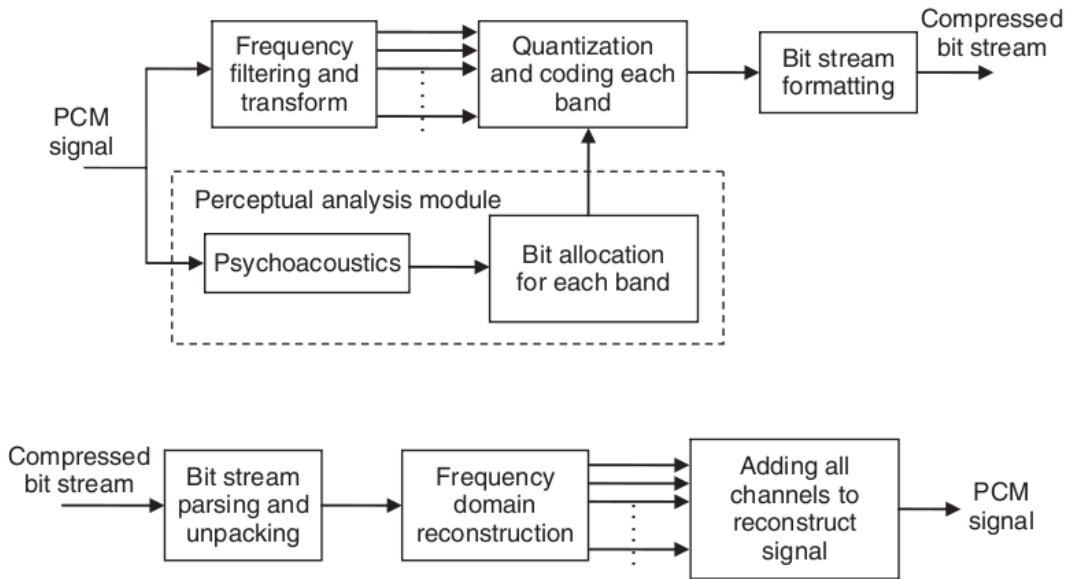
- 1 **rozdělení signálu do oken** s typicky od 512 do 4096 vzorky, nezávislé zpracování oken
- 2 **transformace z časové do frekvenční domény**
- 3 **filtrace** – odstranění neslyšitelných frekvencí a rozdělení slyšitelných **do pásem**
- 4 **paralelně kódování signálu a úpravy psychoakustického modelu** podle signálu
 - kódování signálu = **kvantizace** frekvenčních koeficientů pásem – narozdíl od statických tabulek u obrazu (dominantní = nižší frekvence) dynamická na základě **dynamického** psychoakustického modelu (dominantní = maskující frekvence, měnící se v čase)

Obrázek: Obr. Distribuce bitů v pásmech



- úpravy psychoakustického modelu = **úpravy křivky prahu slyšitelnosti** frekvencí pásem **a kvantizačních tabulek**, na základě analýzy signálu

Obrázek: Obr. Komprese a dekomprese





- komprese = analýza a **modelování (parametrizace) zdroje zvuku ze signálu**, parametry např. frekvence (výška), délka, amplituda (intenzita), periodičita, barva aj.
- dekomprese = **syntetizace zvuků** na základě modelu
- použitelné pouze pro **známé zdroje zvuku**, např. hlas, siréna
- hlas \approx fyzikální proces průchodu zvuku z plic hrdlem přes hlasivky a ústy – model rezonující trubky/píšťaly, parametry **rezonance** a residuální frekvence a amplituda

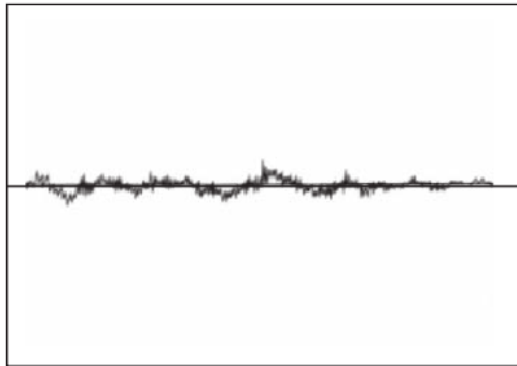
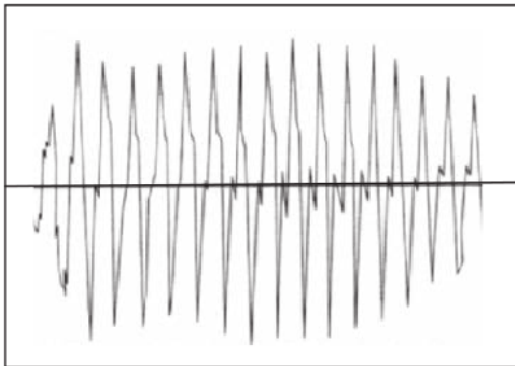
- = metoda **analýzy řeči**, použití pro imitaci (syntetizace)
- rozdělení signálu do bloků
- analýza na **hlas = samohlásky** = nízké frekvence a vyšší amplituda, a **šum = souhlásky** = vyšší frekvence a nízká amplituda

Obrázek: Obr. Samohlásky a souhlásky v řeči

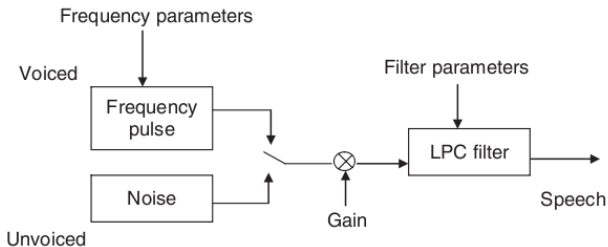
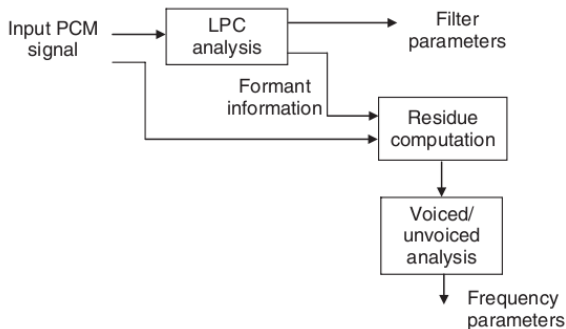
- kódování parametrů: typ, frekvence nebo šum, amplituda, filtrační pro reziduum

Obrázek: Obr. Komprese a dekomprese

Linear Predictive Coding (LPC)

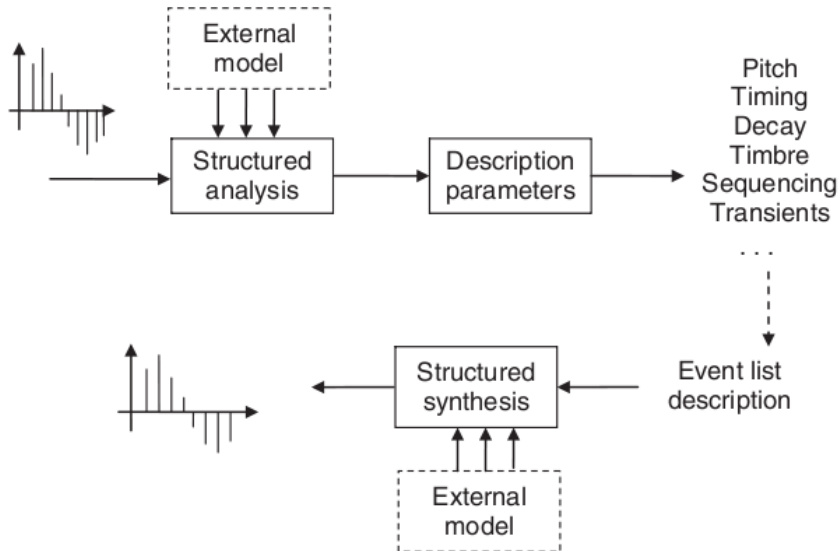


Linear Predictive Coding (LPC)



- komprese = analýza a sémantický popis strukturovaného audia jako posloupnosti **kombinací zvuků generovaných nástroji**, tj. hudby, hudební parametry frekvence (noty), časování (délka), amplituda (síla), barva (nástroj), mixážní efekty atd.
- kódování posloupnosti **hudebních parametrů**
- dekomprese = **syntetizace** zvuků podle parametrů interpretovaných/generovaných **externím modelem (nástroj, emulace oscilátory)**, např. piano

Obrázek: Obr. Komprese a dekomprese





- **aditivní syntéza** – skládání základních frekvencí, s amplitudou proměnných v čase, pro (kvazi)periodické zvuky (neimpulsivních) nástrojů
- **subtraktivní syntéza** – filtrace bohatě harmonického zvuku z oscilátorů, jako modelování hlasu (samohlásek)
- **frekvenční modulace** – modulace nosné (vyšší) frekvence pomocí modulační frekvence zvuku
- **fyzické modelování** – emulace fyzické struktury zdroje a zvuku, parametry popisují materiál a konstrukci zdroje a jeho použití, např. buben a bouchnutí



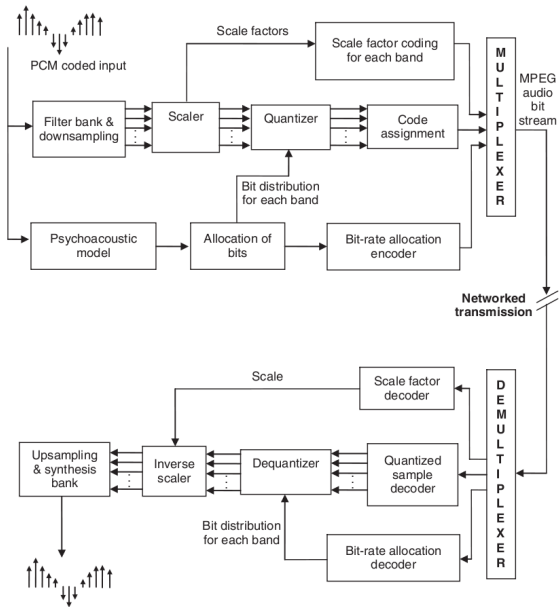
- úspornost – velice nízké bit rate
- snadná tvorba a modifikace hudby – formální vysokoúrovňový popis, flexibilita, podobné **komponování hudby**
- interaktivita – např. **automatické generování hudby** „na pozadí“ při akci člověka, např. hraní
- sémantický popis – analýza, **kategorizace** apod. **hudby**

- některé standardizované – ISO MPEG-x, ITU G.7xx, Dolby AC-x, MIDI
- spolupráce s výzkumnými institucemi a průmyslem

MPEG-1

- 3 zpětně kompatibilní **vrstvy**: **Layer I (MP1), II (MP2) a III (MP3)**
- 4 **módy**: mono, stereo, dual channel a joint (intensity) stereo
- filtrace do 32 stejně širokých pásem, podvzorkování 1 : 32, bloky s 32×12 normalizovanými vzorky a dynamická kvantizace vzorků bloků
- paralelně aplikace psychoakustického modelu pro dynamickou kvantizaci – 512/1024-prvková **FFT** u Layer I/II, jedna z 15 kvantizačních tabulek

Obrázek: Obr. MPEG1 Layer I a II



MPEG-1 Layer III (MP3)

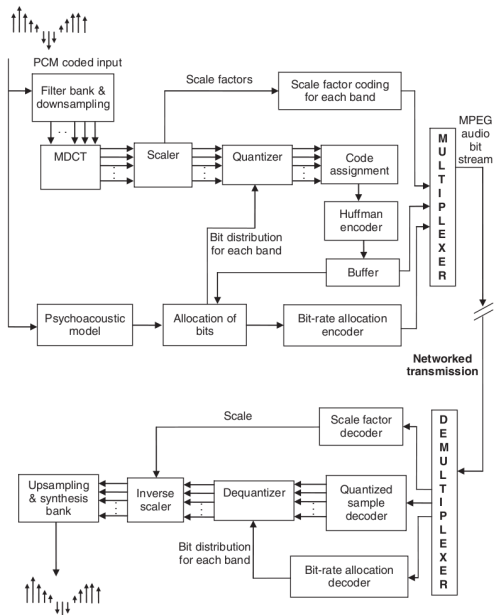
- 18-prvková **Modifikovaná DCT (MDCT)** bloků před normalizací – řešení problému nespojitosti signálu na hranicích oken (\rightsquigarrow „lupání“) pomocí 50% překrytí bloků:

$$F(u) = 2 \sum_{x=0}^{x=N-1} f(x) \times \cos \left[\frac{2\pi}{N} \left(x + \frac{N/2 + 1}{2} \right) \left(u + \frac{1}{2} \right) \right], \text{ pro } u = 0, \dots, N/2 - 1$$

- **Huffmanovo kódování** neuniformně kvantizovaných frekvenčních koeficientů \rightarrow variabilní bit rate

Obrázek: Obr. MPEG1 Layer III

- kvalita transparentní při bit rate 384/296/96 kb/s pro Layer I/II/III na kanál – sluchem nerozeznatelné od originálu

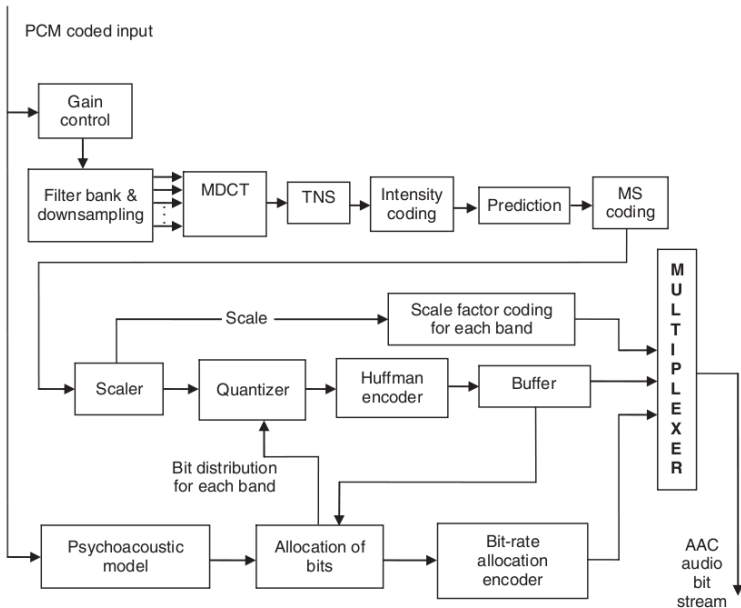


MPEG-2 (AAC)

- podpora 5.1 **surround zvuku**
- 2 skupiny: **BC** (zpětně kompatibilní s MPEG-1, ale pro různé vzorkovací frekvence) a **AAC (Advanced Audio Coding)** – pro DVD-Audio
- za MDCT: temporal noise shaping (pro redukci kvantizačního šumu), stereo intensity coding (redukce irelevancí mezi kanály pomocí kombinace kanálů ve vyšších frekvencích), predikce korelovaných sousedních bloků z předchozího, M/S coding (kódování součtu a rozdílu levého a pravého kanálu)

Obrázek: Obr. MPEG-2 AAC

- vysoká kvalita při bit rate 128 kb/s pro stereo, 320–430 kb/s pro 5.1 (méně než polovina BC)



MPEG-4

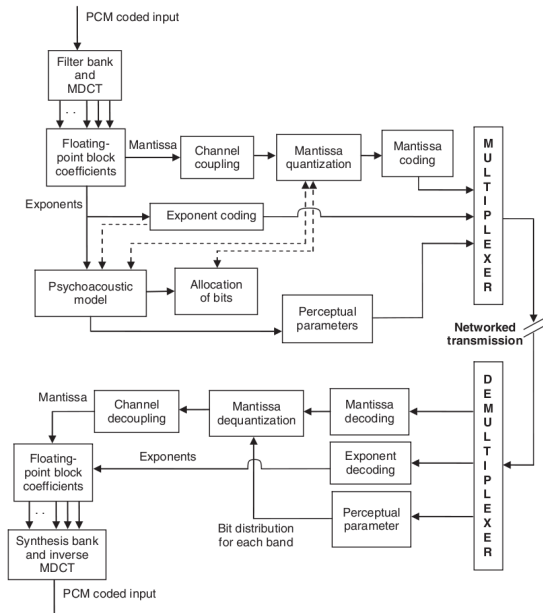
- více než 20 částí: i video, 2D/3D grafika, formát aj. („multimedia standard“), část 3 pro **řeč, hudbu i stukturované audio (SAOL a SASL)**
- řeč: **HVXC** (2 kb/s) a **CELP** (Code Excited Linear Prediction, podobné LPC), syntetizační modul
- hudba: vylepšené verze **MPEG-2 AAC – low complexity (LC, původní), high efficiency (HE, replikace pásem), scalable sample rate (SSR, dvojúrovňová 4/32 filtrace do pásem, transparentní kvalita při 64 kb/s), bit sliced aritmetické kodování (BSAC)**

Dolby AC-2/3

- reprezentace **frekvenčních koeficientů v plovoucí řádové čárce** (s více mantisami) po filtraci a MDCT
- využití exponentů v psychoakustickém modelu a kvantizace mantis – dekódovatelná z exponentů a parametrů modelu

Obrázek: Obr. Dolby AC-3

- bit rate 128–192/32–640 kb/s pro AC-2/3 na kanál



ITU G.711/G.722/G.721,6,7

- pro telefonní signál řeči a hudby vzorkovaný na 3/7 kHz, bit rate 64/48–64/16–40 kb/s
- neuniformní kvantizace PCM signálu, podobná A/ μ -law (G.711), rozdělení signálu na pásma nízkých a vyšších frekvencí kódovaných různě (G.722), ADPCM (G.721, G.726, G.727)

ITU G.723,9/G.728

- pro signál 8 kHz po telefonu (s modemem, G.723 – spolu s H.263 pro video H.324, nebo ISDN, G.728), bit rate 5–8/16 kb/s
- aplikace modelů hlasu (CELP), kombinace s ADPCM (G.728)
- **robustní** vůči ztrátám při přenosu a šumu (G.728)



- 1980 (spolu s digitálním audiem a CD), standard pro **elektronické nástroje produkující digitální/elektronickou hudbu** (syntetizátory, sekvencery aj.) – specifikace HW i SW protokolů pro komunikaci
- sémantický model strukturovaného audia, hudební parametry jako **MIDI zprávy**, systémové a kanálové – pro 16 kanálů = nástrojů, zprávy nota zapnuta/vypnuta, pitch-wheel = změna frekvence, (poly)pressure = intenzita (aftertouch, změna při notě), kontrolní (tremolo, vibrato, sustain atd.), změna nástroje aj.
- nevýhody: jen instrukce pro tvorbu zvuků (z uložených vzorků), ne digitální signál ⇒ **různý signál** z různých zařízení (jako různí hudebníci), omezení intenzity nástroje (7b) a časování (sekvenční nesoučasná interpretace zpráv např. plynule měnícího se zvuku flétny) – vhodnější pro diskrétní nástroje jako např. piano

Kompresa grafiky

- 1990, na interaktivních CD a ve hrách 2D grafika a animace, dnes narůstající využití 3D grafiky – hry, filmy (efekty, reklama), design, GUI, prezentace (modelování) informací ... → **realistické modelování objektů** z reálného světa
- (přesná) vektorové reprezentace, (lepší) možnosti interakce, skenovací a modelovací (authoring) nástroje, výkonný grafický HW ⇒ rozvoj **počítačové (3D) grafiky**
- v poslední době nárůst objemu (pro přenos/zobrazování) a složitosti geometrické reprezentace (vizuální detaily vs. čas renderování) ⇒ **potřeba komprese** – závislé na reprezentaci
- více reprezentací, nejběžnější **polygonální** pomocí sítě trojúhelníků – de facto standard
- např. postava v 3D hře až 5 tis. polygonů ~ 100 kB (10× víc u modelů v 3D efektech filmů), CAD/CAM modely až 2 mil. polygonů ~ 30 MB – jen geometrie bez obrazových textur!
- např. formáty DODELAT



- z jednoduchých tvarů (např. rovina, krychle, koule aj.) pomocí **modelovacích operací** (geometrické transformace, množinové, vyhlazení aj.)
- **skenování** = vzorkování povrchu objektu, vzorky = body sestaveny do reprezentace



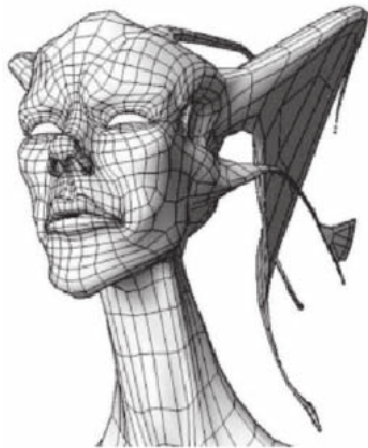
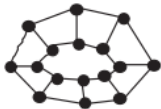
- více, v závislosti na aplikaci a požadované přesnosti
- **vektorové**, pro zobrazení **renderované do rastrového obrazu**
- body: souřadnice (x, y, z) + souřadnice normálového vektoru, barva/souřadnice v textuře, stínování, materiál atd. pro renderování
- 2D oblasti: hraniční body na řádcích souřadnicové mřížky, **quad-tree** = hierarchie nejmenších čtverců postupně dělených na 4 podčtverce obsahujících celý a pak část regionu, hraniční křivky
- křivky: (první) bod a relativní pozice dalších bodů (8/26 pozic v souřadnicové mřížce), **aproximace** propojenými polynomickými (**interpolačními** nebo **spline**) podkřivkami různých stupňů – seznam řídicích bodů (získaných iterativně přidáváním čar nebo rekurzivně dělením čar mezi body křivky)

Polygonální síť (Polygonal meshes)

= síť polygonů – seznam bodů, hran (= dvojice bodů), nebo stěn (= seznam bodů nebo hran)

Obrázek: Obr. Polygonální síť

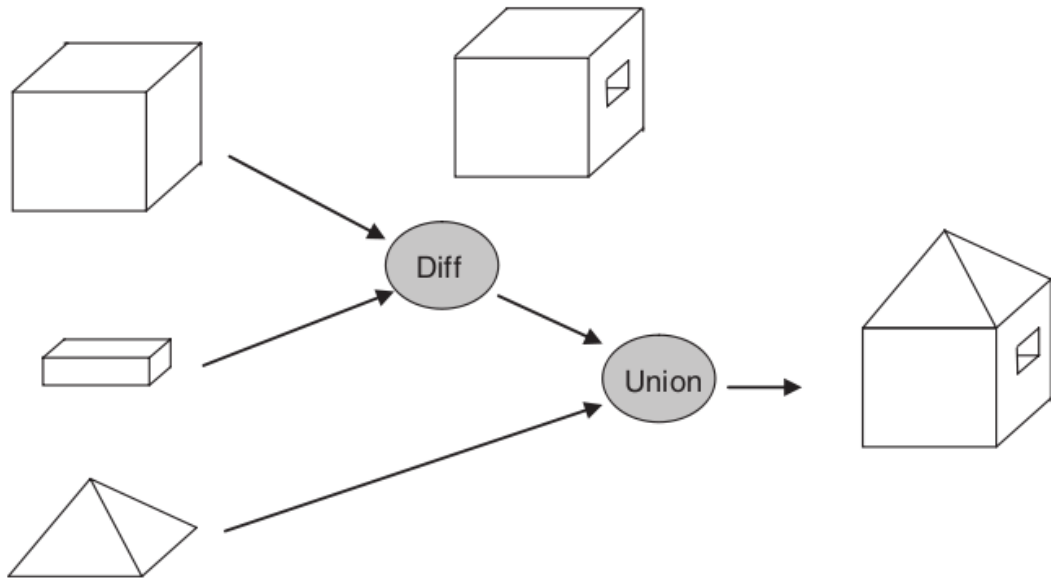
- polygony nejčastěji **trojúhelníky** (popř. čtyřúhelníky, ale lze převést) = **trojúhelníkové síť (triangular meshes)** – de facto standard pro modelování povrchů
- nejpoužívanější, podporována grafickým HW a SW (polygony = atomické jednotky)
- plochy: **aproximace** propojenými polynomickými podplochami (**Bézier**, **NURBS**) různých stupňů – řídící polygonální síť



Konstruktivní geometrie těles (Constructive Solid Geometry)

- pro **přesné modelování** těles v CAD/CAM, nevhodné pro animace
- = úpravy základních parametricky popsaných těles (např. krychle, koule, kvádr, kužel, jehlan atd.) pomocí **množinových operací**

Obrázek: Obr. Konstruktivní geometrie těles

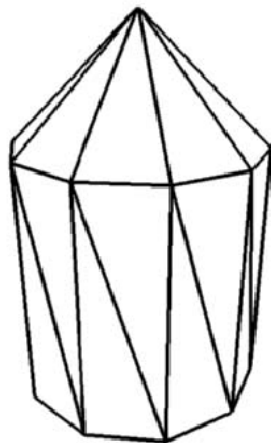
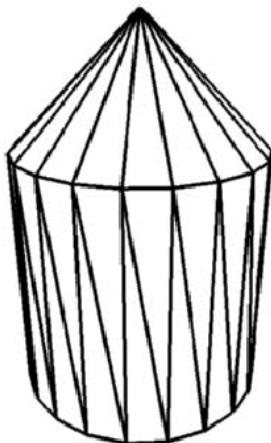
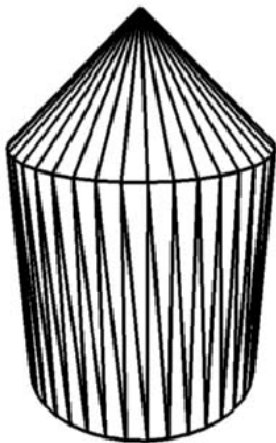


- relativně nová oblast (nárůst objemu a složitosti 3D grafiky), reprezentace objektů trojúhelníkovou sítí
- informace: **pozice a propojení bodů do trojúhelníků**
- redundance:
 - **reprezentační** – vícenásobné reprezentace společných hran a bodů trojúhelníků
 - **povrchová** – body nepřispívající k detailnější aproximaci (při skenování), např. příliš trojúhelníků v rovině

- bezztrátová (přesnost v CAM/CAD) a ztrátová (zobrazování, animace):
 - **komprese parametrů** – kvantizace parametrů bodů, RLE posloupností bodů
 - **komprese kódování konektivity** – redukce informací o propojení trojúhelníků, využití koherence, reprezentační, např. topological surgery
 - **zjednodušení mnohostěnu** = zrušení bodů, hran a trojúhelníků při zachování (vnímání) geometrického tvaru (minimalizace zkreslení), povrchová, např. progresivní síť

Obrázek: Obr. Zjednodušení mnohostěnu

- **podpásmová povrchová komprese** – kvantizace frekvence trojúhelníků na různých úrovních detailů, např. waveletová

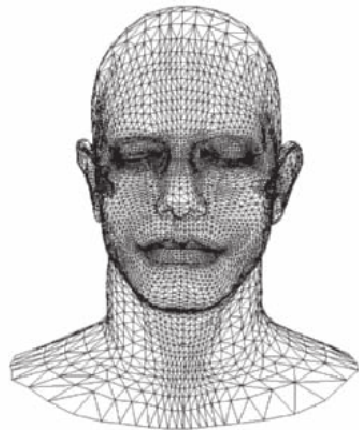
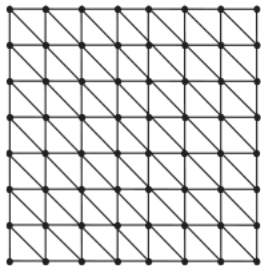


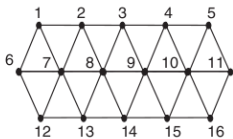
- neexistuje formule pro počet trojúhelníků jako funkce počtu bodů
- **trojúhelníkové mřížky** – strukturované (implicitní) propojení trojúhelníků, např. pro jednoduché parametrické plochy (rovina, koule), terény v 3D kartografii

Obrázek: Obr. Trojúhelníkové mřížky vs. obecná plocha

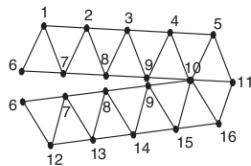
- obecné 3D plochy – potřeba **explicitní** reprezentace propojení trojúhelníků pomocí 3 vrcholů, redundance
- **pruhy trojúhelníků** – implicitní propojení (trojúhelník = tři vrcholy „cikcak“ po sobě), reprezentace seznamem vrcholů, použití např. v OpenGL

Obrázek: Obr. Pruhy trojúhelníků

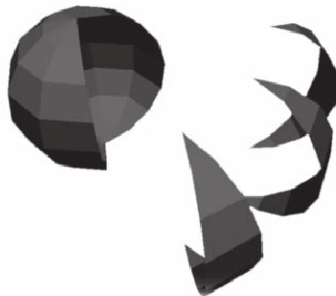
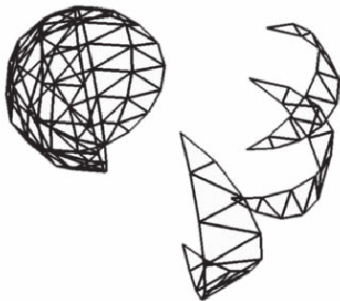




Triangle mesh specified by 16 points and 18 triangles. Each triangle is a triplet of indices (1,6,7) (7,1,2) (2,7,8) (8,2,3) ... (13,8,7) (7,12,13) (12,7,6)—a total of 54 indices.

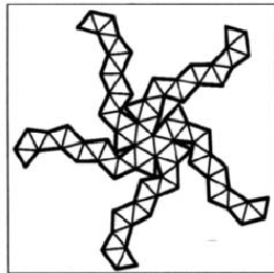
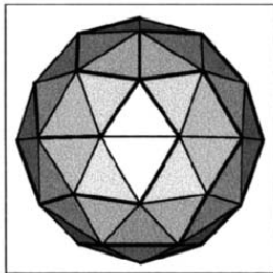
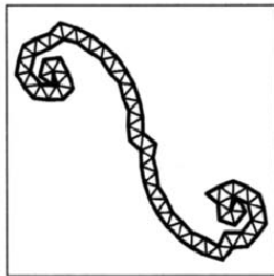
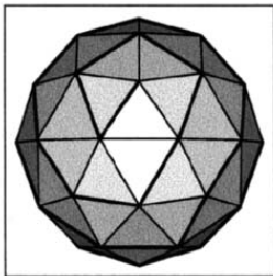


Same mesh is cut open and approximated by a triangular strip with indices 6,1,7,2,8,3,9,4,10,5,11,16,10,15,9,14,8,13,7,12,6—a total of 21 indices. The triangles are implicitly defined as a moving window of three indices—6,1,7 – 1,7,2 – 7,2,8.



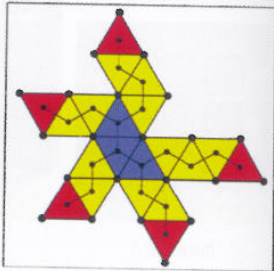
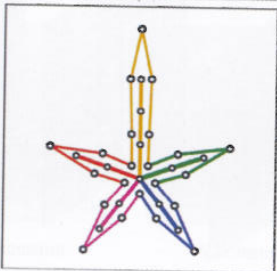
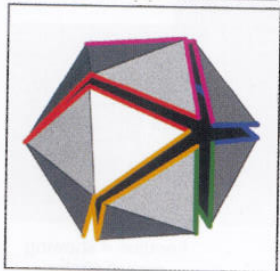
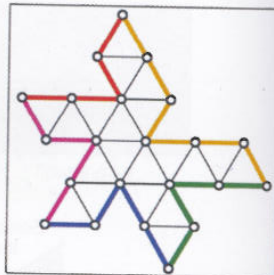
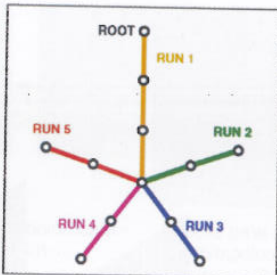
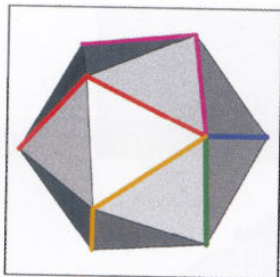
- metoda bezztrátové komprese kódování konektivity, součást MPEG-4 (3D komprese)
- = **dekompozice (rozřezání) trojúhelníkové sítě na pruhy a kódování pruhů** – ideálně 1 pruh, prakticky graf propojených pruhů
- 1** konstrukce **spanning stromu vrcholů** – **heuristické strategie** průchodu vrcholů a výběru následujícího vrcholu (vrcholy mají více sousedů), určuje míru komprese
- 2** **rozřezání sítě** podle cest ve stromu vrcholů od kořene k listům do grafu pruhů = **stromu trojúhelníků, minimalizace počtu pruhů** (a maximalizace jejich délky) u strategie průchodu vrcholů v konstrukci spanning stromu vrcholů, např. do hloubky, šířky, kombinovaný (A^* search)

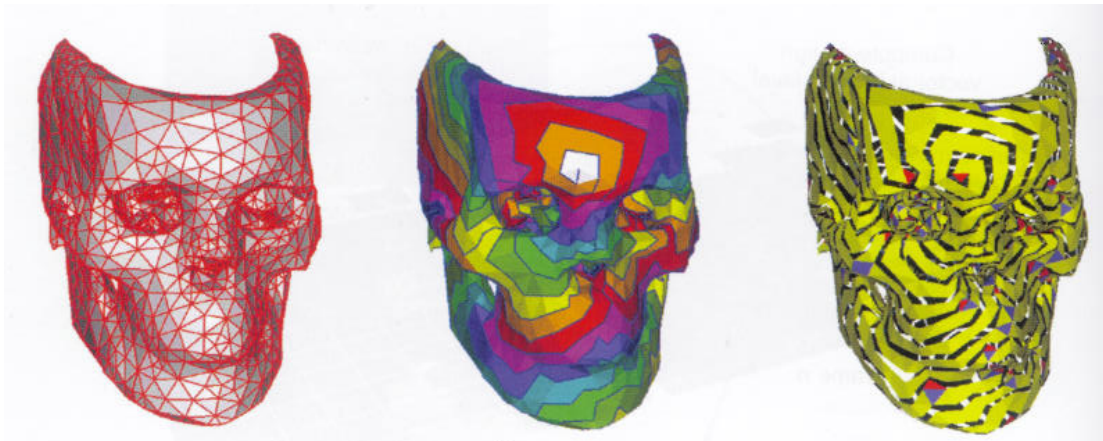
Obrázek: Obr. Strategie rozřezání sítě trojúhelníků



- 3 kódování spanning stromu vrcholů** – uspořádaný seznam vrcholů → diferenční a statistické kódování souřadnic, **minimalizace rozdílu (chyby predikce)** při výběru následujícího vrcholu v konstrukci spanning stromu vrcholů, např. nejbližší (lokální podobnost povrchu)
- 4 kódování stromu trojúhelníků** – větvící (3 sousedi), pruhové (2 sousedi) a listové (1 soused) trojúhelníky – tzv. **marching pattern**, kódování trojúhelníků jako vrcholů stromu

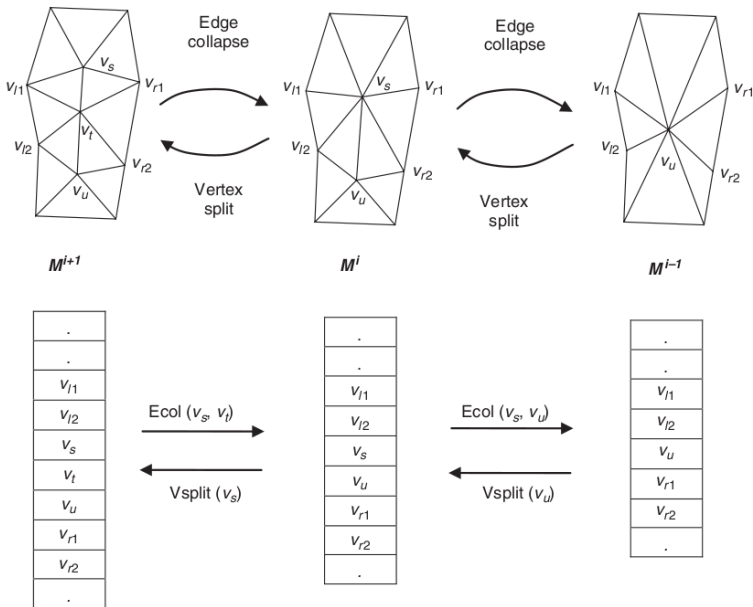
Obrázek: Obr. Topological surgery





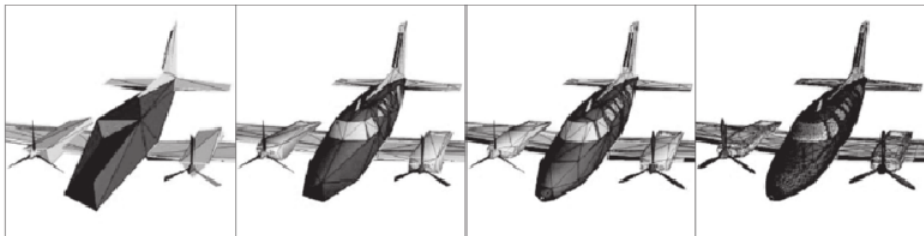
- bezztrátová i ztrátová metoda zjednodušení mnohostěnu
- = nejhrubější **aproximace trojúhelníkové sítě**, příp. spolu s informacemi pro **postupné** zjemnění do původní, transformace:
 - **rozdělení bodu (vertex split)** – přidání nového bodu propojeného s původním, případně posunutým, přidání 2 trojúhelníků sdílejících hranu, operace zjemnění pro dekompresi
 - **zhroucení hrany (edge collapse)** – sloučení bodu hrany s druhým, případně posunutým, zrušení 2 trojúhelníků sdílejících hranu, operace zhrubění pro kompresi, komplementární k rozdělení bodu

Obrázek: Obr. Transformace rozdělení bodu a zhroucení hrany



- diferenční kódování přidaného/zrušeného bodu a posunutí bodu, kódování trojúhelníků jako bodu a kombinace 2 sousedů
- **minimalizace lokálního zkreslení sítě** při výběru hrany ke zhroucení – např. nejbližší body, ale zároveň nedůležité např. pro zobrazení při dekompresi (neviditelné, vzdálené apod.)

Obrázek: Obr. Progresivní síť

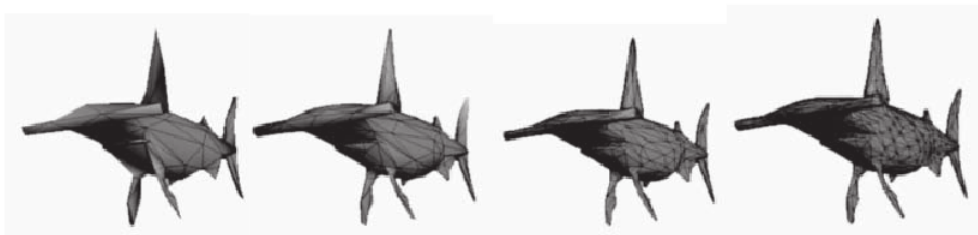


(a) Base mesh M^0 (150 faces)

(b) Mesh M^{175} (500 faces)

(c) Mesh M^{125} (1,000 faces)

(d) Original $M=M^0$ (13,546 faces)



250 faces

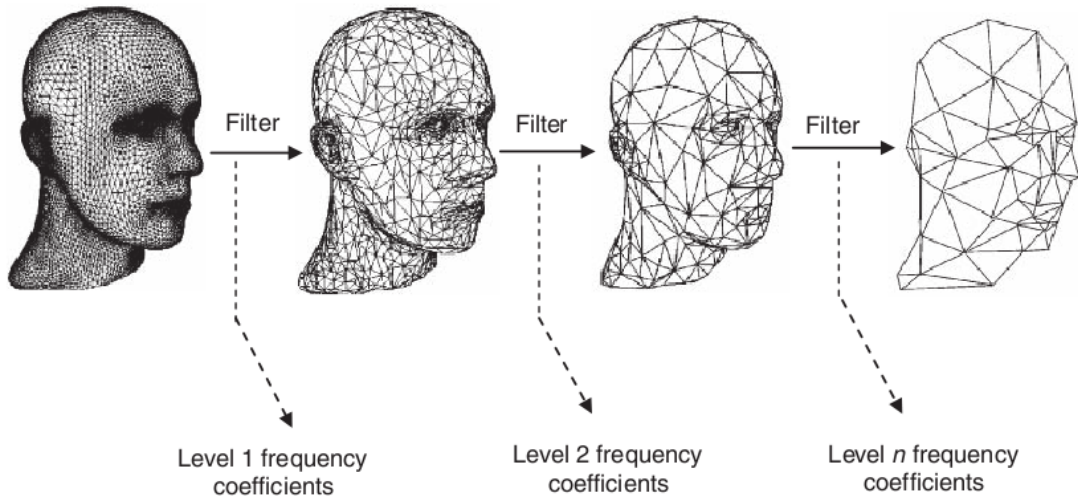
500 faces

1000 faces

2000 faces

- globální bezeztrátová i ztrátová komprese
- také hrubá aproximace trojúhelníkové sítě spolu s informacemi pro postupné zjemnění, ale 3D frekvenční **(waveletové) transformace na celé síti**
- rozdělení (filtrace) sítě na části s nízkým a postupným vyšším rozlišením (frekvencí) trojúhelníků
- kvantizace frekvenčních (waveletových) koeficientů změn mezi pásmy

Obrázek: Obr. Waveletová komprese





- = uložení/přenos **nejhrubější aproximace** objektů spolu s informacemi pro **postupné zjemnění** do detailnějších reprezentací
- pro dosažení real-time přenosu a interaktivity/renderování, např. 3D síťové hry, vizualizace (vědecké, komerční aj.), animace – práce s hrubějšími objekty, pak renderování do jemnějších
- přirozeně např. progresivní sítě (lokální zjemnění), waveletová komprese (globální)

