

Boja i zvuk u video kompresiji – osvrt

Student: Marin Petraš

Datum: 28.05.2021.

Kompresija podataka je važan čimbenik kod prijenosa video i zvučnog signala. Prijenos podataka (Bit rate) nekomprimiranog video signala je u rasponu od 270 Mb/s za SD TV (Standard Definition TV), 1,5 Gb/s za HD TV ili 3 ili više Gb/s za Ultra HD TV. Takva količina podataka prevelika je za efikasnu pohranu tj. transmisiju pa svi podaci moraju biti komprimirani tj. sažeti pomoću **perceptualnog kodiranja**. Perceptualno kodiranje uzima u obzir karakteristike ljudskih organa za primanje signala, a to su naše oči i uši, da bi se ustanovili limiti naše percepcije i shodno tome reducirali podaci koje mi ne možemo primjetiti. Te operacije izvode se prilikom transmisije video signala ili prilikom samog snimanja kamerama, fotoaparatom, mobilnim uređajima i sl. Svi oni imaju mali kapacitet pohrane i moraju izvoditi neku vrstu kompresije da bi mogli spremiti podatke. Kod transmisije tj. streamanja video-zvučnog signala moramo voditi računa o količini podataka koju šaljemo kako bi primatelj signala mogao neometano pratiti signal bez prevelikih gubitaka. Kada govorimo o sažimanju podataka direktno prilikom snimanja govorimo o kodiranju izvora (source coding) jer sažimamo analogni signal koji dolazi na senzore uređaja za snimanje, za razliku od komprimiranja podataka koje se radi u postprocesiranju. Nakon kodiranja izvora količina podataka bit će svedena na 1 do 15 Mb, ovisno o algoritmu kompresije koju koristimo – MPEG-1, MPEG-2, MPEG-4, H-264 ili H-265. Tako da jačina kompresije ovisi o video codec-u i željenoj rezoluciji slike/zvuka. Postoji reduciranje suvišnih (redundantnih) i nevažnih podataka (irelevantnih). Suvišni podaci su oni koji se ponavljaju više puta, a nevažni oni koje ljudsko oko neće primjetiti da ne dostaju. Imamo dva načina kompresije – kompresija bez gubitka (lossless) i kompresija s gubitcima (lossy). Kod kompresije bez gubitaka dijelovi koda se zamjenjuju kraćim kodovima koji ih matematički opisuju. S obzirom da bolje vidimo promjenu svijetlo-sjena nego promjenu boja, boje mogu biti reducirane, a time i količina podataka o boji. Također oko ne može vidjeti fine strukture na slici (npr. jako tanke linije, male promjene kontrasta...). Na tom principu djeluju metode redukcije podataka kao što su JPG i MPEG za video, gdje se grublje strukture tj. veći kontrasti prenose s puno većom preciznošću nego finije strukture. Takvi podaci se zovu nevažni podaci i oni se

izostavljaju u procesu kodiranja. Takav način kompresije zove se lossy tj. kompresija s gubitcima.

Video kamera daje nam izlazne signale u RGB colour sustavu. Signali se pretvaraju u luminantne i krominantne signale tj. u komponente koje se odnose na svjetlinu (luminancija – oznaka Y) i komponente koje se odnose na ton (krominancija – oznaka Cb ili Cr). Takav sustav boja označavamo YCbCr ili YUV. Konverzija se radi jednostavnim matematičkim operacijama - Y signal dobiva se zbrajanjem signala primara RGB (crveni je 30%, zeleni 59%, a plavi 11%), Cb je $0.56 * (B - Y[\text{luminacija}])$ i Cr $0.71 * (R - Y)$. Često možemo naići na podatak omjera triju brojeva - 4:4:4, 4:2:2 ili 4:2:0. Oni nam predstavljaju odnos frekvencija uzorkovanja za jednu luminantnu i dvije krominantne komponente video signala. Taj proces zovemo uzorkovanje boje (Chroma subsampling). 4:4:4 originala je rezolucija slike nakon preračunavanja iz RGB u YCbCr sustav. Prva četvorka se odnosi na veličinu uzorka (4 piksela). Druga dva broja se odnose na chroma komponente, a oba su relativna prema prvom broju i definiraju vertikalno i horizontalno uzorkovanje. Kod redukcije boje uvijek govorimo o redukciji chroma podataka, a luminacija ostaje ista. Kod omjera 4:2:2 reduciramo svaki horizontalni piksel (horizontalno sampliranje). Spajanjem luma i chroma komponente dobivamo 4 različite boje s 8 različitih svjetlina. Kod 4:2:0 imamo još veći redukciju – izbacuje se svaki drugi piksel s vertikalne osi (reduciranje slike 4 puta). Spajanjem luma i chroma dobivamo 2 boje i 4 različite svjetline. Dakle, video signal može podnijeti puno veću kompresiju slike zbog tromosti oka da primjeti sitne razlike u detaljima dok se dešava pokret. Kod statične slike moramo dobro pogledati detalje slike kako bi primijenili odgovarajuću kompresiju.

Slični principi funkcioniraju i kod zvuka. Ljudsko uho ima dinamički raspon od otprilike 140 dB, s rasponom frekvencija od 20 Hz do 20.000 Hz. Zvukovi iznad 20.000 Hz su ultrazvukovi, a ispod 20 Hz infrazvukovi. Audio signal koji želimo transmitirati digitalnim putem mora imati karakteristike koje poštuju raspone čujnosti ljudskog uha. Najveća osjetljivost uha je u rasponu od 3 do 4 kHz.

Analogni signal snima se različitim uređajima te se putem filtera limitira njegov raspon prije nego što se on digitalizira. Taj proces zove se sampliranje ili uzorkovanje signala. Analogni zvuk koji je kontinuiran pretvaramo u uzorke koji su mjerljivi (analogni signal nije mjerljiv, no preko filtera ga činimo mjerljivim) te na kraju dobivamo njihov konačan broj koji još nazivamo i

indiskretni broj uzoraka. Sampling rate ili brzina uzorkovanja mjeri se kao broj snimljenih audio uzoraka unutar jedne sekunde (sample per second). Kada govorimo o videu, ta mjera slična je brzini izmjene sličica u sekundi (frame per second). Za audio signal najmanji broj uzoraka po sekundi je 8.000 Hz. Ako koristimo manju frekvenciju, zvuk će se činiti isprekidanim, slično kao što bi se i slika trzala pri manjem frame-rate-u. Dakle, sampling rate je mjera na x-osi zvučnog vala koja nam govori kolika je finoća uzoraka u jedinici vremena.

Druga mjera o kojoj govorimo kod zvučnog signala je sample depth tj. preciznost uzorka. Mjeri se u bitovima po uzorku, a jednako je tako važna kada pričamo o kvaliteti zvuka. Određuje nam koliko mogućih razina amplitude može postići zvučni signal. Dakle, ona nam predstavlja finoću podjele jačine zvuka na y-osi.

Treća veličina je bit rate, koji je kod audio signala produkt preciznosti uzorka (Sample depth) pomnožen s brzinom uzorkovanja (Sample rate). Mjerna jedinica je bit po sekundi. Komprimirani audio signal uvijek ima bit rate manji od umnoška Sample rate-a i Sample depth-a tj. frekvencije i amplitude koje ljudsko uho ne može čuti režu se i komprimiraju. Najčešće korišteni sampling rate-ovi su: 8 kHz kod telefonije, 44.1/48 kHz kod TV-a/CD-a, a 96/192 kHz kod blue-ray-a. Sampling rate se čini visok (s obzirom da čovjek čuje do 20 kHz), no Nyquist-Shannonov teorem kaže da se kontinuirani izvor signala može uzorkovati i savršeno rekonstruirati stih uzoraka najmanje dvostrukom frekvencijom od najveće moguće frekvencije u signalu. Nyquistov teorem tvrdi da kako bi se spriječio gubitak informacije kada se signal digitalno uzorkuje, sample rate mora biti najmanje dvostruke veličine od najveće moguće očekivane frekvencije signala. Ljudsko uho ima najveći domet frekvencije do 20 kHz pa je 44 kHz uzeto kao prikladna frekvencija uzorkovanja. Veće frekvencije se uzimaju samo zbog dodatne obrade signala kod koje može doći do gubitaka (veća frekvencija osigurava bolju kvalitetu zvuka nakon obrade).

