

Videotechnika gyakorlat III.

2015

1. Objektív video minőségi metrikák

1.1. PSNR

A legelterjedtebb videominőségi metrika. A négyzetes középhibából (MSE) származtatható

$$MSE = \frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} [A(i, j) - \hat{A}(i, j)]^2 \quad (1)$$

ahol A egy $M \times N$ -es képmátrix, és \hat{A} annak kódolt/tömörített változata. Ebből a PSNR

$$PSNR = 10 \cdot \log_{10} \left(\frac{MAX_I^2}{MSE} \right) \quad (2)$$

$$= 20 \cdot \log_{10} \left(\frac{MAX_I}{\sqrt{MSE}} \right) \quad (3)$$

$$= 20 \cdot \log_{10} (MAX_I) - 10 \cdot \log_{10} (MSE) \quad (4)$$

ahol MAX_I a maximális dinamikatartomány (amplitúdó), az adott bitmélység mellett. Jellemzően $MAX_I = 2^b - 1$ b bites ábrázolás esetén.

Jellemzően 30 dB PSNR fölött elfogadható, 40 dB PSNR fölött jó/kiváló kéminőségről beszélhetünk. A PSNR érték a szubjektív minősítési tesztek eredményeivel 0.5-0.8-as korrelációt mutat.

Színes képek/videók minősítésére elterjedten használt a következő súlyozott metrika:

$$WPSNR = 0.8 PSNR_Y + 0.1 PSNR_{Cb} + 0.1 PSNR_{Cr} \quad (5)$$

1.2. SSIM

A PSNR mellett a másik elterjedt, és a szubjektív minősítési tesztek eredményeivel jelentősen jobb korrelációt (0.9 fölötti korreláció) mutató metrika az Structural Similarity Index. Az MSE alapú hibamértékek esetén azt feltételezzük, hogy az érzékelt minőségromlás szoros kapcsolatban áll a hibajel (pl. négyzetes különbség) észlelésével. A legtöbb perceptuális objektív metrika különböző súlyozást használ a hibajelre, attól függően, hogy a hibajel különböző komponensei (pl. térbeli komponensek, vagy dekorrelált színsatorna komponensek) mennyire észlelhetőek. Ezen modellek különböző

pszichofizikai méréseken alapulnak, de ezen mérések általánosíthatósága erősen korlátozott. Ezért a legbonyolultabb HVS-modell alapú metrikák sem feltétlenül korrelálnak sokkal jobban a szubjektív értékelésekkel, mint az egyszerűbb modellekre épülő metrikák.

A természetes képek erősen strukturáltak: a képek szomszédos pixeljei között jelentős a korreláció, és ez a korreláció jelentős információt hordoz a képen ábrázolt objektumok térbeli struktúrájáról. Az emberi látórendszer alapvetően arra adaptálódott, hogy a látómezőben észlelt objektumok struktúráját feldolgozza, így a struktúrát illető információ változás a képminőség romlásának jó becselője lehet.

A látómezőben észlelt objektumok felületéről a szemünkbe érkező fény spektrális eloszlása a környezeti megvilágítás spektrális eloszlásának és a felület reflektanciájának szorzata, de az agy által értelmezett struktúra független a környezeti megvilágítástól. Ezért ha a strukturális információ kinyerésére törekszünk, akkor a környezeti megvilágítás hatását el kell különítenünk a képi információban. A strukturális információ a képi adat alapján úgy definiálható, hogy függetlenítjük a képi adatot az átlagos fénysűrűségtől, és az átlagos kontraszt értéktől.

A képi hasonlóság ezután három attribútum alapján jellemezhető: fénysűrűség, kontraszt és strukturális hasonlóság. A megvilágítás hatása az átlagos fénysűrűséggel jellemezhető. A strukturális információ kinyeréséhez kivonjuk a kép fénysűrűségéből az átlagos fénysűrűséget, majd az átlagos kontraszttól való függetlenítés céljából a képi adatot elosztjuk a szórással, vagyis normalizáljuk egységnyi szórásértékre.

Két kép megfelelő $N \times N$ -es blokkja közötti hasonlóságot tehát a blokkon belüli világosságértékek (fénysűrűség) statisztikai jellemzői (μ_x és μ_y várható érték, σ_x^2 és σ_y^2 szórásnégyzet és a két jel σ_{xy} kovarianciája) alapján becsülhetjük meg, három jellemzőt tekintve: Világosság- (l), a kontraszt- (c) és a strukturális (s) hasonlóság:

$$l(x, y) = \frac{2\mu_x\mu_y}{\mu_x^2 + \mu_y^2} \quad (6)$$

$$c(x, y) = \frac{2\sigma_x\sigma_y}{\sigma_x^2 + \sigma_y^2} \quad (7)$$

$$s(x, y) = \frac{\sigma_{xy}}{\sigma_x\sigma_y} \quad (8)$$

ahol

$$\sigma_{xy} = \frac{1}{N-1} \sum_i^N (x_i - \mu_x)(y_i - \mu_y) \quad (9)$$

Ezek szorzatából már adódik egy hasonlósági index, de numerikusan instabil a metrika, ha bármely nevező közel van a nullához.

$$\text{SSIM}(x, y) = l(x, y) c(x, y) s(x, y) \quad (10)$$

Ezért a gyakorlatban használt alakja:

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (11)$$

ahol, $c_1 = (k_1L)^2$ és $c_2 = (k_2L)^2$ és L a maximális dinamikatartomány $L = 2^b - 1$ b bit esetén.

Az SSIM indexet mozgó ablakkal (általában 8x8-as ablakmérettel) átlagolva a teljes képre kiszámítják.

Értéke 0 és +1 között változik, +1 érték esetén a két kép megegyezik.

Hasonlóan a WPSNR-hez, hasonlóan heurisztikus módon definiálhatjuk a WSSIM-et színes képek minőségromlásának minősítésére:

$$WSSIM = 0.8 SSIM_Y + 0.1 SSIM_{Cb} + 0.1 SSIM_{Cr} \quad (12)$$

2. Video tömörítések vizsgálata

A gyakorlat során az MPEG-2 (Long-GOP és Intra), valamint a H.264/AVC (Long-GOP és Intra) tömörítési eljárásokat fogjuk megvizsgálni, alapvetően a következő szempontok alapján:

- Szekvencia átlagos PSNR értéke a videóbitsebesség függvényében (mind a 4 kódolási típusra), néhány tipikus (rövid) tesztszekvencia esetére
- Szekvencia átlagos SSIM értéke a videóbitsebesség függvényében, néhány tipikus (rövid) tesztszekvencia esetére
- Ugyanazon szekvenciák esetén azonos PSNR, SSIM illetve szubjektív képminőséghez szükséges bitsebesség a 4 kódolási típus esetében
- Ugyanazon szekvenciák esetén Long-GOP struktúra beállításainak (GOP hossza, B képek száma, mozgáskeresési módszer, és annak pontosság, stb.) hatása a szubjektív képminőség mellett a PSNR (és SSIM) értékekre, és a bitsebességre
- A H.264 kódoló összetett statisztikai outputjának elemzése különböző tesztszekvenciákra

2.1. Az FFMPEG MPEG-2 kódolójának beállításai

```
ffmpeg -i inputfile.avi -v:c mpeg2video parameterek outputfile.m2v
```

Lehetséges paraméterek:

-bf : B képek száma

-g : GOP méret

-b:v : Videó bitsebesség (bit/s - ban !), vagy pl. 3000K és 3M is használható

-qpel : 1/4 pixeles mozgásbecslés

-qprd : Rate/distortion alapú optimalizálás a qpel használatánál

-intra : Csak intra kódolás

-qmax : Maximális kvantálási skálafaktor (VBR kódolás esetén) (1-31)

-qscale: Kvantálási skálafaktor (1-31)

-qmin : Minimális kvantálási skálafaktor (VBR kódolás esetén) (1-31)

-minrate: Minimális videó bitsebesség (bit/s - ban !)

-maxrate: Maximális videó bitsebesség (bit/s - ban !)

CBR / VBR : Vajon hogyan definiáljuk ez alapján ?

-psnr: psnr kijelzés kérése

-me_method xxx: mozgáskeresési módszer

-me_range: Mozgáskeresés max. sugara (pixelben)

dia: (diamond) a legegyszerűbb és leggyorsabb kereső, 4 irányba indul el, 1-1 pixelt és kiválasztja a legjobb vektort

hex: (hexagon) a fentihez hasonló de lényegesen hatékonyabb, 6 pontot vizsgál meg 2 pixel távolságban

umh: (uneven multi-hex) a hexnál lényegesen lassabb, mert egy complex hexagon alakú mintán halad végig a nehezebben megtalálható vektorok érdekében. Az me_range parameter szabályozza, milyen messzire mehet

esa: (exhaustive) intelligens keresőalgoritmus, mely a me_range opció által határolt területet teljesen végigkeresi, az umh-nál jóval lassabb ám nincs gyakorlati előnye

full : Teljes mozgáskeresés

2.2. Az FFMPEG libx264 kódolójának fontosabb beállításai

- crf (crf): Konstans bitsebesség-faktor tömörítő használata (Constant Ratefactor), a célzott minőséget lehet paraméterként megadni. (0-51) 0: veszteségmentes tömörítés

- qp (qp): konstans kvantálási skálafaktor

- b : bitsebesség állítása

- bf : egymást követő B képek száma

- qmax (qpmax) Maximális kvantálási skálafaktor

- qmin (qpmin) Minimális kvantálási skálafaktor

- refs : referenciaképek száma

- me_range : Mozgáskeresés tartománya

- me_method : Mozgáskeresés módszere:

dia (dia) Diamond search with radius 1 (fastest). epzs is an alias for dia.

hex (hex) Hexagonal search with radius 2.

umh (umh) Uneven multi-hexagon search.

esa (esa) Exhaustive search.

tesa (tesa) Hadamard exhaustive search (slowest).

- subq : A szubpixel mozgásbecslés finomságát határozza meg. 0-9 között állítható, az alapérték a 7.

0. fullpel only

1. QPel SAD 1 iteration

2. QPel SATD 2 iterations

3. HPel on MB then QPel

4. Always QPel

5. Multi QPel + bi-directional motion estimation
6. RD on I/P frames
7. RD on all frames
8. RD refinement on I/P frames
9. RD refinement on all frames

2.3. x264 kódoló kimenteti információi

Tipikus x264 kimenet a kódolást követően (magyarázat mindig a kimeneti szekció alatt):

```
avis [info]: 1280x720 @ 1.77 fps (40997 frames)
yuv4mpeg: 640x480@30/1fps, 0:0
x264 [info]: 352x288 (given by file name) @ 25.00 fps
```

A bemeneti forrástól függően a fenti sorok egyike jelenik meg, és az avisynth, y4m illetve yuv bemenetre vonatkozik.

```
x264 [info]: using cpu capabilities: MMX2 SSE2Fast SSSE3 Cache64
Az x264 jelöli, hogy milyen elérhető processzor parancskészleteket talált.
x264 [info]: profile High, level 4.0 bit depth 9
```

Információ az alkalmazott H.264 profilról és szintről ami a tömörítési konfigurációs beállítások eredménye. 8 bitnél magasabb csatornánkénti kvantálás esetén ez is kijelzésre kerül.

```
x264 [info]: frame I:461 Avg QP:16.16 size: 51489 PSNR Mean Y:45.96 U:50.20
V:49.59 Avg:46.85 Global:44.69
x264 [info]: frame P:11836 Avg QP:17.35 size: 28198 PSNR Mean Y:42.89 U:47.43
V:46.95 Avg:43.81 Global:42.82
x264 [info]: frame B:28700 Avg QP:16.99 size: 11522 PSNR Mean Y:41.85 U:45.69
V:45.32 Avg:42.68 Global:42.27
```

A három képtípusra vonatkozó statisztikai adatok: a kódolás során készített összes képkocka száma, az átlagos makroblokk kvantáló, az átlagos képkocka méret és azkapcsolódó PSNR statisztikák.

```
x264 [info]: consecutive B-frames: 3.5% 10.8% 20.6% 28.2% 16.8% 4.5% 15.6%
Az egymást követő Bframeek száma és eloszlása, egytől sorozatban.
x264 [info]: mb I I16..4: 18.3% 52.4% 29.2%
x264 [info]: mb P I16..4: 3.6% 9.4% 3.0% P16..4: 45.3% 18.5% 10.0% 0.0%
0.0% skip:10.3%
x264 [info]: mb B I16..4: 0.3% 1.0% 0.2% B16..8: 35.2% 1.4% 1.6%
direct: 6.1% skip:54.3% L0:50.6% L1:43.0% BI: 6.3%
```

A három makroblokk típusra vonatkozó statisztikák: azt jelölik, hogy a bennük lévő partíciók mekkora része I illetve natív. Az I partícióra vonatkozó 3 szám 16x16, 8x8 and 4x4 méreteket jelöl, míg a P és B partícióknál található 5 szám 16x16, 16x8/8x16, 8x8, 8x4/4x8 and 4x4 méretekre vonatkozik. A direct érték a skip vektort használó, residual nélküli makroblokkok aránya, míg a skip érték a residual skip vectoros makroblokkok aránya.

```
x264 [info]: final ratefactor: 27.45
```

Csak ABR menetnél kijelzett érték, a kódolás átlagos kvantálási paramétereit jelöli

```
x264 [info]: 8x8 transform intra:59.0% inter:63.9%
```

```
x264 [info]: direct mvs spatial:99.3% temporal:0.7%
```

Képkockák aránya a különböző direct/skip vector számítási metódus tekintetében.

```
x264 [info]: ref P L0 46.2% 18.9% 13.6% 7.7% 6.9% 6.6%
```

```
x264 [info]: ref B L0 51.2% 20.8% 13.7% 9.2% 5.2%
```

```
x264 [info]: ref B L1 84.7% 15.3%
```

Makroblokkok aránya a különböző képtípusoknál attól függően, hogy mennyi referenciát használnak.

```
x264 [info]: coded y,uvDC,uvAC intra: 39.9% 60.7% 56.2% inter: 29.3% 47.8% 37.6%
```

Intra és inter makroblokkok aránya világosságjel, színkülönbség jel DC és AC szerint

```
x264 [info]: i16 v,h,dc,p: 47% 42% 8% 3%
```

```
x264 [info]: i8 v,h,dc,ddl,ddr,vr,hd,vl,hu: 3% 87% 7% 0% 0% 1% 0% 0%  
1%
```

```
x264 [info]: i4 v,h,dc,ddl,ddr,vr,hd,vl,hu: 24% 27% 18% 6% 5% 5% 4% 5% 7%
```

```
x264 [info]: i8c dc,h,v,p: 33% 51% 12% 4%
```

Kijelzi a különböző intra becslő módok statisztikáit 4x4, 8x8 világosságjel, 16x16 és 8x8 színkülönbségjel makroblokkok szerint, ahol v - vertical h - horizontal dc - mean of corner pixels touching this macroblock ddl - diagonal down-left ddr - diagonal down-right vr - vertical-right hd - horizontal-down vl - vertical-left hu - horizontal-up p - plane

```
x264 [info]: Weighted P-Frames: Y:5.2%
```

Súlyozott P referenciaképkockák aránya.

```
x264 [info]: kb/s:237.1
```

```
encoded 300 frames, 119.24 fps, 722.37 kb/s
```

Az utolsó értékek a kódolt képkockák számát, a kódolás sebességét és bitsebességét tartalmazzák.

2.4. Feladatok

Végezzük el egy tetszőlegesen választott HD szekvencia kódolását MPEG-2 és H264 kódolással a következő bitsebességekre: 0.5, 1, 3, 5 és 10 Mbit/s mind CBR, mind VBR kódolással. Szubjektív módon elemezzük a kapott tömörített szekvenciákat.

Példa command-line parancs (MPEG-2):

```
ffmpeg -i lupo.mkv -c:v mpeg2video -b:v 500K lupo_mpeg2_500_Kbps.mkv
```

A nyílt forráskódú libx264 H264 kódoló számos presetet kínál a kódolási komplexitás (kódolási idő) és kódolási hatékonyság beállítására. A http://dev.beandog.org/x264_preset_reference.html oldalon tanulmányozzuk az egyes presetek által módosított paramétereket. A default preset a medium preset. Valamely alacsony bitsebességen (pl. 0.5 vagy 1 Mbits) végezzük el ugyanazon szekvencia kódolását a superfast és a veryslow presetek segítségével. Mit tapasztalunk kódolási idő és képminőség tekintetében.

Példa command-line parancs (H264):

```
ffmpeg -i lupo.mkv -c:v libx264 -preset veryslow -b:v 500K lupo_h264_500_Kbps.mkv
```

Alacsony bitsebesség esetén vizsgáljuk meg a különböző színkülönbségi-jel alulmintavételezési struktúrák hatását. Az alulmintavételezési struktúrák a következő módon állíthatók:

```
ffmpeg -i lupo.mkv -c:v libx264 -pix_fmt yuv444p -b:v 500K lupo_h264_500_Kbps_444.mkv
```

```
ffmpeg -i lupo.mkv -c:v libx264 -pix_fmt yuv422p -b:v 500K lupo_h264_500_Kbps_422.mkv
```

```
ffmpeg -i lupo.mkv -c:v libx264 -pix_fmt yuv420p -b:v 500K lupo_h264_500_Kbps_420.mkv
```

Mit tapasztalunk ?