

Vääräväri-ilmakuvien pakkausmenetelmien vertailu

Jukka Rahkonen
4. huhtikuuta 2012

Johdanto

Maanmittauslaitos vapauttaa 1. toukokuuta 2012 ortoilmakuvansa julkiseen käyttöön. Koko maan kattava aineisto on varsin suuri, sillä 6 x 6 km suuruisia karttalehtiä on 12840 kappaletta. Tämän suuruisen pakkaamattoman värikuvan tiedostokoko on puolen metrin pikselikoolla 420 Mt. Koko maan ilmakuva-aineiston kooksi tulee siten lähes 5400 gigatavua. On luonnollista, että näin suurilla tietomäärillä kuvien pakkaamisvaihtoehdot tulevat herättämään kiinnostusta.

Maanmittauslaitos on päättänyt käyttää ortoilmakuvien tiedostolatauspalvelussa häviöttömästi pakattua JPEG2000-tiedostomuotoa.

<http://maastotiedotauki.blogspot.com/2012/03/ortokuville-ja-laseraineistoille-uudet.html>

Häviötön JPEG2000-pakkaus pienentää tiedostokoon noin kolmasosaan alkuperäisestä. Tämä on tuntuva tilansäästö, mutta häviöllisillä pakkausmenetelmillä tiedostokoko on mahdollista pienentää vielä huomattavan paljon enemmän. Kuvien käyttötarkoituksesta riippuu, onko alkuperäiseen kuvaan sisältyneen tiedon lopullinen häviäminen siedettävää vai ei. Jos kuvaa käytetään silmävaraiseen työhön kuten kuvaruudulta digitoimiseen, niin kuvan sävyihin liittyvän tiedon häviäminen ei haittaa ennen kuin kuva sumentuu niin paljon, etteivät kuvan yksityiskohdat erotu. Väriäinen ortoilmakuva kestää yleensä vielä suhteessa 1:20 tehtävän pakkaamisen.

Ortoilmakuvien tehokkaaseen pakkaamiseen soveltuvia menetelmiä

ECW (Enhanced Compression Wavelet)

Alun perin ER Mapper -yhtiön luoma pakkausmenetelmä, joka kuuluu yrityskaupan jälkeen nykyisin Erdas-yhtiölle. Yksi- ja kolmikanavaisille kuville optimoitu pakkausmenetelmä, jonka pakkaussuhde ja kuvanlaatu on hyvä. Sovellukset käyttävät ECW-kuvia yleensä nopeasti, koska niitä luetaan aina samojen ECW-kirjastojen kautta. Avoimen lähdekoodin ohjelmissa ECW:n käyttö oli ER Mapperin aikaan varsin helppoa, mutta Erdas'in lisenssipolitiikka on tiukempi.

MrSID (Multiresolution Seamless Image Database)

LizardTech-yhtiön kehittämä pakkausmenetelmä, joka on sen verran samanlainen ECW-pakkausmenetelmän kanssa, että nämä kaksi yhtiötä kävivät monivuotisen oikeussodan siitä, kieltäkö MrSID:n käyttämä patentoitu, saumattoman loppukuvan tuottava triviaali menetelmä kaikki muut saumattoman lopputuloksen tuottavat menetelmät. ER Mapper voitti oikeudessa mutta hävisi rahansa eikä sitä enää ole. MrSID on yleisesti käytössä Yhdysvalloissa mutta ei niinkään muualla maailmassa. Myös MrSID'in käyttäminen avoimen lähdekoodin ohjelmilla on hankalaa.

JPEG2000

Kolmas vastaavanlaiseen pakkausmatematiikkaan perustuva tiedostomuoto kuin ECW ja MrSID. JPEG2000 on ISO-standardi eikä samalla tavalla riippuvainen yhden yhtiön patenteista kuin nämä kaksi muuta. JPEG2000 on kuitenkin hyvin monipuolinen ja samalla hyvin monimutkainen tiedostomuoto, eikä toistaiseksi ole ilmestynyt erityisen hyviä avoimen lähdekoodin kirjastoja JPEG2000-kuvien käsittelyyn. Ei varsinkaan suurten JPEG2000-kuvien käsittelyyn, joita ortoilmakuvat ja varsinkin ortoilmakuvamosaiikit ovat. Siitä syystä avoimen lähdekoodin ohjelmien käyttäjät ovat heikoilla myös JPEG2000-tiedostomuodon kanssa.

JPEG-pakkausta käyttävä TIFF

TIFF:in käsitetään yleensä olevan tilaa tuhlaava, pakkaamaton tiedostomuoto. TIFF on kuitenkin varsinaisesti vain kuori, joka sulkee kuvadatan sisäänsä, ja sisäisesti TIFF-tiedosto voi käyttää pakkausta. Ortoilmakuville erityisen hyvin sopiva TIFF-muoto on sisäisesti 256x256 pikselin kokosiin laattoihin jaettu kuva, jonka laatat on pakattu normaalilla JPEG-pakkausmenetelmällä. Sisäisen laatoituksen avulla isostakin kuvasta voidaan avata nopeasti ja tehokkaasti vain pieni rajattu alue, ja JPEG-pakkaus tarjoaa valokuvamaisille aineistoille hyvän pakkaussuhteen ja ihmissilmälle hyvän laadun. Sekä TIFF- että JPEG-käsittelyyn on olemassa erittäin hyviä avoimen lähdekoodin kirjastoja, minkä takia niin sanottu ”JPEG-in-TIFF” -pakkaus on esimerkiksi MapServer-WMS-palvelimen lähtöaineistoille tavallisimmin käytetty pakkausmenetelmä.

Aineistot ja menetelmät

Tähän raporttiin liittyy arkistotiedosto http://latuviitta.org/documents/ortokuvien_pakkaus.zip. Irfan View -kuvankatseluohjelma avaa useimmat näytteet, jos siihen on asennettu kaikki laajennokset.

Alkuperäinen kuva, tiedostokoko 2935 kt

1000x1000 pikselin suuruinen pakkaamaton GeoTIFF-kuva

JPEG2000test.tif

Häviöttömästi pakattu JPEG2000 kuva

Pakkaaminen tehtiin sekä YCBCR-muunnoksen kanssa että ilman sitä

JPEG2000test_lossless_YCBCR.jp2 1140 kt (39 %)

JPEG2000test_lossless_RGB.jp2 1456 kt (50 %)

Pakattu Kakadun kdu_compress-ohjelmalla, komento

```
kdu_compress -i JPEG2000test.tif -o JPEG2000test_lossless_YCBCR.jp2  
Creversible=yes
```

```
kdu_compress -i JPEG2000test.tif -o JPEG2000test_lossless_RGB.jp2
```

Creversible=yes Cycc=no

Häviöllisesti pakatut JPEG2000 kuvat

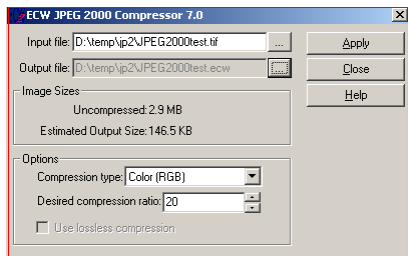
JPEG2000test_lossy_10_prosenttia.jp2	294 kt (10 %)
JPEG2000test_lossy_5_prosenttia.jp2	147 kt (5 %)
JPEG2000test_lossy_2_prosenttia.jp2	60 kt (2 %)

Häviöllisesti pakattujen JPEG2000-tiedostojen luomiseen käytetyt komennot

```
kdu_compress -i JPEG2000test.tif -o JPEG2000test_lossy_10_prosenttia.jp2 -rate 2.4
kdu_compress -i JPEG2000test.tif -o JPEG2000test_lossy_5_prosenttia.jp2 -rate 1.2
kdu_compress -i JPEG2000test.tif -o JPEG2000test_lossy_2_prosenttia.jp2 -rate 0.48
```

ECW-pakattu kuva, ECW Image Compressor 7.0, tavoitepakkaussuhde 1:20

Komento annettu sovelluksen ikkunassa näin:



JPEG2000test.ecw 154 kt (5,2 %)

Sisäistä JPEG-pakkausta käyttävät TIFF-tiedostot

JPEG-pakkauksen laatuasetus (GDAL:in oletus) 75 %, pakattu kahdella eri värikäsittelytavalla. Toinen (RGB) pitää värikanavat erillisinä, toinen (YCBCR) tekee jakaa ne luminanssi- ja krominanssikanaviin.

JPEG2000test_jpeg_in_tiff_RGB.tif	408 kt (14 %)
JPEG2000test_jpeg_in_tiff_YCBCR.tif	153 kt (5 %)

Muunnokset tehty gdal_translate -ohjelmalla seuraavilla komennoilla:

```
gdal_translate -of GTiff -co COMPRESS=JPEG JPEG2000test.tif
JPEG2000test_jpeg_in_tiff_RGB.tif
```

```
gdal_translate -of GTiff -co PHOTOMETRIC=YCBCR -co COMPRESS=JPEG
JPEG2000test.tif JPEG2000test_jpeg_in_tiff_YCBCR.tif
```

Tulosten tarkastelu

Häviöttömästi pakatun JPEG2000-tiedoston kuvan laatu on tietysti moitteeton. Tällä menetelmällä ei kuitenkaan pystytä säästämään levytilaa kuin noin 60 prosenttia.

ECW-, JPEG2000 ja YCBCR-muunnosta hyödyntävä JPEG-TIFF:ssä -pakkaus tuottavat kaikki kuvaruututyöhön riittävän hyvän tuloksen pakkaussuhteella 1:20, mikä merkitsee 95 prosentin säästöä levytilan tarpeessa. MrSID-pakkausta ei testattu, mutta voidaan perustellusti olettaa, että se käyttäytyy samalla tavalla kuin ECW- ja JPEG2000-pakkaukset.

Johtopäätökset

Vääräväri-ilmakuvia voidaan käyttää kuvaruututyöskentelyssä jopa 95 prosenttia pakattuina. Tähän pakkaussuhteeseen päästään ECW-, JPEG2000- ja JPEG-TIFF:ssä pakkausmenetelmillä. Käytettäessä JPEG-TIFF:ssä -menetelmää kannattaa ehdottomasti käyttää YCBCR-värimuunnosta. Tästä muunnoksesta on selvää hyötyä myös JPEG2000-pakkauksessa. Avoimen lähdekoodin ohjelmien käyttäjille on luvassa ongelmia kaikilla näillä tiedostomuodoilla.

Monien Javalle tehtyjen GIS-ohjelmien mukana tulee vielä vanhalla ECW-kirjaston versiolla 3.x tehty tuki ECW:lle ja se toimii yleensä hyvin. Uusimpien 3.x -version kirjastojen mukana tulee saman tien tuki myös JPEG2000:lle. Tällaisia Java-ohjelmia ovat ainakin gvSIG, Kosmo GIS ja OpenJUMP:in Plus-versio.

JPEG-Tiffissä -pakkausta Java-ohjelmat puolestaan eivät yleensä pysty avaamaan. GDAL:ia käyttäville ohjelmille, kuten Quantum GIS:lle tämä pakkausmenetelmä sen sijaan sopii hyvin.

Varoitus

Edellä esitetyt esimerkkikuvien luomiseksi käytetyt komennot eivät sovellu sellaisenaan oikeiden isojen ilmakuvatiedostojen pakkaamiseen, lukuunottamatta ECW Image Compressor 7.0 -ohjelmaa. Sekä JPEG2000-pakkauksessa että suurten TIFF-kuvien luomisessa on käytettävä lisäparametreja, jotta valmist tiedosto saadaan tukemaa nopeaa pienen osa-alueen poimimista kuvan sisältä.

Testissä käytetyt ohjelmistot

- GDAL versio 1.9
- ER Mapper ECW JPEG 2000 Compressor 7.0
- Kadadu v. 6.0 demo-ohjelmapakettiin kuuluva kdu_compress-pakkausohjelma