sievenna Testing

Onni Aarne

January 16, 2018

sievenna's Huffman coding has been tested on the Large Text Compression Benchmark [2] as well as an uncompressed tarball of the standard Calgary corpus [1].

Image compression benchmark used was an 8-bit RGB photograph called nightshot_iso_100 from imagecompression.info [3].

An uncompressed 8-bit RGB image called nightshot_iso_100.ppm was compressed from 22.128 MB to 17.843427 MB giving a size reduction of 19.36%.

The first 100 MB of the Large Text Compression Benchmark compressed down to 63.862 MB, giving a size reduction of 36.14%.

References

[1] Timothy Bell, Ian H Witten, and John G Cleary. Modeling for text compression. **ACM Computing Surveys (CSUR)**, 21(4):557–591, 1989.

[2] Matt Mahoney. Large text compression benchmark, 2011.

[3] Rawzor. The new test images. `http://imagecompression.info/test_images/`, 2008. Accessed 26.12.2017.

File	Size	Compressed	Ratio	Comp. Time	Decomp. Time
nightshot_iso_100.ppm	22.128 MB	17.843427 MB	1.240	2.654 s	1.989 s
enwik8	100 MB	63.862	1.566	9.100 s	6.900 s
enwik9	1000 MB	648.370 MB	1.542	68 s	59 s
Calgary Corpus	3.154 MB	2.125 MB	1.484	0.385 s	0.321 s

Table 1: Performance statistics for Huffman coding. enwik8 and enwik9 are different sizes of the Large Text Compression Benchmark.