Research Report on JPEG 2000 for Video Archiving

By Ian Gilmour, Media Matters

Abstract
Motion-JPEG2000 lossless data reduction is implemented as a key technology for reducing file size in storage and for reducing data rates in networked file transfer.

Aims
To determine typical values of compression or bit-rate reduction for a range of program content. To assess whether average and peak values affect accuracy of predicting data capacities for planning digitizing projects. To investigate particular characteristics which may significantly increase or reduce encoded file sizes.

Background
The primary goal for the M-JPEG2000 encoder is to maintain 100% image quality with significant reduction in file size. The coding regime is largely unique among popular CODECs in that it operates with a fixed parameter of applying a completely reversible transform, resulting in a variable reduction ratio, rather than aiming for a fixed data rate by introducing variable rates of lossy perceptual coding, according to picture complexity. Adopters of this technology require a means of predicting typical file sizes relative to the types of images which they intend to encode, so that they can plan for storage, network capacities, workflows and time-constants.

Test Method
Seven short clips were selected to provide a range of image complexities, based upon the known characteristics of the integer transform, and of the subsequent entropy coding stages of JPEG2000. The clips were already compiled onto one show-reel from one production house. They appear to have been finished on Avid Symphony or similar, so previous compression is likely, but generally high quality standard definition.

Source tape was high-band U-matic NTSC; Player: Sony VO-9800 VTR

SAMMA TBC set to 0 dB gain, 0mV black level, no dropout compensation.

Results

Title	Content	Duration	8-bit encoding			10-bit encoding		
			File Size					
		Seconds	MBytes	MB/min	GB/hr	MBytes	MB/min	GB/hr
Sol Beer	Cardboard animatic, colour, texture	30	314	628	36.80	462	924	54.14
Girl washing Dog	Daylight outdoor action, water drops	12	125	625	36.62	187	935	54.79
Haunted house	Animation	10	96	576	33.75	142	852	49.92
Bhutan	Detail outdoor composit	31	274	530.3	31.07	422	816.8	47.86
Blue Train	CG blue, white backgr'd	60	479	479	28.07	749	749	43.89
Nikon	Indoor action	30	229	458	26.84	363	726	42.54
Bass Beer	Red/amber with bubbles	30	221	442	25.90	349	698	40.90
Hendrix	Live concert + CG	290				3113	644.07	37.74

Analysis

The highest bit rates occurred on the Sol Beer ad [314Mbytes/30sec or 36.8 GBytes/hr] in which every frame was filled edge-to-edge with rich textures, complex color gradients, and fine detail. This clip also had constant movement which does not affect the JPEG2k transform, but would be a good test for comparison with MPEG GOP encoders.

For comparison with the DCT coding rates the file size of conventional JPEG coding was recorded. The first screenshot [right] is 19kBytes as a JPEG, while the 2nd screenshot [below] is 27 kiloBytes.

A Disney clip, had the 2nd highest data rate, [125Mbytes/ 12sec or 36.62 GBytes/hr] comprising mostly real life outdoor footage of girl washing dog, pluse CG overlay, so fine detail and chroma variation every few pixels in fir, trees and grass. Sky shots fairly constant luminance and chroma, with low gradients, but small percentage of total pixel count per frame, with

limited opportunity to exploit ROI capabilities of encoder. The screen-grab above was also the second largest JPEG overall at 21.3 kBytes.

Another Disney clip had the 3r^{d} highest video data rate, [96Mbytes/ 10sec or 33.75 GBytes/hr]. Image data was virtually all CG, with lots of stars and sparkly effects on dark backgrounds.

The JPEG screenshot was one of the smaller DCT files, at 13.8 kBytes.

The fourth clip, a travel ad for Bhutan had some of the most complex images, mostly outdoor and close-ups of local architecture and artifacts, all heavily composited, but overall file size [274Mbytes/31sec or 31.07 GBytes/hr] was reduced by grey borders.

This is best illustrated by the comparison of two screenshots. The full-frame image [right] was 19.1kBytes as a JPEG DCT, or roughly the same as the first Sol beer shot above, which is less complex, but sharper than the Bhutan frame. By comparison, the grey bordered Bhutan frame [below] is only 13 kBytes.

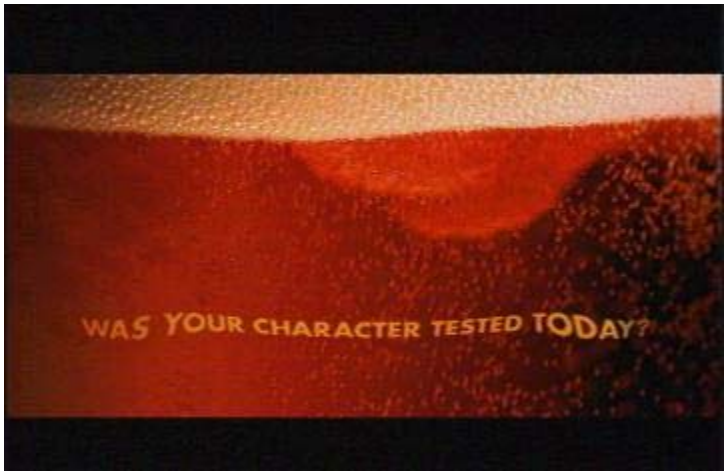
The biggest surprise was the blue train to Newark airport ad, which has a very restricted palette, large expanses of white background, and mostly line art, with no texture and limited detail in graphics and character overlays. Rather than being the lowest data rate, this 60 second clip came within the lower group, but 3rd from the bottom, at [479Mbytes/60sec or 28.07 GBytes/hr]. The screenshot JPEG at left is about as complex as the clip ever gets, and compresses to 13.7 kBytes, roughly the same as Haunted House, and Nikon.

The Nikon clip reveals the savings in encoded data through plain borders, and simple backgrounds, to a greater extent than Bhutan. A typical shot with busy P-in-P, in this frame [right] 63% of total, compresses to 13.8 kBytes in conventional JPEG, while simpler background with smaller subjects[below right] is only 10.3 kBytes.

Lossless MJ2 video was 2nd lowest at [229Mbytes/30sec or 28.07 GBytes/hr].

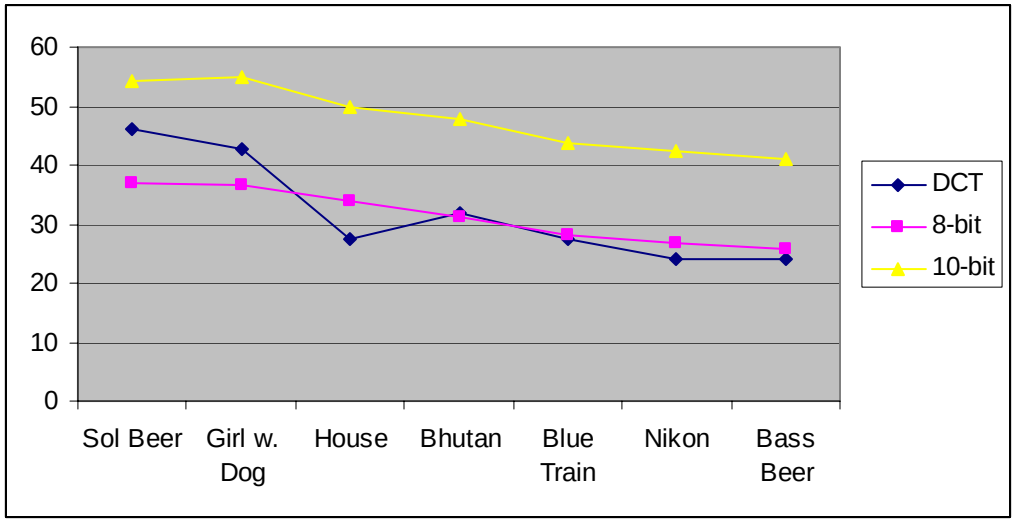
The lowest bit rate occurred on the Bass Beer ad [221Mbytes/ 30sec or 25.9 GBytes/hr] which locked in a close up of a reddish-amber liquid with lots of bubbles [maybe beer?] and overlaid swirling text. The restricted palette with limited gradients is easy for both DWT and DCT encoders, while the circular bubbles are more readily process by wavelet than by DCT coders. As a JPEG, the still frame was 12.1 kBytes, larger than Nikon3 above.

The image was also letterboxed from 16:9, and the additional black bars [roughly 25% of the total picture area] have very low percentage coding overhead.



Note that all screenshots were initially subsampled at ¼-screen resolution [320 W x 240 H] and are displayed in this document as variable rate [constant-quality] lossy JPEG.

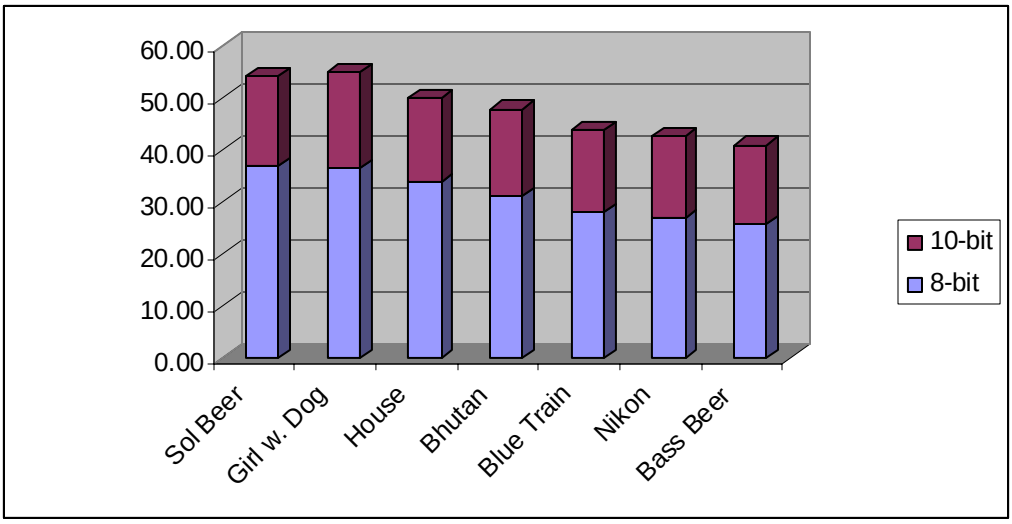
Comparisons between lossless M-JPEG2000 and the conventional, lossy DCT coding in JPEG are illustrated below. JPEG still sizes have been rescaled by a factor of 2000 for this graph. JPEG DCT works with relatively greater efficiency on the smooth gradients in the CG animated Haunted House and Bass Beer, as well as and on the matt borders and backdrops in Bhutan, Blue Train, and Nikon. DCT is less efficient on the fine detail and textures in Sol Beer and the fir/leaves/water droplets in Dog Wash.



8-bit vs 10-bit

The 10-bit files were 47% to 58% larger than the 8-bit files, rather than the 25% increase which simple math would predict. This may be due to file packing, and needs further investigation. Because of this, the detailed analysis above refers only to the 8-bit coding.

More recent tests were done using two longer video sequences. A 3-minute compilation of Jimi Hendrix live concerts and CG resulted in a 3113 MByte 10-bit file. A full hour of Lawrence Welks variety show was 30640 MB for 8-bit and 47800 MB in 10-bit.



Conclusions

The range of image content in this preliminary study of lossless JPEG encoding has provided valuable data on file sizes which generally accords well with theoretical predictions, and known behavior of integer-wavelet transforms and entropy coding. The results can be used to provide guidance for clients in planning large-scale digitizing projects, and to predict likely storage capacities, network performance and costs. In a recent test of lossless video CODECs, CS MSU Graphics & Media Lab Video Group determined typical compression rations to be in the range of 1.5:1 to 3:1[1].

In extreme cases of low image complexity, artificial signals such as black or color bars encode in as little as 3 GigaBytes per hour. With real-world programs, material which is letterboxed or surrounded by plain borders will achieve greater data reduction, as will plain backgrounds such as cyclorama shots with small subjects. Actual frame composition should be compared with theoretical predictions.

Further analysis is required on data packing of 10-bit files within the body payload, or to determine other possible reasons for the file-size of 10-bit encoding. As a benchmark, raw MJ2 streams were encoded from some clips and saved as binaries. Several clips were also encoded also as uncompressed YUV as a cross-check to determine actual compression ratios.

[1] Vatolin, D. et al 'Lossless Video Codecs Comparison' CS MSU Graphics & Media Lab Video Group. Moscow, 25 Jan 2005 http://www.compression.ru/video/