Constant Quality JPEG2000 Rate Control for Digital Cinema

Michael D. Smith[a*] and John Villasenor[b†]

[a]Consultant, Los Angeles, CA, USA;
[b]UCLA Electrical Engineering, Los Angeles, CA, USA;

ABSTRACT

An uncompressed Digital Cinema Distribution Master (DCDM) image typically has dimensions of up to 4096x2160 (4K) or 2048x1080 (2K) with 12-bit pixel data for each of the X'Y'Z' color planes. At a frame rate of 24 frames per second, this gives uncompressed data rates of 7.6 and 1.9 Gbps for 4K and 2K respectively. Even after compression, average data rates in the hundreds of Mbits/sec. are encountered. Recently, the Society of Motion Picture and Television Engineers' (SMPTE) has chosen JPEG2000 as the standard to be used for digital cinema compression. Thus, methods to appropriately trade off rate and quality for JPEG2000-compressed movies will have high importance in the next several years as systems are designed and deployed.

In this paper we describe a new distortion-based framework for rate control that enables a JPEG2000 encoder to achieve a user-specified quality, and therefore makes it possible to produce constant quality from frame-to-frame in an image sequence. The new method makes direct use of the same JPEG2000 coding pass data as the traditional approaches, and thus can easily be adopted at the back end of JPEG2000 encoding engines. We compare the new method with two other common rate control techniques for JPEG2000.

Keywords: JPEG2000, rate control, constant quality, constant distortion, digital cinema, DCI, compression, coding

1. INTRODUCTION

JPEG2000[1] is the most advanced still image compression standard and has the potential to impact still image coding over a wide range of commercial applications. The standard is very flexible and, when applied to a single image frame, offers a wide range of rate-distortion choices and enables substantially improved compression efficiency over the older DCT-based JPEG standard, particularly at low bit rates. JPEG2000 represents the end product of very significant research and standardization efforts on the part of the participating institutions and the image processing community in general, and as a result, offers rate-distortion performance that is unlikely to be surpassed in the foreseeable future, particularly if reasonable constraints on complexity are imposed.

While new opportunities for completely new still image compression methods are likely quite limited, the issue of how JPEG2000 can best be used for frame-by-frame video compression remains open. At first glance, the application of JPEG2000 to video may seem inappropriate, particularly in light of the availability of advanced video coding algorithms such as MPEG-4 and H.264 that specifically exploit the interframe redundancy found in video sequences. However, for very high rate, high quality encoding, the benefits of exploiting this redundancy are lower. In the limit of high coding rate, the bandwidth costs of coding motion compensated prediction error can approach the costs of simply directly representing the desired image content without any predictive coding.[2] In addition, when compared with still image coding, video coding of course involves significant additional computational complexity and memory associated with generating and utilizing prediction data. These factors and others have led the cinema industry to choose frame-by-frame JPEG2000 compression as the basis for digital cinema. Substantial commercial efforts are already underway to prepare for the inevitable transition to digital cinema, and algorithmic methods that can lead to lower cost, higher efficiency solutions thus will have high importance.

While there is a long history of work on rate control for traditional video encoders, there has been almost no attention paid to the issue of how to manage rate control on a video sequence in which each frame is compressed

*miksmith@attglobal.net; †villa@icsl.ucla.edu

Visual Communications and Image Processing 2007, edited by Chang Wen Chen, Dan Schonfeld, Jiebo Luo,
Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 6508, 65081B,

independently but where consistent frame to frame post-encoding quality is desired. Similarly, while there have been extensive efforts to develop rate distortion optimal approaches to wavelet still image coding, many of which have led to specific techniques in JPEG2000, those efforts have by definition been aimed at coding of stand-alone images. Even in the stand-alone image case, methods for targeting a specific post compression quality have not been a focus of attention. Thus, from a coding standpoint, the combination of JPEG2000 and digital cinema creates a unique opportunity. When bandwidth is not at a premium, satisfactory visual quality can be obtained using very simple fixed- or variable-rate coding schemes. For example, a fixed-rate scheme with a high per-frame bit allocation or a variable rate approach that targets very small residual distortions will both ensure very high visual quality. However, approaches such as this tend to utilize far more bits than are necessary. It is desirable, therefore, to have a scheme that enables constant, high quality and simultaneously makes economical use of bits.

The present paper describes such an approach, and further extends a method recently described in the SMPTE Motion Imaging Journal.[3] The SMPTE Motion Imaging Journal publication included a concept-based presentation of the algorithm framework and presented results with 4K scanned-film and 4K computer generated imagery (CGI) sequences. In this work, we describe a mathematical framework for the algorithm and present results using a 2K CGI image sequence.

2. JPEG2000 RATE CONTROL BACKGROUND

Currently available JPEG2000 encoders usually implement either "rate-based" or "efficiency-based" rate control algorithms. In this section, we provide an overview of some key building blocks of a JPEG2000 code-stream and briefly review these common rate control methods.

The fundamental unit of data in the JPEG2000 compression standard is the code-block. A code-block is simply a spatial grouping of wavelet coefficients and is chosen to be of size 32x32 for digital cinema applications. Each code-block is further decomposed into "fractional bit-planes". As the term implies, this decomposition is related to the bit planes in the binary representation of the quantized wavelet coefficients. There are typically three fractional bit-planes for each bit-plane in a code-block. The fractional bit-planes are compressed with a context adaptive arithmetic coder. Compressed fractional bit-planes are often called "coding passes", and contain the actual bits that comprise a JPEG2000 code-stream. For a 2048x1080 (2K) 3-color 12-bit digital cinema image, decomposed using an 5-level discrete wavelet transform (DWT), there are approximately (2048/32)*(1080/32)*3 $\approx$ 64*34*3 = 6528 code-blocks. The number of coding passes per code block is a function of various factors including the quantization precision used. For example, in a case where there are 45 coding passes per code block, this means there are approximately 6528*45 = 293760 coding passes in a 2K digital cinema image. If all the coding passes are retained in the output code-stream, lossless or nearly lossless (depending on the DWT filters used) compression will result. In contrast to a lossless compressor, a typical lossy compressor will discard a large number of coding passes. It is the lossy compressor's rate control algorithm which specifically determines which of the many coding passes to include in the final output code-stream and which to discard.

A rate-distortion optimized compressor typically calculates an efficiency measure for each coding pass of each code-block. This efficiency measure is often called "distortion-rate slope".[4] Each coding pass has a certain size, ΔR, measured in bits or bytes. The inclusion of each coding pass reduces the resulting distortion by an amount ΔD. The quantities ΔR and ΔD are used to calculate the distortion-rate slope of the coding pass, $S = \Delta D/\Delta R$, which is essentially a measure of the efficiency of the bits in that particular coding pass in reducing distortion. The distortion-rate slope is calculated for each coding pass of each code-block. JPEG2000 places some restrictions on the order in which coding passes can be included, assuring, for example, that the less significant bits of a coefficient are not placed in the code-stream before the more significant bits,[5].[6]

Given this framework, the two traditional methods for rate control can be referred to as rate-based and efficiency-based. A rate-based rate control algorithm specifies a target size for the output code-stream, R_{target}. The coding passes with the steepest distortion-rate slopes are included before including other coding passes with lower distortion-rate slope. Coding passes are included in this manner until the target size, R_{target}, is met. This results in an output code-stream that meets specific rate goals.[5]

As noted in[5] the truncation point of the i^{th} code-block is denoted n_i. The rate and distortion associated with this truncation point in the i^{th} code-block are $R_i^{n_i}$ and $D_i^{n_i}$ respectively. The Post Compression Rate Distortion

Optimization (PCRD-opt) algorithm selects the optimal truncation point, n_i, for each code-block that minimizes the overall reconstructed image distortion D where

$$D = \sum_i D_i^{n_i}$$

subject to the constraint

$$R = \sum_i R_i^{n_i} \leq R_{target}$$

where R_{target} is the target bit rate. This optimization problem can be solved using the Lagrange multiplier technique where the optimization process is equivalent to minimizing the cost function

$$J = D + \lambda R = \sum_i \left(D_i^{n_i(\lambda)} + \lambda R_i^{n_i(\lambda)} \right)$$

with due attention to the rate constraint.

The other common JPEG2000 rate control method, efficiency-based rate control, involves specifying a distortion-rate slope threshold, $S_{threshold}$, and including all coding passes with a steeper slope than $S_{threshold}$ in the output code-stream. This means that all coding passes will have an efficiency equal to or greater than the specified distortion-rate slope threshold. While efficiency-based rate control is usually applied to still images, it has also been considered for image sequences in systems with varying amounts of buffer constraints.[7]

3. DISTORTION-BASED JPEG2000 RATE-CONTROL FOR CONSTANT QUALITY

Both rate-based and efficiency-based rate control achieve results that in many environments are quite satisfactory. However, the distortion-rate slope is a highly local measure that pertains to individual code-blocks. By contrast, what is of interest in many applications, including digital cinema, is the ability to obtain one or more images having a specific desired PSNR after encoding. In such constant distortion environments, the goal is to have the same residual overall distortion in the images obtained after considering data from all the code-blocks and taking the inverse wavelet transform. The residual distortion in a coded image is most directly related to the distortion reductions from the coding passes that were **not** included in the code-stream, not the distortion associated with the coding passes that were included. Thus, it is more intuitive, as the results below show, and more accurate, to utilize an approach that specifically accounts for distortion that will not be mitigated by the data in the coding passes that are utilized in constructing the encoded output.

The distortion-based optimization is similar to the rate-based optimization, except the roles of distortion and rate are exchanged. In rate-based optimization, distortion is minimized subject to the rate constraint, while in distortion-based optimization the opposite occurs. More formally, a distortion-based optimization selects the truncation points, n_i, for each code-block that minimize the overall rate

$$R = \sum_i R_i^{n_i}$$

subject to the constraint

$$D = \sum_i D_i^{n_i} \leq D_{target}$$

where D_{target} is the target reconstructed image distortion. This optimization problem can be solved using the Lagrange multiplier technique where the optimization process is equivalent to minimizing the cost function

$$J = \lambda D + R = \sum_i \left(\lambda D_i^{n_i(\lambda)} + R_i^{n_i(\lambda)} \right)$$

4. CONSTANT-QUALITY EXPERIMENTAL RESULTS

The desire for constant quality encoding is of particular interest in digital cinema applications. Even given all the customary caveats regarding the risks of using quantitative measures such as PSNR to gauge "quality", it is undeniable that PSNR and other similar measures, when used properly and with due consideration of other more subjective factors are valuable tools in assessing quality. From the standpoint of the user - in this case the person viewing a movie in a digital cinema - quality is clearly of far more direct relevance than is encoded data rate. From the standpoint of equipment and communications infrastructure providers, however, proper management of average and instantaneous compressed data rate are of paramount importance as well.

A practical system needs to respond to both of these factors. A system that involves only rate constraints can lead to significant and visible quality degradations for isolated images or sequences containing particularly difficult content. An implementation constrained only by quality may produce unacceptable local rate spikes. Thus, to be most effective an encoding system should consider both rate and quality. While the literature and the JPEG2000 standard itself gives ample treatment to rate-constrained encoding, there has been very little attention given to methods, such as that presented here, that allow targeting a specific quality, or equivalently, a constant D_{target} across multiple frames in an image sequence.

The constant quality algorithm was tested on a number of sequences of varying length and complexity, including a 600-frame sequence from the Disney film Treasure Planet. The Treasure Planet content is distributed with the Standard Test Evaluation Material (StEM) Mini-Movie, available through the Society of Motion Picture and Television Engineers (SMPTE).[9] Six representative frames from the sequence are shown in Figure 2. To allow fair comparisons between the three rate control methods, comparisons were conducted using the same average bit rate over the entire 600-frame sequence. The results shown here were obtained for a 1.0 bit per pixel (bpp) average rate using a 2048x1080 12-bit luma version of Treasure Planet.

The Treasure Planet content is well suited for illustrating the difference between the three types of rate control algorithms because the content varies significantly across the 600 frames as shown in Figure 2. The sequence begins with a black sky containing a moon crescent and a few stars (see Figures 2(a) and 2(b)). The camera then zooms into the moon which reveals a city with buildings (see Figures 2(c) and 2(d)). The sequence ends with a closer view of the city on the moon showing fine details of the buildings and flying ships (see Figures 2(e) and 2(f)).

Figure 1 shows the PSNR vs. Frame Number and Rate vs. Frame Number plots for the three different rate control methods. Note that the new constant quality algorithm (labeled "dist-based" in the figure) has the smallest variation in PSNR and the largest variation in rate of the three methods tested. The small residual variations in PSNR for the distortion-based curve is due in part to the non-orthogonality of the DWT and thus the fact that energy correlations between the DWT and image domains are approximate but not exact.

For the results presented in this paper, we used the standard distortion metric, mean square error (MSE), and plot our results in terms of PSNR, which for 12-bit image data is calculated as $\text{PSNR} = 10 \cdot log_{10}((2^{12}-1)^2/\text{MSE})$. It should be noted that other perceptually-based distortion metrics can be used with the distortion-based rate-control method, like those based on the contrast sensitivity function or a visual masking model.[8] Both the rate-based and the distortion-based rate control methods only require a single frame to be buffered at a time. Dagher, Bligin and Marcellin's slope-based method[7] requires a number of compressed frames to be buffered so the appropriate distortion-rate slope threshold can be determined.

The results in Figure 1 confirm that as expected, the rate-based method achieves constant rate, the distortion-based method achieves constant distortion (constant-quality), and the efficiency-based method leads to variations in both rate and quality. The efficiency-based method is simultaneously variable-rate and variable-quality. There has been some confusion in the JPEG2000 user community regarding efficiency-based coding, where it is sometimes assumed that the efficiency-based method produces constant quality. While it does lead in some environments to lower variations in quality (for example see Dagher et al.[7]), in general there can be significant quality variations as shown in Figure 1.

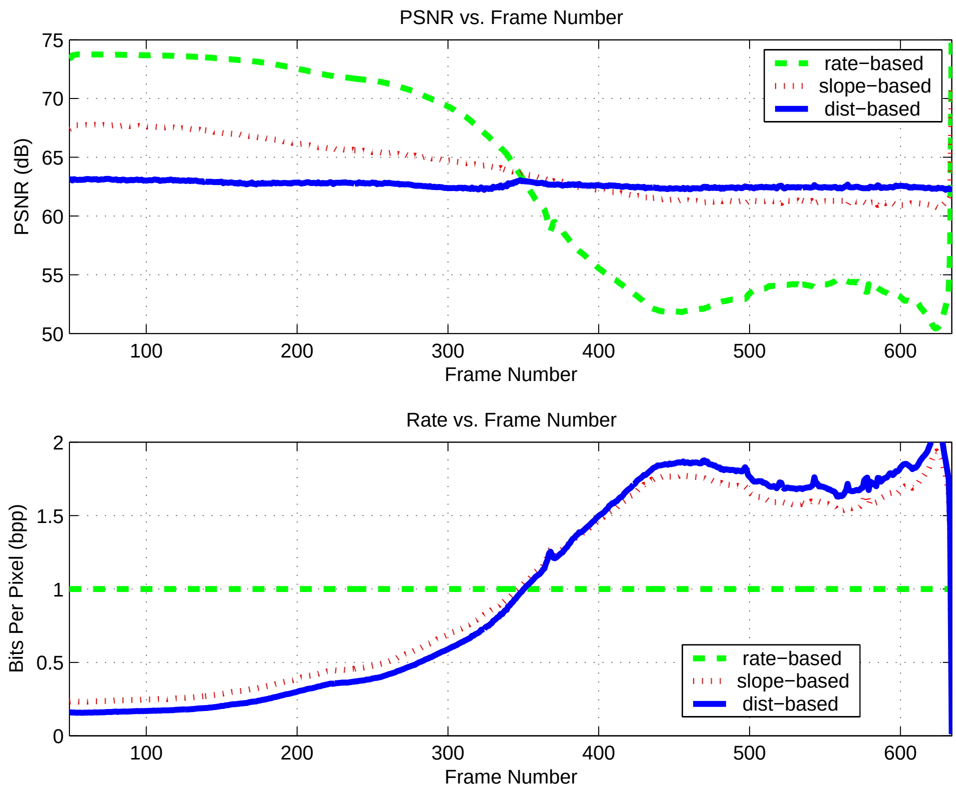


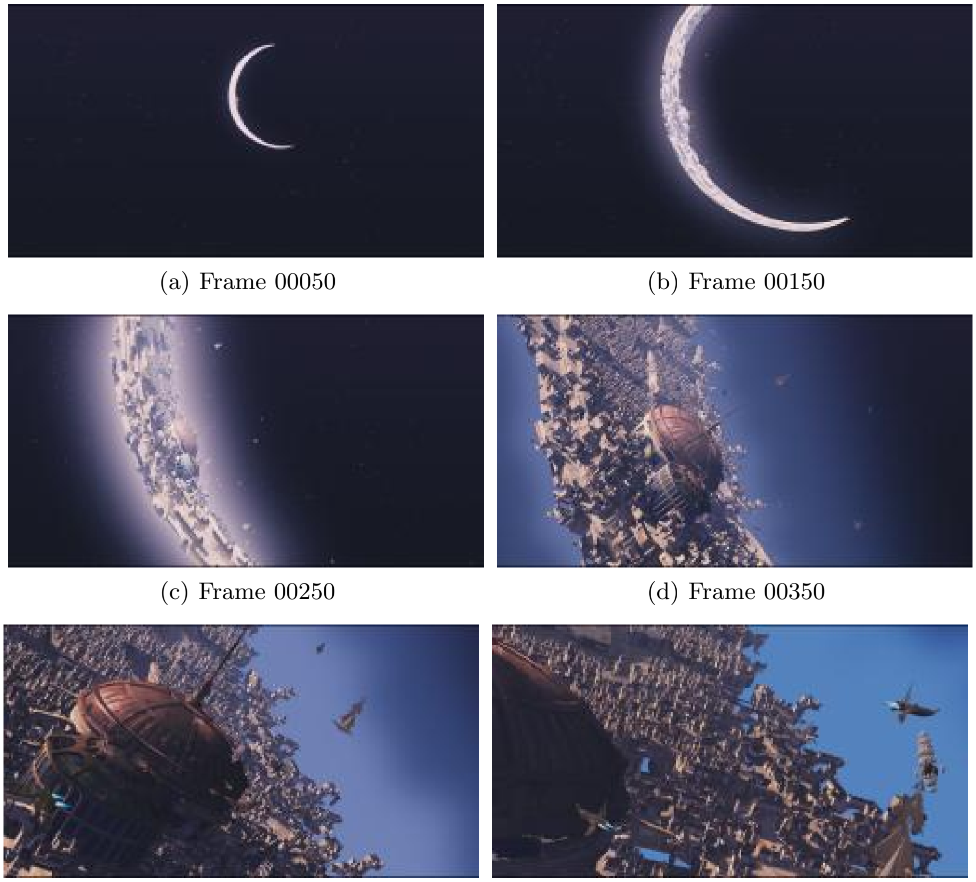
Figure 1. Rate Distortion performance of the three rate control methods. Note that the curve labeled "dist-based" achieves nearly constant PSNR.

5. JPEG2000 PROFILES FOR DIGITAL CINEMA

Two special digital cinema distribution profiles have been created by the JPEG committee in collaboration with SMPTE. Profile-3 is for 2K content and Profile-4 is for 4K content. The profiles have very specific constraints related to the organization and structure of the JPEG2000 code-stream. The main attributes of the Profiles for Digital Cinema are the following:

- code-blocks have size 32x32
- precincts are size 256x256 except those at the lowest resolution level which are 128x128
- the irreversible 9/7 wavelet filters are required
- A single tile is used for the whole image
- The progression order is CPRL
- The tile-part lengths, main header (TLM) marker must be included
- For 24 fps content, each code-stream may not exceed 1,302,2083 bytes, which corresponds to 250 Mbs.
- For 4K content, the 2K portion of the image must precede the 4K data in the code-stream.

Further details of the Digital Cinema profiles can be found in the official JPEG document.[10]



(a) Frame 00050 (b) Frame 00150

(c) Frame 00250 (d) Frame 00350

(e) Frame 00450 (f) Frame 00550

Figure 2. Frames from Treasure Planet sequence

6. CONCLUSION

We have described a JPEG2000 encoding method that enables a user to achieve specified distortion on an encoded image. When the same distortion constraint is applied to all the frames in an image sequence, the result is a sequence of images with nearly constant quality. The algorithm can be implemented on one frame at a time, so no multi-frame buffering is necessary. Experimental results confirm that the new method has much less PSNR variation than rate- and efficiency-based methods. Thus, it has strong potential for application in digital cinema where it can guarantee consistent image quality levels while also making efficient use of bits.

REFERENCES

1. ISO/IEC, *ISO/IEC 15444-1 Information Technology JPEG 2000 Image Coding System*, 2000.
2. M. D. Smith and J. D. Villasenor, "Intra-frame JPEG-2000 vs. Inter-frame Compression Comparison: The benefits and trade-offs for very high quality, high resolution sequences", in *Proc. SMPTE Technical Conf.*, Pasadena, CA, Oct. 2004, pp. 1-9.
3. M.D. Smith and J. Villasenor, "JPEG-2000 Rate Control for Digital Cinema," SMPTE Motion Imaging Journal, Vol. 115, No. 10, pp.394-399, Oct 2006.
4. J. Li and S. Lei, "An Embedded Still Image Coder with Rate-Distortion Optimization", *IEEE Trans. Image Processing*, vol. 8, no. 7, pp. 913-924, July 1999.
5. D. Taubman, "High Performance Scalable Image Compression with EBCOT", *IEEE Trans. Image Processing*, vol. 9, no. 7, pp. 1158-1170, July 1999.

6. D. S. Taubman and M. Marcellin *JPEG2000 - Image Compression Fundamentals, Standards and Practice*. Norwell, MA: Kluwer, 2002.
7. J. C. Dagher, A. Bilgin, and M. W. Marcellin, "Resource-Constrained Rate Control for Motion JPEG2000", *IEEE Trans. Image Processing*, vol. 12, no. 12, pp.1522-1529, Dec. 2003.
8. M. J. Nadenau, J. Reichel, and M. Kunt, "Wavelet-Based Color Image Compression: Exploiting the Contrast Sensitivity Function", *IEEE Trans. Image Processing*, vol. 12, no. 1, pp.58-70, Jan. 2003.
9. DCI and ASC, Standard Test Evaluation Material (StEM) Mini-Movie, *http://www.smpte.org/smpte_store/test_materials/*.
10. ISO/IEC 15444-1:2004/FDAM 1 - Information technology - JPEG2000 Image Coding System: Part 1 - Core Coding System, AMENDMENT 1: Profiles for Digital Cinema Applications.