While rate control in general has received significant attention in the academic and commercial communities, with a few notable exceptions there has been almost no formal research aimed at addressing the problem when a still image coding method such as JPEG-2000 is applied to successive frames in an image sequence. A new framework is introduced for rate control that enables a JPEG-2000 encoder to achieve a user-specified quality and therefore makes it possible to produce constant quality from frame to frame. The new method makes direct use of the same JPEG-2000 coding pass data as the traditional approaches and thus can easily be adopted at the back end of JPEG-2000 encoding engines. The proposed method is compared with two other common rate-control techniques for JPEG-2000.

Introduction

JPEG-2000 is the most advanced still-image compression standard and has the potential to affect 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. JPEG-2000 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 to derive improved frameworks for still image compression are quite limited, the issue of how JPEG-2000 can best be used for frame-by-
frame video compression remains open. At first glance, the application of JPEG-2000 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 inter-frame 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. 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 JPEG-2000 compression as the basis for digital cinema distribution. Substantial commercial efforts are already under way 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.

There is a long history of work on rate control for traditional video encoders; however, almost no attention has been paid to the issue of how to manage rate control on a video sequence in which each frame is compressed independently but where consistent 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 JPEG-2000, those efforts have by definition been aimed at coding of standalone images. Even in the standalone 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 JPEG-2000 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 ensure very high visual quality. However, approaches such as this tend to use far more bits than are necessary. It is therefore desirable to have a scheme that enables constant high quality and simultaneously makes economical use of bits subject to the quality constraint.

Rate Control for JPEG-2000

Currently available JPEG-2000 encoders usually implement either “rate-based” or “efficiency-based” rate-control algorithms. This section provides an overview of some key building blocks of a JPEG-2000 codestream and briefly reviews these common rate control methods.

The fundamental unit of data in the JPEG-2000 compression standard is the code-block. A code-block is simply a spatial grouping of wavelet coefficients, which have size 32 x 32 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 JPEG-2000 codestream. For a 4096 x 2160 3-color 12-bit digital cinema image, decomposed using a 5-level discrete wavelet transform (DWT), there are approximately (4096/32)*(2160/32)*3 = 128*68*3 = 26112 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 in which there are on average 45 coding passes per code block, this means there are approximately 26112*45 = 1175040 coding passes that result from the 4k digital cinema image. If all the coding passes are retained in the output codestream, lossless or nearly lossless compression will result (depending on the DWT filters used). 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 that specifically determines which of the many coding passes to include in the final output codestream 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 sometimes called “distortion-length slope.” Each coding pass has a certain size, ΔL, measured in bits or bytes. The inclusion of each coding pass reduces the resulting image distortion by an amount ΔD. The quantities ΔL and ΔD are used to calculate the distortion-length slope of the coding pass, S = ΔD/ΔL. The distortion-length slope is essentially a measure of the efficiency of the bits in that
particular coding pass in reducing distortion. The 
distortion-length slope is calculated for each coding pass of 
each code-block. JPEG-2000 places some restrictions 
on the order in which coding passes can be included, 
ensuring, for example, that the least significant bits of a 
wavelet coefficient are not placed in the codestream 
before the most significant bits.4,5

Given this framework, the two traditional methods for 
rate control are often referred to as efficiency-based and 
rate-based. A rate-based rate-control algorithm specifies 
a target size for the output codestream, L. The coding 
passes with the steepest distortion-length slopes are 
includes before including other coding passes with 
lower distortion-length slope. Coding passes are includ-
ed in this manner until the target size, L, is met. This 
results in an output codestream that meets specific 
length goals. A thorough explanation of this commonly 
used rate-based rate-control algorithm is available.4

An efficiency-based rate-control algorithm specifies a 
certain distortion-length slope threshold, Sthreshold, and all 
coding passes with a steeper slope than Sthreshold are 
included in the output codestream. The task of deter-
mining the appropriate Sthreshold was addressed for 
image sequences subject to buffer constraints.6 This 
approach ensures that all coding passes that have effi-
iciency greater than the threshold are included.

Constant Quality Rate-Control for JPEG-
2000

The traditional approaches have sound motivations 
and achieve results that in many environments are quite 
satisfactory. However, the distortion-length slope is a 
highly local measure that pertains to individual coding 
blocks. By contrast, what is of interest in many applica-
tions, including digital cinema, is the ability to obtain one 
or more images having a specific desired peak signal-
to-noise ratio (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 code-blocks that were not included 
in the codestream, not the distortion associated with the 
code-blocks that were included. Thus, it is more intu-
itive, as the results below show, and more accurate, to 
utilize an approach that specifically accounts for distor-
tion that will not be mitigated by the data in the coding 
passes that are used.

We propose a new constant-quality rate-control algo-
rithm, which delivers a specified target distortion for the 
output codestream, DTarget. The coding passes with the 
steepest distortion-length slopes are included before the 
coding passes with the lower slopes. In contrast with the 
earlier approaches, the cutoff is based on a global mea-
sure of total distortion, DTarget, as opposed to local mea-
sures based on the distortion-length slopes of individual 
code blocks. The total amount of distortion reduction 
possible for a code-block for which there are a total of N 
coding passes available is the summation of all N distor-
tion reductions corresponding to each coding pass.

![Image](https://via.placeholder.com/150)

If M coding passes from a given code-block are 
included in the output codestream, then the remaining 
distortion in the code-block, DCBRemain, is calculated as:

\[ D_{CBRemain} = D_{CBTotal} - \sum_{i=0}^{M-1} \Delta D_i \]

The total remaining distortion in the image is the sum-
mation of the remaining distortion of each code-block; in 
other words, it represents a measure of the distortion 
that can be expected in the image due to the coding 
passes not included in the encoder output. If there are B 
code-blocks in the image (B is approximately 26,112 for 
the example 4k image with a 5-level DWT considered 
earlier), then the total remaining distortion, DTotal, Can be 
expressed as follows:

\[ D_{Total} = \sum_{b=0}^{B-1} D_{CBRemain}(b) \]

where D_{CBRemain}(b) represents the remaining distor-
tion in code-block b. Coding passes are added until the 
total remaining distortion, D_{Total}, equals the target dis-
ortion, D_{Target}. If the same target distortion, D_{Target}, is 
applied to all the images in an image sequence, the 
result is a constant-quality per frame across the whole 
image sequence. A flow chart illustrating the imple-
mentation of the constant-quality approach is given in 
Fig. 1.
Implementation and Results

Two 4k clips were used to demonstrate the proposed quality-based rate-control method. The first clip, shown in Fig. 2, contains 366 frames from the DCI StEM content; this sequence was referred to as “Clip 2” during the DCI compression tests. The second clip, shown in Fig. 3, contains 586 frames from Disney’s “Treasure Planet” content; this sequence was referred to as “Clip 6” during the DCI compression tests. Both clips are 4:4:4 12-bit 4k content, the DCI StEM content has dimensions 4096 x 1714 and the Treasure Planet content has dimensions 4096 x 2160. The PSNR metric is used for the quality comparisons in units of decibels (dB). For 12-bit content, PSNR is calculated as PSNR=10*log_{10}(4095*4095/MSE), where MSE is the mean square error between the original and decompressed image. Megabits per second (Mbs) is used for the rate comparisons. Rate results are often presented in units of bits per pixel (bpp) in the image compression literature and kilobits per second (Kbs) or megabits per second (Mbs) in the video compression literature. In these experiments, 100 Mbits/sec corresponds to approximately 0.594 bpp for the DCI StEM content and 0.471 bpp for the Treasure Planet content.

The compression experiments were performed using the luma (Y’) color channel. To make fair comparisons between the three rate-control methods, the average bit rate was kept at 100 Mbits/sec for each sequence. The compression software used for these tests is C++ based. The quality-based method has also been implemented in Java software, and the method is currently being ported into a hardware implementation.

The compression results for the three rate control methods are shown in Figs. 4 and 5. Note that the proposed quality-based approach has the smallest variation in PSNR of the three methods. The PSNR results are also described statistically in Tables 1 and 2.

The small residual variations in PSNR for the “quality-based” curves shown in Figs. 4 and 5 are due in part to the non-orthogonality of the discrete wavelet transform (DWT) and thus the fact that energy correlations between the DWT and image domains are approximate but not exact. The experiments that were performed minimize the mean square error between the original and decompressed image. It is well known that mean square error is not the best perceptual quality metric, but it is used here for simplicity and comparison purposes. It is reasonably straightforward to adapt the JPEG-2000 rate-control framework.
to use other perceptually-based quality metrics such as those based on the contrast sensitivity function (CSF) or visual masking.\textsuperscript{7}

Both the rate-based and the proposed quality-based rate-control methods require only a single frame to be buffered at a time. The efficiency-based method introduced in\textsuperscript{6} requires a number of frames to be buffered so the appropriate distortion-length slope can be determined. The number of frames in the rate-control buffer has a direct effect on memory usage as well as the degree of parallelism that can be exploited by the encoder. The rate-based and proposed quality-based methods achieve frame-level parallelism, (meaning each frame can be independently encoded). The efficiency-based method\textsuperscript{6} requires access to the rate-distortion statistics of all the frames in the sequence. From an implementation point of view, the rate-based and proposed quality-based methods are much easier to parallelize than the efficiency-based approach.

**JPEG-2000 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 JPEG-2000 code-stream. The main attributes of the Profiles for Digital Cinema are as follows:

- Code-blocks have size 32 x 32.
- Precincts are size 256 x 256, except those at the lowest resolution level, which are 128 x 128.
- 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 frame/sec content, each code stream may not exceed 1,302,2083 bytes, which corresponds to 250 Mbits/sec.
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 are available.8

Conclusion

An encoding method has been described that enables JPEG-2000 encoding to achieve a user-specified quality 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 multiframe buffering is necessary. Experimental results confirm that the new method has much less PSNR variation than earlier rate- and efficiency-based methods when applied to successive frames in an image sequence. 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


Presented at the SMPTE and VidTrans Joint Conference, Hollywood, CA, Jan. 29-Feb. 1, 2006, Copyright © 2006 by SMPTE.

THE AUTHORS

Michael Smith is a consultant in the area of digital imaging and signal processing, with recent work for organizations including Digital Cinema Initiatives (DCI), Warner Brothers Technical Operations, Dolby Laboratories, Cinea Inc., Path1 Networks, PhatNoise Inc., and various law firms. He is a member of SMPTE and AES. Smith received BS and MS degrees in electrical engineering from UCLA in 2001 and 2004, respectively (miksmith@attglobal.net).

John D. Villasenor received a BS degree in 1985 from the University of Virginia, a MS degree in 1986 from Stanford University, and a PhD degree in 1989 from Stanford, all in electrical engineering. From 1990 to 1992, he was with the Radar Science and Engineering section of the Jet Propulsion Laboratory in Pasadena, CA, where he developed methods for imaging the earth from space.

Villasenor joined the Electrical Engineering Dept. at the University of California, Los Angeles (UCLA), in 1992 and is currently a professor. He served as vice-chair of the department from 1996 to 2002. At UCLA, his research efforts lie in communications, computing, imaging and video compression, and networking. Villasenor is a senior member of the IEEE (villa@icsl.ucla.edu).