

**Intra-frame JPEG-2000 vs. Inter-frame Compression Comparison:
The benefits and trade-offs for very high quality, high resolution sequences**

Michael Smith and John Villasenor

For the past several decades, most of the research on image and video compression has focused on addressing highly bandwidth-constrained environments. However, some of the most compelling applications of image and video compression involve extremely high-quality, high-resolution image sequences where the primary constraints are related to quality and flexibility.

The interest in flexibility, in particular the potential to generate lower resolution extractions from a single higher-resolution master was a primary driver behind the recent evaluation and consideration of JPEG2000 compression for digital cinema sequences. JPEG2000 is an intra-frame coding method where a single frame is coded at a time and no frame-to-frame motion compensation is applied.

While the efficiency of exploiting inter-frame motion compensation is well established for lower bit rates, far less attention has been given to this issue for digital cinema resolutions. This paper addresses the specific coding efficiency differences between inter and intra frame coding for very high quality, high-resolution image sequences and at a wide variety of bit rate ranges. The information presented here will be useful to system designers aiming to balance the flexibility and benefits of intra-frame approaches against the bandwidth advantages of inter-frame methods that exploit frame-to-frame correlation.

Michael Smith (miksmith@icsl.ucla.edu) is a consultant in the area of digital imaging. He received his B.S. and M.S. degrees in Electrical Engineering from UCLA. John Villasenor (villa@icsl.ucla.edu) is on the faculty of the UCLA Electrical Engineering Department, and conducts research in signal processing and communications.

1. Introduction

All of the commonly used video coding standards including MPEG1, 2, 4, H.261, H.263, and the new H.264/AVC [1] (also known as MPEG4 Part 10) utilize motion compensation to exploit the redundancy between nearby video frames. While these standards differ slightly in the way motion compensation is performed, they share a common approach of performing interframe coding based on motion vectors that predict frame content using blocks drawn from one or more nearby frames, obtaining a difference between the predicted and actual pixel values, and utilizing the Discrete Cosine Transform (DCT), or DCT-like transforms, to represent some of this prediction error. Interframe coding therefore uses both the spatial redundancy present in neighboring pixels within each frame and the temporal redundancy relating pixels in nearby frames. This stands in contrast to intraframe methods such as JPEG and JPEG2000 [2] that code a single image on a standalone basis and rely on spatial redundancy alone to achieve compression.

Interframe coding is commonly believed to be dramatically more efficient than interframe coding in terms of the bits/pixel needed to deliver a given post-compression quality. As a result, the enormously higher computational cost of performing motion compensation is viewed as justified, particularly since many applications are highly asymmetric in that decoding complexity (which is much more modestly increased in interframe coding) is far more important than encoding complexity. While there are instances when intraframe techniques are applied to video sequences, such as when motion-JPEG is used, this is typically viewed as representing a major concession in compression efficiency in order to achieve other advantages such as easier random access and lower complexity systems for encoding and decoding.

The compression efficiency superiority of interframe coding is indeed very significant at the low to moderate bit rates that have characterized almost all applications of image sequence coding. For example, video over the internet, which involves data rates typically ranging from the high tens of kbits/sec up to around 1 Mbit/sec, would simply not be practical without interframe coding. DVD and direct broadcast satellite, both of which deliver video at several Mbits/sec, are also dependent on interframe methods. Interframe coding holds similar (though less significant) compression efficiency advantages for handling of HDTV signals.

Digital cinema, however, is different in several fundamental ways from these other applications. First, the resolution is higher, with a single frame of 4K data containing approximately 8 million pixels. Secondly, the quality requirements are more stringent. Because cinema viewers tend to sit within 1.5 to 3 screen heights of the projected image as opposed to the 3 or more screen heights that characterize home television viewing environments, there is greater artifact sensitivity in a cinema environment. To achieve "visually lossless" coding, post-compression data rates of 100-300 Mbits/sec (for 4K) and 30-75 Mbits/sec (for 2K) are needed.

At these data rates and quality levels, the assumption that motion compensation is more efficient needs to be reexamined. That assumption is rooted in the understanding that coding of the motion vectors and the prediction error obtained after motion compensation is usually superior to direct coding of the pixels concerned. However, prediction errors are quite difficult to code. By definition, the prediction error is derived by

subtracting two images (or parts of images), and is therefore often noisy and lacking in the very correlation and structure that make transform coding attractive in the first place. Put simply, transforms are quite efficient for coding well correlated data, but are not particularly well suited to noisy, uncorrelated data.

In what follows we explore the relative compression efficiency advantages for coding of digital cinema image sequences using 1) interframe coding using motion-compensation + DCT and 2) intraframe coding. While recognizing that there are many interframe and intraframe algorithms to choose from, the experiments we perform are based on H.264 and JPEG2000, as these are the latest-generation versions of each of these approaches.

2. Experiment description

2.1 Test content

The primary test content used was from the DCI 4K Minimovie. Two clips were used. The first, Clip 1 “Confetti Title Screen” is shown in Figure 1. This sequence has a resolution of 4096x1714, with 12 bits per color and 4:4:4 color sampling. To reduce the amount of processing during the experiments, a cropped region of size 512x208 was used as indicated by the white box in Figure 1 (the upper left corner corresponds to pixel coordinates (700,1200) on the original frame). This region contains approximately 1/64 of the total number of pixels. Provided that the content within this region is similar in compression difficulty to the full uncropped frame, compressed data rates obtained in the cropped region can be scaled to obtain corresponding rates for the entire frame. The results obtained using the cropped image data are referred to in the remainder of this paper as the “4K resolution” results, since no downsampling, averaging, or other modification of the original DCI pixel data values was performed prior to compression.

We also obtained a “512 resolution” image sequence with pixel dimensions 512x208 by downsampling the entire frame by a factor of 8 in each direction. The rationale behind creating image sequences using both unmodified 4K-sampled pixel values and low-resolution data is to compare the impact of sample density in a scene on compression. This allows direct evaluation of the relative performance of interframe and intraframe coding at low and very high resolutions.

The second clip was Clip 6 used in the DCI compression tests, which is a sequence from Treasure Planet. As with Clip 1, both a representative cropped subimage (512x272) and a downsampled image (512x272) sequence were used to create 4K resolution and 512 resolution inputs to the compression experiments. The upper left corner of the cropped subimage corresponds to pixel coordinates (1400,800) in the original frame.



Figure 1. DCI Clip 1 - "Confetti Title Screen"

2.2 Color Transform

We used the ICT (Irreversible Color Transform), which is the color transform used in the lossy mode of JPEG2000. This contrasts with the Reversible Color Transform (RCT), which is used in the lossless mode. The main difference between the two is that the ICT uses floating point values, while the RCT uses integer arithmetic only, which is necessary if exact lossless reconstruction is required.

The Irreversible Color Transform (ICT) is described by the following equations:

$$\mathbf{a}_R = 0.299 \quad \mathbf{a}_G = 0.587 \quad \mathbf{a}_B = 0.114$$

$$x_Y = \mathbf{a}_R x_R + \mathbf{a}_G x_G + \mathbf{a}_B x_B$$

$$x_{Cb} = \frac{0.5}{1 - \mathbf{a}_B} (x_B - x_Y)$$

$$x_{Cr} = \frac{0.5}{1 - \mathbf{a}_R} (x_R - x_Y)$$

The Irreversible Color Transform is very similar to the color transform described in BT.601 [3], which is often used for digital video. In the H.264/AVC experiments, we offset the Cb and Cr components by $2^{(B-1)}$ where B is the bit depth of each color plane (12 for the work presented in this paper).

2.3 Chroma subsampling

Often in video compression, the chroma components (Cb and Cr) are subsampled by a factor of two because the human visual system is less sensitive to high frequency chroma components. When chroma samples are subsampled in the horizontal dimension, the color sampling is commonly referred to as "4:2:2". When both horizontal and vertical dimensions of the chroma samples are subsampled, the color sampling is

DCI Clip 1 - Confetti Title Screen - 4K

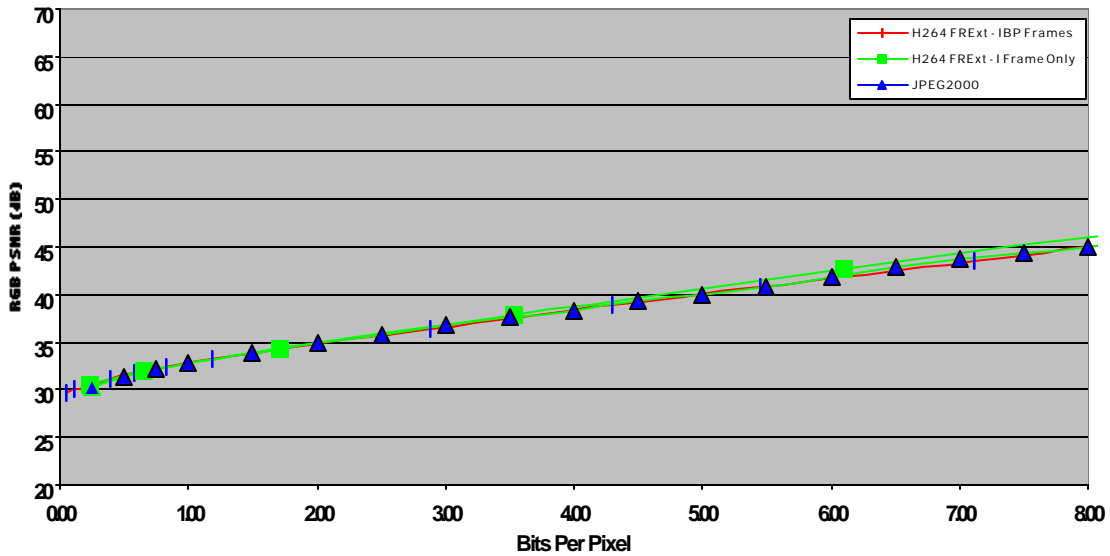


Figure 3a

DCI Clip 6 - Treasure Planet - 4K

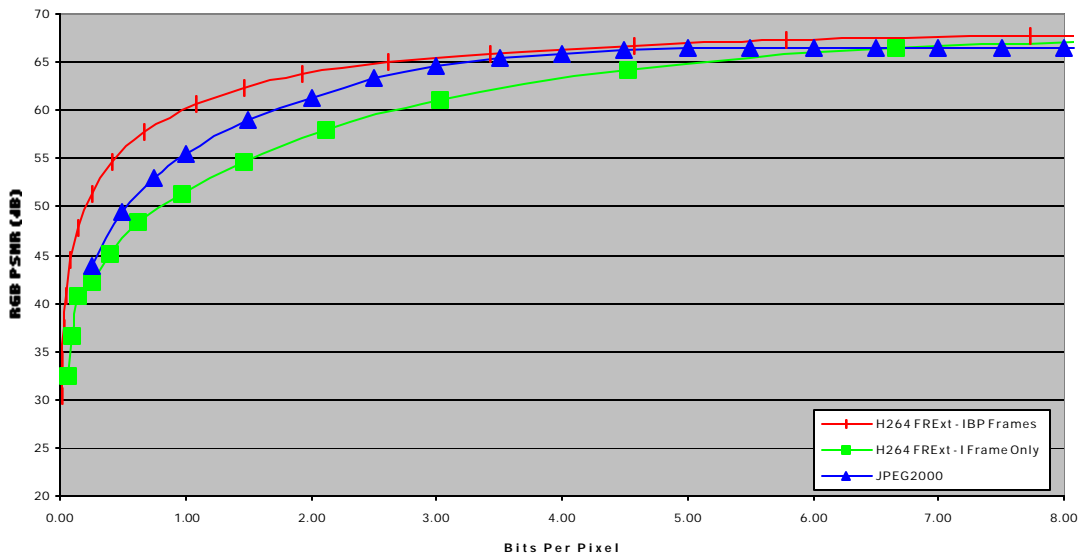


Figure 3b

The most notable feature in Figure 3a is the relatively small difference in PSNR performance between all three algorithms. This nearly identical performance holds over the entire bit rate range that was tested. The results are quite different for the computer generated imagery from Treasure Planet, shown in Figure 3b, which indicate that for this

clip interframe coding outperforms intraframe coding by 5-10 dB over much of the bitrate range. Figure 3b also shows that of the two intraframe methods, JPEG2000 significantly outperforms H.264 with I frames only for low bit rates, which is exactly what would be expected given the general superiority of JPEG2000 over DCT-based methods for lower rate coding.

The contrast between Figures 3a and b raises the obvious question of why Clips 1 and 6 should lead to such different results. Film grain is almost certainly a factor, as it is present in Clip 1 and absent in Clip 6. Since film grain is random from frame to frame, it obviously can not be coded using motion compensation. At visual quality levels where the compression algorithm is called upon to replicate or approximate film grain, it follows that interframe coding will not be particularly advantageous. This is also a possible explanation for the slightly superior performance of intraframe coding (H.264 with I frames only) over interframe coding for very high bit rates for Clip 1. At extremely high quality levels, predictive coding of film grain can be detrimental from a compression standpoint, since the motion compensation provides no coding value for grain but still costs bits (and computer cycles) to perform.

DCI Clip 1 - Confetti Title Screen - 512

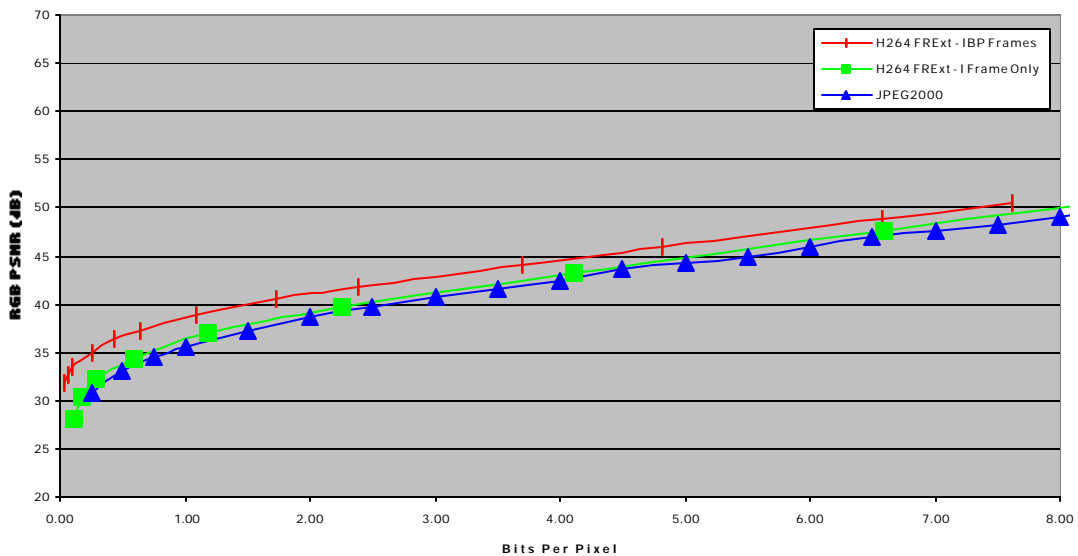


Figure 4a

DCI Clip 6 - Treasure Planet - 512

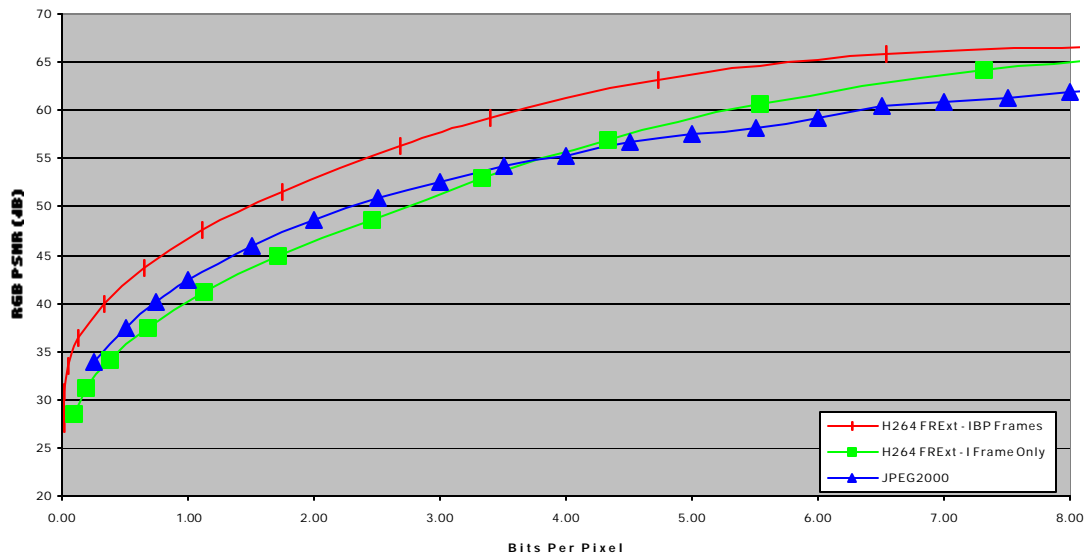


Figure 4b

Figures 4a and b contain results for the low-resolution sequences (512x208 for Clip 1 and 512x272 for Clip 6). As Figure 4a shows for Clip 1, in this case interframe coding is roughly 2 dB better than intraframe coding for most of the bitrate range, though the margin increases dramatically at the very low bit rates. This is consistent with what would be expected, as the lower sampling density of lower resolution images creates higher spatial frequencies that, in contrast with film grain, do maintain frame-to-frame correlation and thus are suited to interframe coding. While film grain is also present in the lower resolution images, the percentage of the energy in the predicted image that is uncorrelated is much lower than in the high resolution images.

Figure 4b shows the low-resolution results for Clip 6. As with the Clip 1 results, interframe coding outperforms intraframe coding by a significant margin over the entire range of bit rates. One difference between the Clip 1 and 6 results is that for Clip 6 JPEG2000 is markedly better than intraframe H.264 until about 4 bits/pixel, while for Clip 1 it is intraframe H.264 that gives slightly (but perhaps not statistically significant) better performance.

3. Conclusions

We have presented experimental results indicating that the coding efficiency advantages of interframe coding are significantly reduced for 4K digitized film content at the data rates, and quality levels associated with digital cinema. For CGI content, interframe coding appears to maintain a significant PSNR advantage over intraframe coding, though if the overall PSNR is high enough for both inter and intraframe coding for CGI, this difference may not be visible. For the digitized film content, at least, this indicates that there may be no particular benefit to interframe coding, as it is much more computationally complex, creates data access complications due to the dependencies among frames, and does not evidence superior compression performance. For the lower

resolution (512x208 and 512x272) experiments, interframe coding was more efficient than intraframe coding, as expected. While the experiments here were performed using H.264 and JPEG2000, replacing either or both of these algorithms with other advanced inter or intraframe methods would be unlikely to change the basic nature of the results.

We emphasize that the results presented here are based on testing of only one clip of each type (digitized film and CGI), and that further work should be performed to establish the variability of the results across different types of digitized film, different resolutions (4K and 2K), different forms of CGI. Assuming that these results can be generalized, they provide strong justification for utilizing JPEG2000 or other intraframe coding methods for digital cinema content derived from digitized film. For CGI, while we do find that intraframe coding gives some PSNR loss with respect to interframe coding, if the overall quality is sufficiently high this difference would not be visually problematic. Under those circumstances, then JPEG2000 would also be an excellent choice for digital cinema.

4. Acknowledgements

The authors thank Eric Gsell and Walter Gish of Dolby Laboratories and Tao Tian of Qualcomm for useful discussions. This work was supported in part by the U.S. Office of Naval Research.

5. References

- [1] Advanced Video Coding for Generic Audiovisual Services, ITU-T Recommendation H.264, 2004.
- [2] Information Technology – JPEG2000 – Image Coding System – Part 1: Core Coding System, ISO/IEC 15 444-1, 2000.
- [3] ITU-R BT.601-5 (10/95), Studio Encoding Parameters of Digital Television for Standard 4:3 and Wide-Screen 16:9 Aspect Ratios.
- [4] David Taubman, Software Kakadu V4.2.1, www.kakadusoftware.com