RATE CONTROL FOR FLICKER ARTIFACT SUPPRESSION IN MOTION JPEG2000

Athanasios Leontaris, Yoshihide Tonomura*†*, and Takayuki Nakachi*†*

*Information Coding Laboratory, University of California, San Diego, La Jolla, CA 92093, U.S.A.
†NTT Network Innovation Laboratories, Yokosuka, Kanagawa 239-0847, Japan

ABSTRACT

Video sequences encoded with JPEG2000 exhibit flicker artifact. This artifact is not perceivable in still images. In contrast, it is perceived in the temporal domain and is the result of two factors: (a) the image content, and (b) the EBCOT post-compression rate allocation. In this work, we address the second factor. First, we propose an intra-frame rate-allocation scheme that ensures uniform error energy distribution and improves perceptual quality. We then extend this scheme to the temporal domain and further modify it to address flicker. Experimental results show that our algorithm suppresses the flicker artifact without sacrificing good PSNR performance.

1. INTRODUCTION

Intra-frame coders compress each frame of a video sequence as a separate still image. Unlike predictive encoders (MPEG), they do not exploit temporal correlation and they sacrifice compression efficiency for random access and error resilience. In addition, their encoding and decoding complexity is lower compared to predictive video codecs. The JPEG2000 image coding standard is such a coder, that features SNR, spatial, and resolution scalability. An emerging application of JPEG2000 is the compression of digital cinema content which is characterized by very high spatial resolution (e.g. 4096×2160 pixels). The combination of wavelets and intra-frame coding lead to flicker.

Flicker artifact can be categorized into two types: (a) flicker artifact due to small temporal variations in the input signal luminance, that give rise to large differences in the de-quantized coefficients [1], and (b) flicker artifact due to uneven rate allocation among collocated code-blocks across subsequent frames. A code-block is the smallest coding unit in JPEG2000, and usually consists of 32×32 wavelet coefficients. Even though flicker is more visible at low bit rates, it can also be observed at medium bit rates. Furthermore, it is perceived solely in the temporal domain. Type (a) flicker artifact is perceived as small ripples in decoded frames. The main cause is the coarse quantization of high frequency subbands which in wavelet coding results to ringing artifact around objects. Due to object motion the phase of the video signal changes across subsequent frames, and a temporal flicker noise signal is produced. Type (a) flicker artifact is a result of *motion*, the wavelets, and the image content.

In contrast, type (b) flicker artifact is perceivable in areas that appear static across subsequent frames; in particular when intra-frame coders (JPEG2000) are used. In [2], the relationship of the flicker artifact with the post-compression (PC) quantization of Motion JPEG2000 was studied. It was found that the main cause of type (b) flicker artifact is the temporal variation in PC quantization. The EBCOT rate allocation scheme of JPEG2000 performs PC quantization through code-block bit-plane truncation on an *intra-frame* basis. As a result, collocated JPEG2000 code-blocks, corresponding to identical content in subsequent frames, are often truncated at different lengths. It was observed in [2] that the temporal variation of the distortion of static regions increases considerably when post-compression EBCOT allocation is used. In [3] it was noted that even if most of the frame is static, a slight content change (e.g. object motion) can significantly alter the rate allocation among static blocks of two subsequent frames. In summary, type (b) flicker artifact is perceivable in *static* regions and is a result of PC rate allocation.

Kuge [1] and Kato et al. [2], studied the type (a) and (b) artifacts, respectively, but did not address the artifact. More recently, Becker et al. [3] investigated both artifacts and discussed an approach for type (a) artifact suppression. In this work, we address the type (b) flicker artifact. We propose a JPEG2000 standard-compliant solution, that modifies the post-compression rate control scheme.

The paper is outlined as follows: We first address the problem of intra-frame non-uniform error distribution and point to perceptual problems of EBCOT for certain types of content in Section 2. This intra-frame scheme is then extended to inter-frame to address the type (b) flicker artifact in Section 3. The method is further refined in Section 4 to help improve the image quality in areas not affected by this type of artifact. Experimental results are presented and discussed in Section 5. Finally, the paper is concluded in Section 6.

2. NON-UNIFORM BIT-PLANE TRUNCATION

EBCOT rate allocation can assign different code-block (coding-pass) truncation lengths for similar or repetitive content within *still* images. While EBCOT is efficient in a rate-distortion sense, it inadvertently causes non-uniform distribution of the quantization error within the frame. A characteristic example is shown in Fig. 1(a). The circular zone pattern has been encoded with EBCOT. The top half exhibits block-like artifacts. In addition, while the bottom half is a mirror image of itself, identical spatial content is not encoded in the same way. Different artifacts are visible on the rightmost part. Despite good PSNR results, perceptual quality suffers.

In previous work [4], it was proposed to use the same coding pass truncation length throughout the image. Uniform truncation resulted to a uniformly distributed quantization error over all code blocks. This intuitive rate allocation improved perceptual quality considerably. However, experimenting with uniform truncation of all code-blocks, we verified that PSNR performance suffers on average by 0.7dB compared to EBCOT. The reason is the failure to account for the influence of the decomposition *level*, the colour *component*, and the *wavelet kernel* on the energy gain of each subband.

Let $e_{s,r,c,w}(l)$ denote the *effective* quantization error which is a result of a bit at bitplane l not being transmitted in a code-block in subband s, resolution level r, and color (RGB or YCrCb) com-

This work was performed while the first author was a summer intern at NTT Network Innovation Laboratories. We would like to thank Prof. Pamela Cosman of U.C. San Diego for her helpful feedback on the manuscript.

ponent c, for wavelet kernel w. Energy term $e_{s,r,c,w}(l)$ is calculated as: $e_{s,r,c,w}(l) = G_{s,c,w} \times W_{r,c}^2 \times (\delta(l) \times 2^{K_{max}-1})^2$ where $\delta(l)$ is the quantization step size ($\delta(l+1) = \delta(l) \times 2$), K_{max} is the maximum bitplane in the image, $G_{s,c,w}$ is the subband energy gain factor, calculated from the kernel and the color transform gains, and $W_{r,c}$ is the component and level weight. All the above energy weights are defined in the Kakadu software. Truncating a code-block of subband HL at length l results to error *energy* $e_{HL,r,c,w}(l)$, while truncating a code-block of subband HH at length l results to error energy $e_{HH,r,c,w}(l)$, which are *not* equal. Thus, the technique in [4] results in non-uniform effective error distribution. We propose to keep the truncation length uniform within the subband's code-blocks (for given (s, r, c, w)), while varying it across subbands to maintain uniform effective error distribution. The ranges for the three parameters are: $s = \{LL, HL, LH, HH\}$, $c = \{Y, Cr, Cb\}$, and $r = \{1, 2, 3, 4, 5\}$. Throughout this study we used 5 wavelet decomposition levels with the 9/7 wavelet kernel.

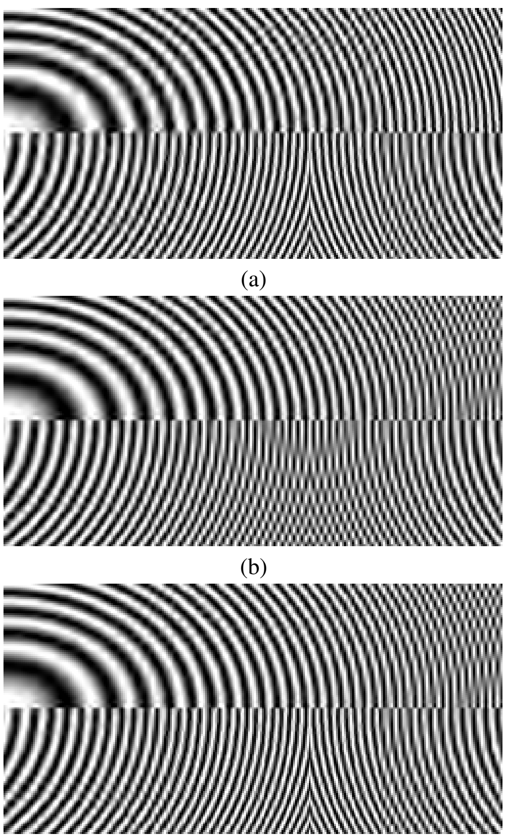


Fig. 1. Decoded frame 51 at 0.40bpp for the circular zone pattern sequence. (a) Standard EBCOT JPEG2000 Rate Allocation, (b) TRUNC inter-frame allocation, (c) HYBRID rate control.

Given error energies $e_{s,r,c,w}(l)$ we seek the subband correspondences. The problem is formulated as: if subband β_i is truncated at bitplane b_i, find the truncation bit-planes b_j, where $j \neq i$, for subbands β_j so that $e_{\beta_j}(b_j) \simeq e_{\beta_i}(b_i)$. We seek a table that for a given truncation bit-plane of an *arbitrary* subband yields the corresponding truncation bit-planes for all other subbands. The relations between different subband energies, within a bitplane, do not depend on the bitplane ($\delta(l)$). Furthermore, incrementing or decrementing l raises or decreases the energies by a factor of 4 regardless of l. As a result, the number of bit-planes that have to be truncated at each subband to ensure similar energies is independent of l, and the table will contain only bitplane up and down shift factors for all subbands. We propose the following approach to populate the table:

Step 1. All code blocks in component 0, resolution 1, and subband LL are truncated up to and including some arbitrary bitplane b. These parameters correspond to a certain error energy level $e_{LL,1,0}(b)$.

Step 2. Given all possible (r, s, c, w) parameters, we seek $\hat{b}$ so that $\mid e_{r,s,c,9/7}(\hat{b}) - e_{LL,1,0,9/7}(b) \mid$ is minimized. The truncation lengths for all combinations of (r, s, c, w) that correspond to b are thus obtained.

Truncation takes place from bottom to top; e.g. a table entry of plane 2 translates to truncating bitplanes 0 (least significant), 1, and 2. The final shift factor table is obtained by subtracting b from all its values. The above calculations are done off-line, stored and used during encoding. The truncation tables obtained depend only on the number of resolution levels, the wavelet kernel and CSF functions. They do *not* depend on the image content.

Each bit-plane in JPEG2000 is coded in three coding-passes: the significance, magnitude refinement, and cleanup pass. We truncate on a pass basis. The bit-plane correspondences are obtained from the table, while the pass remains fixed. If the LL subband is truncated at the cleanup pass of bit-plane 3, and the table points to bit-plane 4 for the HH subband, then HH will be truncated at the cleanup pass of bit-plane 4. During encoding, the pass truncation length l is incremented iteratively, starting from 0. At each iteration, the $(LL, 1, 0)$ subband code-blocks are truncated with the current length l while the other subbands with the lengths obtained from the correspondence table. The search for l is terminated at the first one satisfying the rate constraint. A drawback of this scheme is coarse rate control. Even though high frequencies are less detailed, this coding scheme exhibits better perceptual quality than EBCOT as shown in Fig. 1(b).

3. INTER FRAME EXTENSION

The scheme in Section 2 is extended to the temporal domain to reduce (and in some cases completely eliminate) the flicker artifact. We observed that subsequent frames, with similar content and entropy, often happened to use the same coding pass truncation length. This reduces flicker artifact as observed in the circular zone pattern (CZP) sequence in Fig. 1. As a result, we encoded all frames of the CZP sequence employing the same truncation length and observed that the resulting reconstructed sequence was free of flicker.

Hence we were motivated to maintain the same truncation length for as many *consecutive* frames as possible. However, as the content changes in a sequence, there are going to be variations in the rate expended by subsequent frames. The bit rate may vary significantly so there should be a constraint on the rate variation range. The goal is to keep the truncation length fixed for as long as possible, given the constraint that the rate variation is below some threshold. The previous frame's truncation length is retained if both of the two following conditions are satisfied:

Condition 1. The absolute difference between the resulting rate r_n^p using the previous truncation length p and the target rate t_n for the frame n is sufficiently small: $\mid r_n^p - t_n \mid < \Delta t_n$. The motivation is to avoid unnecessary changes provided the resulting rate is not significantly larger or smaller than the target rate.

Condition 2. The absolute difference between r_n^p and the rate for the previous frame r^p_{n-1} is sufficiently small: $| r^p_n - r^p_{n-1} | < 2\Delta t_n$. Intuitively, this condition ensures low variance of the per-frame bit rate. The "2" in the right side of the expression allows for slightly increased rate variation between consecutive frames.

The parameter Δ could be optimized individually for each sequence, however in our scheme we have set it to 0.1. Since this is an inter-frame scheme, the parameters that are buffered are the previous frame truncation length and resulting bit rate. Once the truncation length for the current frame is determined, the frame is encoded with non-uniform truncation as in Section 2. This inter-frame JPEG2000 rate allocation scheme will be referred to hereon as TRUNC.

4. A HYBRID APPROACH IN THE TEMPORAL DOMAIN

Subjective observations show that flicker artifact is most prominent in *static* image regions. To efficiently address flicker and maintain good SNR performance one can employ a two-tier hybrid encoding strategy where both TRUNC from Section 3 as well as EBCOT are jointly used. Once the flicker-prone areas have been identified, the proposed rate allocation scheme can ensure equal error energy among static subsequent blocks in the temporal domain, by using TRUNC to encode them. *Active* code-blocks that do not contribute to flicker artifact (masked by motion) will be encoded with EBCOT, as it exhibits slightly better PSNR performance than TRUNC, and handles high-frequency areas more gracefully. Due to EBCOT finer rate control is also possible.

A block in the original frame is denoted static if the difference from its collocated counterpart in the previous frame is sufficiently small. Otherwise it is denoted active. The classification map is obtained in the spatial domain at full resolution. Code-blocks of subbands are then classified as static or active, according to their spatial counterparts in the map. In the current scheme, the subband and resolution to which a code-block belongs does not affect its classification. The coding process is now described:

Step 1. The first frame in a sequence is encoded with non-uniform bitplane truncation due to the lack of a previous frame to help classify it as active or static. Jump to step 5.

Step 2. Tier 1: The entire frame is encoded with non-uniform bitplane truncation to obtain the optimal truncation lengths as in Section 2 satisfying the rate constraint. The inter-frame rate control method of Section 3 yields the final truncation lengths. The resulting *truncation length* is stored.

Step 3. Tier 2: A bisection search to find the optimal EBCOT slope threshold for the *entire* frame is conducted. This way one obtains a *slope threshold* that is optimal for the entire frame and not just the active code-blocks.

Step 4. Then the combination of *truncation length* and *slope threshold* obtained in the previous steps are used to encode the static and dynamic blocks, respectively.

Step 5. If there are more frames left to encode, go to step 2.

In this approach we optimize each parameter, the TRUNC truncation length and the EBCOT slope threshold, separately. The reason is low computational complexity. If P is the cardinality of the pass set and S that of the slope thresholds set, then our approach yields complexity of $O(P) + O(S)$ compared to the $O(P \times S)$ of joint optimization. An optimal solution would involve exhaustive joint optimization of both. However, we believe that for real data the performance loss of our scheme is negligible. Incrementing or decrementing the truncation length by one leads to large drops or increases to the rate allocated to the static region. Hence the selected length is in most cases optimal. A simple ± 1 search yields the optimal value in 99% of the cases. The slope threshold can be varied in a finer grain, but under the constraint that similar rate or quality is afforded to both regions, it is close enough to the optimal value.

5. EXPERIMENTAL RESULTS

The proposed HYBRID and TRUNC rate allocation schemes were simulated using the Kakadu software implementation of Motion JPEG 2000 [5]. The block size for the EBCOT code-blocks was 32×32 pixels due to increased granularity in encoding and the improvement of subjective quality of the reconstructed image [5]. In addition, the recently finalized Digital Cinema Initiative's (DCI) Digital Cinema System Specification (DCCS) Version 1.0 [6] requires a 32×32 code-block size. All evaluated sequences had a fixed duration of 10 seconds. We evaluated three video sequences, one digitized from 35mm film, one taken from a high resolution digital video camera, and a synthetic one:

(a) "Opening" sequence. It consists of the first 10 seconds of the "Opening" sequence of the Standard Evaluation Material (StEM), provided by the American Society of Cinematographers (ASC) and DCI. The original consists of three colour components (RGB) each one of 16-bit accuracy. Prior to encoding we downsample it to 12-bit accuracy as required by the DCI DCCS [6]. The original spatial resolution is 4096×1714, however prior to display, it is cropped to 3840×1714. This sequence has been obtained through digitization of an analog source: 35mm film. As a result it exhibits film noise.

(b) "Landmark" sequence. This sequence has a spatial resolution of 3840×2160 and consists of three colour components (RGB) with 8-bits each. It was first presented at the Consumer Electronics Show (CES) 2005 by JVC. The footage was filmed with a JVC digital camera and as a result is free from film noise.

(c) "CZP" sequence. The synthetic Circular Zone Pattern sequence, displayed in Fig. 1 is very useful due to its symmetry and its inclusion of content that spans *all* spatial frequencies. More importantly, it avoids the spatial masking effects of real video content and enables a thorough investigation of the flicker artifact. The spatial resolution is 512×256 and it consists of a single 8-bit component. We evaluated the proposed schemes both subjectively and objectively.

Subjective Evaluation. The subjective evaluation took place at NTT's Digital Cinema Room outfitted with a state-of-the-art SONY 4K digital cinema projector. The evaluated sequences were displayed at a resolution of 3840×2160 and a rate of 24 frames per second. Viewing conditions were identical to that of a cinema theater room. We encoded the "Opening" sequence at 0.10, 0.25, 0.70, 1.00, and 1.18 bits per pixel (bpp). Using the TRUNC rate allocation, flicker artifact was suppressed for low bit rates. For high bit rates it is not visible since the StEM material is characterized by 35mm film noise. At those bit rates it is impossible to tell flicker from film noise. This motivated us to evaluate "Landmark". The conclusions were largely the same in addition to being able to discern flicker at higher bit rates compared to the StEM sequence. With the use of HYBRID the flicker was suppressed in the static areas and the active areas were encoded at a higher accuracy due to EBCOT.

Finally, we evaluated the "CZP" sequence. The upper half of the sequence remains static for the duration of the sequence, while the lower half moves to the left at a constant speed. Frame 51 is being evaluated in Fig. 1. EBCOT, TRUNC, and HYBRID are compared at 0.40bpp. The frame encoded with our methods (TRUNC and HYBRID) is much more pleasant to the human eye with uniformly distributed quality. One cannot help but notice the blockiness in the EBCOT generated image as well. EBCOT has better performance for high-frequency content but exhibits significant artifacts for low-frequency content. TRUNC is good at low frequencies but suffers from aliasing at higher frequencies. HYBRID on the other hand retains the good performance at low frequencies in the active part (low resolutions were encoded with TRUNC) but also suppresses the aliasing due to the use of EBCOT in encoding the HF subbands of the lower half. The reason the lower frequency subbands were encoded with TRUNC is that they were smaller than the code-block size (for 5 levels of wavelet decomposition) and that we bias the static/active selector to prefer static in these ambivalent cases.

Evaluating the moving sequences we observe no flicker artifact at all in the static regions of HYBRID, while extensive flicker is observed in the EBCOT sequence. The TRUNC scheme, in contrast, exhibits *no* flicker artifact. The temporal flicker is traded-off for the spatial aliasing artifact. Subjective tests showed that the temporal flicker is more noticeable than the spatial aliasing artifact. The lower half active part of HYBRID exhibits slightly more flicker than the EBCOT version, but is largely masked by the motion.

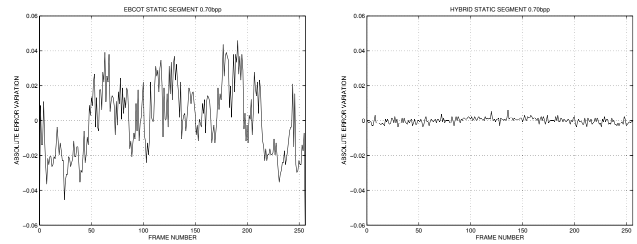
Objective Evaluation. In the first part, we evaluate performance with the help of PSNR. Intuitively, PSNR ought to decrease compared to EBCOT since TRUNC or HYBRID is not optimal in rate-distortion sense. However, the drop in PSNR is an acceptable trade-off for the suppression of the flicker artifact. As shown in Table 1, TRUNC and HYBRID yield average PSNRs very close to EBCOT, in contrast to uniform truncation as proposed in [4]. Due to space constraints we briefly mention that uniform truncation underperforms EBCOT, HYBRID, and TRUNC for the R and G components. The performance for B is better for uniform truncation, however, this is misleading: the Y (in YCrCb space) component is calculated as $Y = 0.299R + 0.587G + 0.114B$; component B is not as significant as R and G. PSNR values were also calculated for sequence "Landmark" and the same conclusions from "Opening" apply in this case as well. They are not included due to space constraints.

bits/pixel	0.1131	0.2500	0.7379	1.0000	1.1274
ebcot *R*	30.23	32.16	34.32	34.85	35.06
ebcot *G*	32.31	34.10	36.36	36.94	37.12
ebcot *B*	27.45	28.30	29.67	30.35	30.66
trunc *R*	29.97	31.99	34.29	34.81	35.32
trunc *G*	32.01	33.93	36.34	36.87	37.17
trunc *B*	27.39	28.22	29.54	29.96	30.46
hybrid *R*	29.99	32.06	34.26	34.78	35.25
hybrid *G*	32.05	33.99	36.32	36.87	37.17
hybrid *B*	27.39	28.26	29.60	30.05	30.57

Table 1. PSNR results for the StEM 240-frame 12-bit sequence "Opening" at 3840×1714-pixels encoded with 32×32 pixels code-blocks. EBCOT rate allocation is compared with HYBRID.

For the synthetic "CZP" sequence the PSNR gap is much larger: The sequence was encoded at 0.7bpp yielding the following PSNR averaged over the entire sequence: EBCOT 26.48dB, TRUNC 20.56 dB, and HYBRID 22.97dB. We observe that the PSNR delta has been reduced in the new scheme. More insight can be gained by looking at the regions: the upper static part and the lower active part. The respective average absolute errors for each codec are: EBCOT 8.88 and 8.89, TRUNC 13.68 and 15.53, and HYBRID 13.66 and 8.99. The errors in the static region are essentially identical for TRUNC and HYBRID.

We also studied the average absolute error over frame index in "CZP". Previous work [2], showed that error variance in static areas is tied to the flicker artifact and is also a direct consequence of



(a) (b)

Fig. 2. Per-frame absolute error statistics for two JPEG2000 rate allocation configurations: (a) EBCOT, (b) HYBRID.

the post-compression rate-control as employed in EBCOT. In Fig. 2, the absolute error variation of the static (upper half) segment of the sequence along the temporal axis (frame number) is plotted. For EBCOT the error varies significantly in time, as a result of rate allocation that behaves differently along subsequent frames even if the content is the same. For HYBRID we observe an almost constant absolute error that is attributed to the unchanged content of the upper half and the almost constant truncation length. This is another indication of flicker artifact suppression.

6. CONCLUSIONS

Flicker is a direct consequence of the post-compression rate-control of JPEG2000 [2]. We suppress flicker by addressing the root cause, the rate allocation algorithm, avoiding pre-processing the source signal as proposed in [3]. We propose a solution to non-uniform error distribution in still images, which is then extended to the temporal domain. We conclude that for *static* or for *translational* motion as in the synthetic "CZP", the TRUNC scheme ensures no flicker artifact. In real video content such as "Opening" and "Landmark" we observe a suppression of the artifact. It is not eliminated completely due to temporal content variations, film noise, and the wavelet kernel length. We note that our proposed schemes are *standard-compliant*.

7. REFERENCES

[1] T. Kuge, "Wavelet picture coding and its several problems of the application to the interlace HDTV and the ultra-high definition images," in *Proc. IEEE International Conference on Image Processing*, June 2002, vol. 3, pp. 217–220.

[2] T. Kato, S. Tanabe, H. Watanabe, and H. Tominaga, "On a relationship between flicker artifacts and quantization error for Motion JPEG2000," in *Proc. FIT2003, J-039*, Sept. 2003.

[3] A. Becker, W. Chan, and D. Poulouin, "Flicker reduction in intraframe codecs," in *Proc. IEEE Data Compression Conference*, Mar. 2004, pp. 252–261.

[4] S. Naito, A. Koike, and S. Matsumoto, "Optimal JPEG2000 encoder mechanism for low delay and efficient distribution of HDTV programs," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, May 2004, vol. V, pp. 81–84.

[5] D. Taubman, "High performance scalable image compression with EBCOT," *IEEE Transactions on Image Processing*, vol. 9, no. 7, pp. 1158–1170, July 2000.

[6] "Digital cinema system specification V1.0," Digital Cinema Initiatives, LLC, Member Representatives Committee, July 2005.