

EFFECTIVE HARDWARE-ORIENTED TECHNIQUE FOR THE RATE CONTROL OF JPEG2000 ENCODING

Te-Hao Chang, Chung-Jr Lian, Hong-Hui Chen, Jing-Ying Chang and Liang-Gee Chen

DSP/IC Design Lab., Graduate Institute of Electronics Engineering and
Department of Electrical Engineering, National Taiwan University, Taipei, Taiwan, ROC
E-mail: {thchang, cjlian, sernc, skii, lgchen}@video.ee.ntu.edu.tw

ABSTRACT

A great deal of computation for JPEG2000 encoding is a redundancy when the compression rate is high. That is because many coded bit-streams will be truncated after the rate control of JPEG2000. In this paper, an effective scheme for JPEG2000 rate control is proposed. Through this scheme, the computation complexity for JPEG2000 entropy coding, that is, EBCOT Tier-1, can be greatly reduced almost without paying any penalty on the image quality, especially at lower bit-rate. Moreover, the proposed method is considered to be suitable for the hardware implementation since the existed techniques for the rate control of JPEG2000 are all software-oriented. By the proposed rate control method, a dedicated hardware of JPEG2000 with high-speed processing and effective rate control ability can be achieved.

1. INTRODUCTION

JPEG2000, the next-generation image compression standard, is well-known as its superior compression performance and novel functions. However, the computation complexity of JPEG2000 is much higher than that of previous image compression standards. According to the profiles reported by [1], EBCOT Tier-1 (Embedded Block Coding with Optimized Truncation [2]) holds the largest portion at the rate of about 45% to 60% of the total encoding time. Therefore, reducing its computation can effectively decrease the processing time of the JPEG2000 system.

Moreover, differing from other lossy compression standards using the quantization scheme, JPEG2000 possesses precise rate control ability by means of the rate-distortion optimization algorithm. Through this optimization, which is performed by EBCOT Tier-2, the image quality of the obtained bit-stream under the specific rate constraint is guaranteed to be optimal. In order to achieve the rate-distortion optimization, however, Tier-1 has to compress all the data even if plenty of them will be truncated by Tier-2 especially under the low bit-rate constraint. That is, a large amount of Tier-1 processing is a waste at high compression rate.

In view of this situation, an estimation scheme was proposed in [3] to predict the truncation points in advance. But the PSNR performance by this scheme is not appreciated since it decreases 1dB or more compared with that of JPEG2000 verification model (briefly VM). It drops to nearly the same quality of the traditional JPEG standard, so the high-quality advantage of JPEG2000 is not existed after applying this scheme.

In this paper, an efficient rate control scheme is presented to considerably reduce the waste of unnecessary computation of

Tier-1. Besides, the method applied in JPEG2000 standard [4] and VM contains iteration operations, which are not suitable for the hardware implementation. However, the proposed method provides the one-pass processing scheme and it is more proper for dedicated hardware design.

This paper is organized as follows. In Section 2, the algorithm of rate control for JPEG2000 is introduced. Section 3 makes a description of the proposed method of Tier-2 rate control, which can greatly reduce the computation of Tier-1 through its feedback processing. The simulation results and analyses are presented in Section 4. Finally, a conclusion is given in Section 5.

2. RATE CONTROL ALGORITHM FOR JPEG2000

JPEG2000 coding system can be separated into three main parts: Discrete Wavelet Transform (DWT), quantization, and then EBCOT. Unlike general compression standards using quantization to perform the rate control, the purpose of quantization in JPEG2000 is close to adjusting the weights among the different frequency sub-bands. In reversible wavelet transform mode, the quantization part even does nothing since the quantization step of each sub-band is set to one. After quantization, the DWT coefficients of each sub-band are partitioned into several non-overlapped code-blocks, which are the independent coding units for EBCOT. EBCOT is composed of Tier-1 and Tier-2 processing. As the name "EBCOT" implies, Tier-1 Embedded Block Coding compresses each code-block into a respective sub-bit-stream by applying the context-based adaptive arithmetic coding technique. After getting all the compressed sub-bit-streams from the Tier-1 process, Tier-2 takes charge of optimized rate control under a given bit-rate and then forms the JPEG2000 file. This algorithm is called Post-Compression Rate-Distortion Optimization (PCRD Opt.), which is the issue of encoder only.

As mentioned above, EBCOT Tier-1 processes independently among the code-blocks. Instead of the pixel-by-pixel scan order for other compression standards, JPEG2000 EBCOT Tier-1 compresses each code-block bit-plane by bit-plane. Furthermore, each bit-plane is scanned by three passes according to the importance of the bit data. That is, the bit data which are more possible to be significant in a bit-plane will be coded in the first pass. The second pass will code the bit data with less importance, and then the remaining data will be handled by the third pass. Each pass of a code-block is a possible truncation point. As the compressed passes from the most significant bit-plane to the least significant bit-plane of a code-block are all connected into a sub-bit-stream, we can expect that no matter which truncation point is selected

for this sub-bit-stream, the quality is optimal since the passes in this code-block is naturally arranged by the order of importance from the most to the least.

At a specific bit-rate, one can set a truncation point for each code-block, and discard all the passes behind them to achieve the target bit-rate. In order to determine the truncation points optimally, JPEG2000 adopts the PCRD Opt. algorithm in EBCOT Tier-2. The concept is to solve the Lagrange criterion to minimize the total distortion of an image at the target bit-rate. Based on PCRD Opt. algorithm, we can calculate the rate-distortion slope value of each pass of the code-block by the following equation,

$$S_i^k = \Delta D_i^k / \Delta R_i^k$$

where ΔD_i^k and ΔR_i^k mean respectively the decrease of Mean Square Error (MSE) and the increase of the number of code bytes between the k-th and (k-1)-th truncation point for the i-th code-block. The slope value of a pass can be considered as the importance of this pass. A pass with larger slope value indicates that receiving this pass can reduce more MSE with fewer byte data. Besides, with the increase of k, the truncation segment is from the first pass to the third pass in a bit-plane, and from the most significant bit-plane to the least significant bit-plane. In this order, the slope values of each code-block have a natural tendency to decrease monotonically, meaning the importance is arranged from high to low. If there are some values of slope that do not follow the rule, we should merge these slopes. That is, we should remove the truncation points to avoid the slopes with the property of $S_i^k > S_i^{k-1}$ for $k-1 > 0$. While all the slopes have been calculated and examined, one can iterate several times finding the minimal threshold value to truncate each code-block for a given bit-rate constraint. In other words, it is to minimize the threshold value and still satisfy the bit-rate constraint after discarding all the passes with the values of slope lower than the threshold value. Afterwards, the image quality of the obtained bit-stream is optimal at this target bit-rate.

3. EFFECTIVE TIER-2 FEEDBACK CONTROL TECHNIQUE

According to the method of rate control mentioned above, we can discover that the slope value of a pass can be calculated just after Tier-1 finishes the arithmetic coding of that pass. However, the iteration for finding the threshold value cannot be launched until all the code-blocks complete Tier-1 coding. For lower bit-rate requirement, a large part of the computation in Tier-1 will be in vain after Tier-2 performs the optimized truncation.

The block diagram of this conventional EBCOT rate control framework is illustrated in Fig. 1(a). To reduce the waste of unnecessary computation in Tier-1, we propose an effective computation reduction scheme through JPEG2000 rate control as shown in Fig. 1(b). The only difference between these two structures is the feedback control module, where the proposed computation reduction technique, called Slope-Byte Table method, is adopted. Fig. 2 shows the detailed flow chart of the proposed method. The main idea is to build a slope-byte table by accumulating the number of bytes of each pass for each slope value and

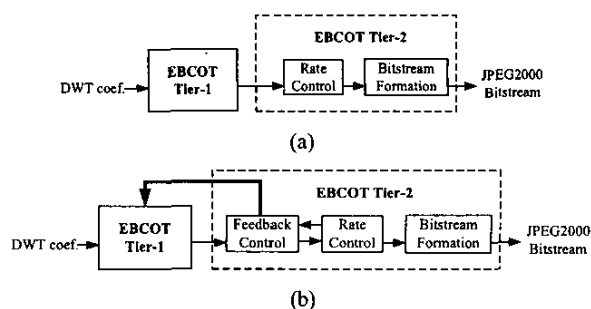


Figure 1. Block diagram of the EBCOT rate control framework.

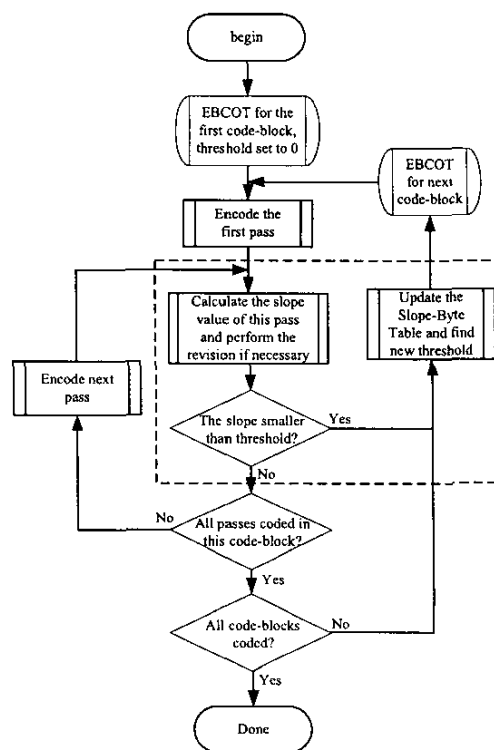


Figure 2. Flow chart of the proposed Slope-Byte Table method.

update it after finishing the coding of each code-block. In VM, a rate-distortion slope value is composed of 15 bits. So there are 32,768 possible values for a slope. First we can construct a table with 32,768 items, each of which accumulates the number of bytes of passes belonging to that value of slope. The slope-byte table has to be updated after each code-block finishes coding. Then we accumulate the number of bytes in the table with the value of slope from the highest to the lowest until the number of accumulated bytes is just larger than the target bit-rate. The corresponding slope will be the threshold of next code-block. If the slope value of the coming coded pass in next code-block is smaller than the threshold value, the passes after that in the code-block need not be encoded since they are not possible to be in

cluded. That is because the current threshold value must be equal to or small than the final optimized threshold value, and we suppose the slope values of the following passes in this code-block are all smaller than that of the current pass based on the monotonically decreasing property of slope values in a code-block. Therefore, after calculating the slope value of each pass, we should check the monotonically decreasing property and perform the revision, namely merging the passes as mentioned before, in case of disobedience.

While the last code-block finishes coding, the final optimized threshold value can be easily obtained just after the last update of the slope-byte table and the target bytes accumulation. As for performing iteration, which is applied by VM, it is not easy to implement on the dedicated hardware. Besides, the redundant computation in EBCOT Tier-1 can also be reduced through the proposed feedback control process. And it is flexible for variant sizes of images because the number of items of the table is fixed.

However, the 32,768 items of the slope-byte table seem to be unnecessarily large. In fact, a truncated slope-byte table can be adopted to reduce the memory requirement. For example, we can build the table with only 1,024 items, that is, taking the front 10 bits of the slope value as the index of the table. After this simplification, the main problem is the precision of the rate control if the table is truncated to be too small. Note that even the rate control is not very precise, the image quality is still optimal for the obtained bit-stream. For general application such as digital still cameras, the rate control is not necessary to be very precise since it may need only three kinds of compression levels. However, if the precision of the rate control is essential for some specific applications, two refinement methods are provided as follows. Then all the statistics and analyses will be detailed discussed in next section.

1. By the accumulating method mentioned earlier, although the exact threshold value cannot be obtained for truncated slope-byte table, we can still get the upper and lower bound of the threshold. In general, the upper bound will be used as the threshold to avoid exceeding the bit-rate. But some of the passes between these two bounds should also be included to achieve the precise rate control. So we can just add these kinds of passes in the order of the sub-bands from lower to higher frequency until the bit-rate is achieved. Because the selection of these few passes by this method does not follow the rate-distortion optimization algorithm, there will be a little PSNR degradation.
2. If the perfect quality performance is also required, a technique, called Minimal Slope Discarding method (briefly Min method) [5], can be applied to achieve perfect rate control without paying PSNR degradation. The concept of the method is to discard the passes with minimal slope value when the buffer overflows during Tier-1 coding. Since this method can also reduce the computation redundancy of Tier-1, by combining it with the proposed table method, the unnecessary computation is reduced even more. However, the Min method needs to search the minimal slope value among all the coded code-blocks, so it will become more complex when the image size is getting larger.

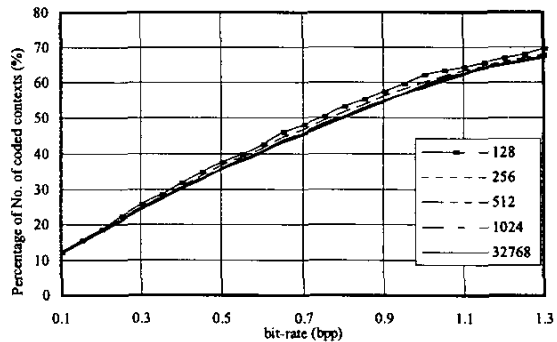


Figure 3. The percentage of computation in Tier-1 through the proposed method with the different table sizes.

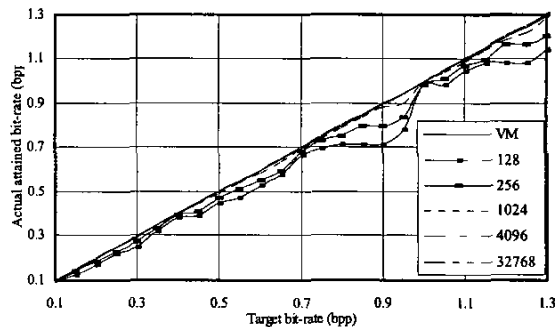


Figure 4. The precision of the rate control through the proposed method with the different table sizes. (The results of the above two figures are the average value of four test images (lena, baboon, pepper and jet) with the size of 512×512. 5 decomposition levels of reversible DWT filter are applied.)

4. EXPERIMENTAL RESULTS AND ANALYSES

By applying the proposed method, the percentage of computation of Tier-1, which is compared with VM 7.2, is shown in Fig. 3. Here, the number of coded contexts is exploited to represent the computation complexity instead of using the run time. This is because the number of coded contexts is more related to the coding cycles of hardware in Tier-1. Through the proposed method, the computation can be greatly reduced since a large part of unnecessary processes are skipped. As for the truncated slope-byte table, the computation reduction abilities among these different sizes are actually quite similar. The difference between the size of 128 and 32,768 is no more than 5%, and it seems to be saturated when the size is larger than 512.

Fig. 4 exhibits the precision of the rate control through the proposed method with different sizes of table without any refinement. Of course, the performance by the non-truncated table (with the size of 32,768) is almost the same as VM. Furthermore, for those applications which do not urge the rate control ability, 1,024 items for the table are quite enough and suitable.

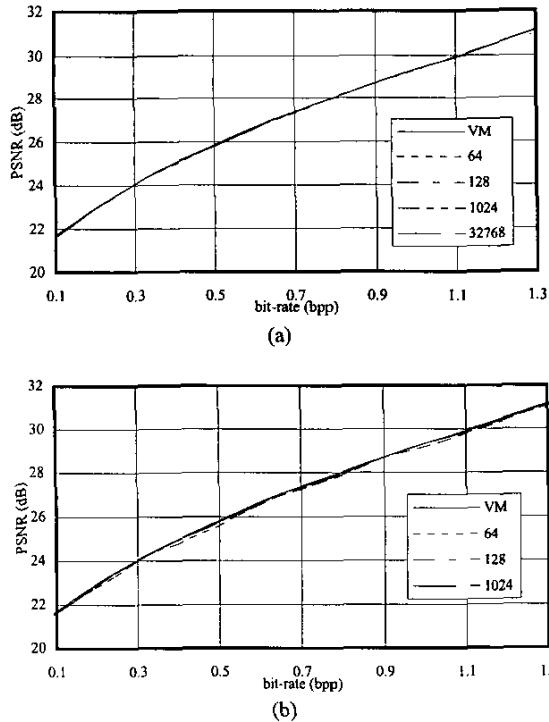


Figure 5. The PSNR performance of the proposed method with different sizes of table compared with VM. (a) without rate control refinement. (b) with rate control refinement. (Test image is "baboon" with the size of 512×512 . Reversible DWT filter with 5-decomposition levels is applied.)

The PSNR performance is illustrated in Fig. 5. Here, we adopt "baboon" as the test image because it is highly textured. As mentioned earlier, the performance between the proposed method and VM is almost the same, with the degradation no more than 0.05 dB. As shown in Fig. 5(a), the PSNR differences among those truncated tables are almost indistinguishable since the problem of the truncated table method is the precision of rate control rather than quality performance. In addition, Fig. 5(b) shows the PSNR performance after applying the first refinement method stated in Section 3. By the refinement, the rate control can be very precise since it is forced to reach the bit-rate. However, a slight number of passes are included but not guaranteed to be optimized. The PSNR descent for the table with the size greatly truncated to 64 is still no more than 0.2dB. The performances for several other test images are all very similar with the presented example, so we don't show here for brevity.

The other rate control refinement scheme for the truncated table method is the combination of the Min method. By blending these two techniques, besides the precise rate control, the computation in Tier-1 can even decrease more, as exhibited in Fig. 6. We can find that the computation of the combination can be reduced about 5% more than that of the table method only. As for the required hardware of Min method, one can utilize the Winner-Take-All

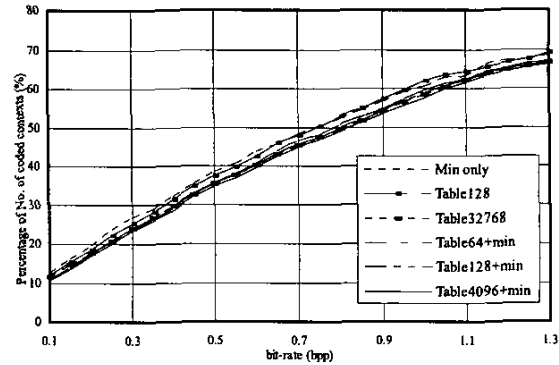


Figure 6. The percentage of computation in Tier-1 through the proposed method with the combination of Min scheme. (The condition is the same as Fig. 4.)

circuits [6] to find the minimal value and its index efficiently. However, it is still not suitable for the images with large or variant sizes.

5. CONCLUSIONS

An effective rate control scheme of JPEG2000 is proposed. Through this method, the computation of Tier-1 coding, which is the heaviest part of JPEG2000, can be reduced greatly by feedback processing. As far as the hardware is considered, the proposed method can perform the rate control with simple operations and constant size of hardware for images with variant sizes. If the table size is considered to be enormous, the truncated table method is also provided. And two refinement methods are supplied for very precise rate control. To sum up, the proposed method can be adopted in a dedicated hardware application of JPEG2000 to achieve the high-speed and lower-power requirement with the rate control ability.

6. REFERENCES

- [1] K. F. Chen et al., "Analysis and architecture design of EBCOT in JPEG2000," in *Proc. IEEE Int. Symp. on Circuits and Systems*, vol. 2, pp. 765-768, 2001.
- [2] D. Taubman, "High performance scalable image compression with EBCOT," *IEEE Trans. on Image Processing*, vol. 9, pp. 1158-1170, July 2000.
- [3] Masuzaki, T. et al., "JPEG2000 adaptive rate control for embedded systems," in *Proc. IEEE Int. Symp. on Circuits and Systems*, vol. 4, pp. 333-336, 2002.
- [4] JPEG-2000 Part I Final Committee Draft Version 1.0, ISO/IEC JTC1/SC29/WG1 N1646R.
- [5] T. H. Chang et al., "Computation reduction technique for lossy JPEG2000 encoding through EBCOT Tier-2 feedback processing," in *Proc. of IEEE Int. Conf. on Image Processing*, vol. 3, pp. 85-88, 2002.
- [6] S. H. Ou et al., "A scalable sorting architecture based on maskable WTA/MAX circuit," in *Proc. IEEE Int. Symp. on Circuits and Systems*, vol. 4, pp. 209-212, 2002.