CONSTANT QUALITY AIMED BIT ALLOCATION FOR 3D WAVELET BASED VIDEO CODING

Zefeng Ni *Jianfei Cai*

School of Computer Engineering
Nanyang Technological University, Singapore 639798
Email: {zfni, asjfcai}@ntu.edu.sg

ABSTRACT

MCTF has been widely used in wavelet based video coding due to its attractive features. For MCTF based codecs, a fundamental question is how to allocate bits to each temporal band so that certain degree of constant quality can be achieved. In this paper, we propose a novel approach for constant quality aimed bit allocation among T-bands for the applications of adaptive stored video streaming. The basic idea of our proposed scheme is to adjust the energy gains based on an empirical model to compensate the different contributions from different types of T-bands, and more or less equally distribute the distortions among the T-bands at the same level. Experimental results show that our proposed bit allocation can greatly reduce the PSNR fluctuation with only slight degradation in average PSNR.

1. INTRODUCTION

With the great success of 2D wavelet in still image coding [1], 3D wavelet that extends 2D wavelet transform to motion pictures has received much attention recently. The current trend of 3D wavelet codecs [2, 3, 4] is to apply motion compensated temporal filtering (MCTF) for temporal decomposition and 2D wavelet for further spatial decomposition. Such codec can provide temporal, spatial and SNR scalabilities simultaneously [4]. Particularly, MCTF has many attractive features. For instance, the lifting design of MCTF allows an efficient implementation of temporal filtering with high orders, and multiple levels of MCTF enables the exploration of multiple-frame redundancies [4]. In addition, the open-loop structure of MCTF significantly reduces the effects of the drifting problem typically associated with traditional hybrid codecs.

For MCTF based codecs, a fundamental question is that, given a bit budget, how to allocate bits to each temporal band (T-band) so that certain degree of constant quality can be achieved. Although the problem of constant quality aimed bit allocation has been well studied for conventional hybrid codecs [5, 6], it is still an open question for MCTF based codecs. Recently, we have seen some research work [7, 8, 9, 10] looking into this issue. In the popular MC-EZBC codec [7], the authors propose to stop bitplane scanning of all the GOPs at the same fractional bit plane. However, it can only help achieve similar distortion among T-bands, which does not lead to constant quality in reconstructed frames. This is because the distortion in T-bands propagate unevenly into reconstructed frames, which is hard to model mathematically. In [8, 9], in order to smooth PSNR behavior, an optimized quantization step is obtained by analyzing the motion behavior during the temporal filtering. In [10], the authors propose to use an adaptive update step for the temporal filtering. Compared with conventional implementation, these methods indeed help reducing the serious PSNR fluctuation in reconstructed frames. However, they did not explicitly address the problem of bit allocation among T-bands.

In this paper, we propose a novel approach for constant quality aimed bit allocation among T-bands for the applications of adaptive stored video streaming. In particular, we develop a 3D wavelet codec based on MCTF and JPEG2000. The reconstructed frames are divided into different groups according to the types of their associated temporal bands. We propose an approximate mathematical model to describe the relationship between the T-band distortions and the distortions of the reconstructed frames. Since we consider stored video in this research, we can offline generate the model parameters. During the online transmission stage, given the current network bandwidth, we first perform the conventional JPEG2000-like optimum truncation. After that, a two-step operation is used for reducing the PSNR fluctuation, where the basic idea is to adjust the energy gains to compensate the different contributions from different types of T-bands and then more or less equally distribute distortion among the T-bands at the same level. Experimental results show that our proposed scheme can greatly reduce PSNR fluctuation with only slight degradation in average PSNR.

2. PROBLEM STATEMENT

In order to study how to allocate bits among T-bands, we need to understand the process of MCTF first. Fig. 1 shows two levels of MCTF with the popular 5/3 wavelet. In particular,

MCTF is implemented by the two-step lifting scheme. In the prediction steps, the input frames are taken to generate high-pass bands H_k, while in the update steps the original even frames and the obtained high-pass bands are used to generate the low-pass bands L_k. Recursively applying this temporal filtering to the low-pass bands obtained in the previous level, we generate many T-bands at different levels. For example, three levels of MCTF produce four types of T-bands: LLL, LLH, LH and H bands. It is clear that the same distortion associated with different T-bands will contribute differently to the distortion in reconstructed frames. Typically, this difference is compensated by introducing synthesis filter gains in the conventional bit allocation schemes[11].

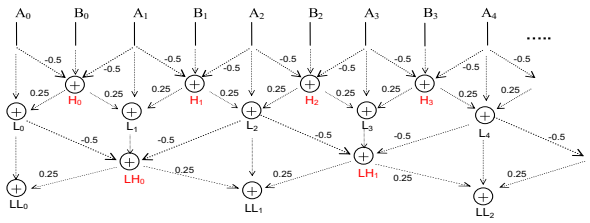

Fig. 1. Two Levels of MCTF with 5/3 wavelet

In order to obtain the synthesis filter gains, a simple mathematic model is commonly used to describe the relationship between the reconstructed frames and the T-bands. In particular, considering one-level MCTF[1] in Fig. 1, we have

$$A_k = L_k - \frac{1}{4}H_k - \frac{1}{4}H_{k-1} \tag{1}$$

and

$$\begin{aligned} B_k &= H_k + \frac{1}{2}A_k + \frac{1}{2}A_{k+1} \\ &= \frac{3}{4}H_k + \frac{1}{2}L_k + \frac{1}{2}L_{k+1} - \frac{1}{8}H_{k-1} - \frac{1}{8}H_{k+1} \end{aligned} \tag{2}$$

Further assuming the distortion in each T-band is *zero-mean additive white noise*, then the distortion of reconstructed frames measured in mean square error (MSE) can be expressed as

$$\sigma^2_{\epsilon_{A_k}} = \sigma^2_{\epsilon_{L_k}} + \frac{1}{16}\sigma^2_{\epsilon_{H_k}} + \frac{1}{16}\sigma^2_{\epsilon_{H_{k-1}}} \tag{3}$$

and

$$\sigma^2_{\epsilon_{B_k}} = \frac{9}{16}\sigma^2_{\epsilon_{H_k}} + \frac{1}{4}\sigma^2_{\epsilon_{L_k}} + \frac{1}{4}\sigma^2_{\epsilon_{L_{k+1}}} + \frac{1}{64}\sigma^2_{\epsilon_{H_{k-1}}} + \frac{1}{64}\sigma^2_{\epsilon_{H_{k+1}}} \tag{4}$$

where $\sigma^2_{\epsilon_F}$ denote the mean error variance of a T-band F. Applying the distortion relationship between reconstructed frames and the T-bands to each level of MCTF, we can derive the energy gains. For example, the energy gain G_L for L bands is

$$G_L = 1 + 1/4 + 1/4 = 1.5. \tag{5}$$

[1]The motion vectors are omitted here.

Similarly, we have the energy gains for other T-bands, such as

$$G_H = 9/16 + 1/16 + 1/16 + 1/64 + 1/64 \approx 0.72, \tag{6}$$

and

$$G_{LH} = G_L \times G_H \approx 1.08. \tag{7}$$

With the energy gains, we can easily extend conventional bit allocation approaches to allocate bits among T-bands. Considering a scalable codec we have developed based on MCTF and JPEG2000, i.e. generating T-bands using MCTF and encoding each T-band using JPEG2000 with many quality layers, a straightforward T-band bit allocation is to consider the energy gains in bit allocation as in JPEG2000's optimum truncation [1] and use each quality layer as a potential truncation point. In fact, we have implemented such straightforward T-band bit allocation scheme and we call it JPEG2000-like bit allocation. Fig. 2 shows the PSNR performance of the Stefan sequence coded at 1 Mbps with two-level MCTF using the JPEG2000-like bit allocation scheme. We can see that there exist large quality fluctuations. In general, the frames, A0, A2,..., have the highest PSNR and the frames, B0, B1,..., have the lowest values.

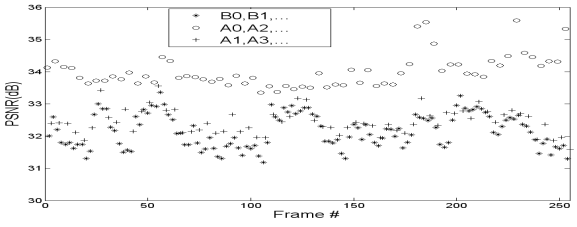

Fig. 2. PSNR fluctuation for CIF Stefan at 1 Mbps.

One obvious reason for PSNR fluctuation is that the optimum bit allocation is for minimizing the average distortion instead of the quality difference. Thus, we change the bit allocation among T-bands according to the distortions calculated by Eqs. (3) and (4) instead of searching the slopes along the R-D curves. However, our experiments show that the large PSNR fluctuation still remains. In fact, some studies [11] have pointed out that, due to the incorrect assumption of white noise and ignoring the motion compensation and the correlations among T-bands, the mathematic model given in Eqs. (3) and (4) is not accurate at all. Fig. 3 shows an example of the distortion relationship in coding the CIF Stefan sequence using our developed codec. The X-axis is the distortion change in L_k while Y-axis is the corresponding distortion change in the reconstructed frame B_k. It indicates that the coefficient $1/4$ in Eq. (4) is inaccurate. Similarly, we can also empirically prove that neither the other coefficients nor the additive property in Eqs. (3) and (4) are accurate. Therefore, the obtained energy gains cannot be accurate.

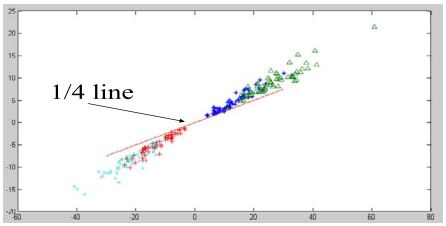

Fig. 3. The distortion relationship between L_k and B_k.

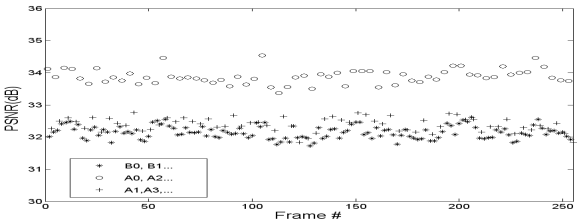

Fig. 4. Smoothing MSE of T-bands for CIF Stefan.

3. CONSTANT QUALITY AIMED BIT ALLOCATION

Ideally, a mathematical model is needed in order to achieve constant quality bit allocation. However, it is non-trivial to develop such a model. First, due to the complex motion compensation process, distortion at different regions of a T-band propagates differently to reconstructed frames. A precise model would require distortion analysis at pixel or block level. Then we will have to consider finer bit allocation scheme, such as bit allocation among small code-blocks of 3D wavelet coefficients. In addition, the unknown correlation among T-bands further complicates the distortion relationship between reconstructed frames and T-bands, and hence makes it difficult to estimate the distortion of individual frames. Here, we propose a simple approach to reduce quality fluctuation.

One reason for PSNR fluctuation is the uneven MSE distribution among the T-bands at the same level. Intuitively, the distortion of the T-bands at the same level should be close to each other since they contribute more or less equally to the distortion of the reconstructed frames. Therefore, for our develop codec, we propose to readjust bit allocation after the JPEG2000-like optimum truncation. In particular, we first perform the optimum bit allocation, through which we obtain not only the selected truncation points for each T-band but also the other feasible truncation points. Then, for a T-band level l, we calculate the average MSE $D_{avg}(l)$ of all the T-bands at this level. After that, for each T-band at level l, we adjust the selected truncation point to a neighboring feasible truncation point that has a closer MSE to $D_{avg}(l)$. Fig. 4 shows the resulted PSNR. We can see that the proposed readjustment indeed helps smoothing the PSNR fluctuation within the same class of frames (such as A0,A2, A4, ...). Since this adjustment is performed towards the average MSE of T-bands at the same level, the total number of bits does not change much (usually less than 5%). Of course, the average PSNR is dropped because the new bit allocation is not targeted to minimize the average distortion any more. However, in our experiment, we observe this drop is usually less than 0.3dB.

Although the adjustment can smooth the PSNR fluctuation within the same class of frames, there still exists large PSNR difference among different classes of frames. For example, the group of A0, A2,... frames has the highest PSNR in Fig. 4. This is because this group of frames are directly related to the lowest T-bands, i.e. LL bands, which has the highest energy gain. One way to solve this problem is to empirically modify the energy gains until the average PSNR difference among different classes of the reconstructed frames becomes very small. However, to search the energy gains for each different bit rate is intolerable for adaptive online transmission. Therefore, in this paper, we propose another way to find the appropriate energy gains. Specifically, we modify the model in Eqs. (3) and (4) into

$$\sigma^2_{\epsilon_{A_k}} = \alpha_{Ak} \cdot \sigma^2_{\epsilon_{L_k}} + \beta_{Ak} \cdot \frac{1}{16}\sigma^2_{\epsilon_{H_k}} + \gamma_{Ak} \cdot \frac{1}{16}\sigma^2_{\epsilon_{H_{k-1}}}, \quad (8)$$

and

$$\sigma^2_{\epsilon_{B_k}} = \alpha_{Bk} \cdot \sigma^2_{\epsilon_{H_k}} + \beta_{Bk} \cdot \frac{1}{16}\sigma^2_{\epsilon_{A_k}} + \gamma_{Bk} \cdot \frac{1}{16}\sigma^2_{\epsilon_{A_{k+1}}}, \quad (9)$$

where $\alpha_{Ak}, \beta_{Ak}, \gamma_{Ak}, \alpha_{Bk}, \beta_{Bk}, \gamma_{Bk}$ are positive scaling factors. Clearly, this modified model is also inaccurate to estimate the distortions of individual frames since we still ignore the correlation among T-bands. Nevertheless, it is sufficient to estimate the average distortion among one class of frames. In particular, we offline compute the unknown parameters in Eqs. (8) and (9) by multiple encoding and decoding, and record these parameters for future adaptation. Then, during online transmission, we can easily test different energy gains and use Eqs. (8) and (9) to calculate the average distortions among different classes of frames. We select the set of energy gains that make the average distortions among different classes of frames close to each other. Note that we only perform multiple-pass coding at the offline stage.

In short, our proposed bit allocation scheme can be summarized as follows.

- Step 1. During the JPEG2000-like optimal truncation, modify the energy gains so that the average estimated distortions for different classes of frames using Eqs. (8) and (9) are close to each other. With this, we bypass complex distortion analysis between T-bands and reconstructed frames.
- Step 2. Readjust the bit allocation to smooth the MSE distribution among the T-bands at the same level. This helps to take the distortion value into account rather than only considering the slope of the rate-distortion curve in typical Lagrange-based optimum bit allocation.

4. EXPERIMENT RESULTS

We test our method for the first 256 frames of the CIF Stefan sequence with two levels of MCTF. Fig. 5 shows the PSNR result. We can see that, compared with the original JPEG2000-like bit allocation, the proposed scheme achieves a much smoother PSNR in the reconstructed frames. Specifically, the PSNR variance is reduced from 0.86 to 0.12 while the average PSNR is dropped about 0.4 dB, which is acceptable in practice. We have also tested other video sequences including CIF Foreman and CIF Table Tennis. Similar performance can be observed. Table 1 summarizes the results for different sequences at different bit rates. In general, our proposed scheme can reduce PSNR variance $3 \sim 10$ times while average PSNR degradation is not more than 0.6 dB.

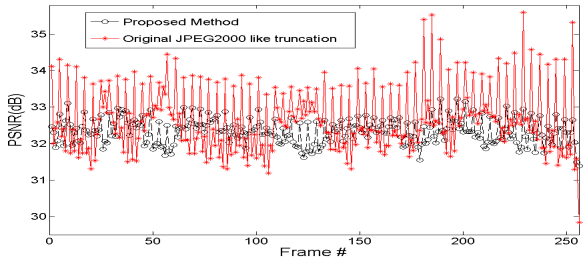

Fig. 5. PSNR result of the CIF Stefan at 1 Mbps.

5. CONCLUSION

In this paper, we have proposed a novel approach for constant quality aimed bit allocation among T-bands. We first analyzed the process of MCTF and studied the conventional JPEG2000-like bit allocation. After observing the problem of the significant PSNR fluctuation, we have pointed out the causes in the conventional bit allocation. Based on our analysis, we have proposed to adjust the energy gains of T-bands at different levels in order to make the distortions of different classes of reconstructed frames close to each other. Furthermore, by smoothing the MSE allocation for T-bands at the same level, we have further reduced the PSNR fluctuation for reconstructed frames in the same class. Experimental results have demonstrated the effectiveness of the proposed method.

Table 1. Change in Variance and average of PSNR

Video (CIF) Sequence	Bit Allocation Scheme	Average PSNR (dB)	PSNR Variance
Stefan (800kbps)	JPEG2000-like	31.5	0.78
	Proposed	31.2	0.13
Stefan (1Mbps)	JPEG2000-like	32.7	0.86
	Proposed	32.3	0.12
Stefan (1.2Mbps)	JPEG2000-like	34.0	1.62
	Proposed	33.4	0.23
Foreman (250kbps)	JPEG2000-like	32.2	1.33
	Proposed	31.9	0.08
Foreman (400kbps)	JPEG2000-like	34.1	0.84
	Proposed	33.7	0.08
Foreman (500kbps)	JPEG2000-like	35.0	0.96
	Proposed	34.7	0.11
Table tennis (300kbps)	JPEG2000-like	30.6	0.68
	Proposed	30.2	0.16
Table tennis (450kbps)	JPEG2000-like	32.4	0.55
	Proposed	32.0	0.17
Table tennis (600kbps)	JPEG2000-like	33.7	0.62
	Proposed	33.1	0.20

6. REFERENCES

[1] "ISO/IEC 15444-1: Information technology - JPEG2000 image coding system - part 1: core coding system," 2000.

[2] A. Secker and D Taubman, "Motion-compensated highly scalable video compression using an adaptive 3D wavelet transform based on lifting," *Proc. Int. Conf. Image Processing*, pp. 1029–1032, 2001.

[3] L. Luo, J. Li, S. Li, Z. Zhuang, and Y.-Q. Zhang, "Motion compensated lifting wavelet and its application in video coding," *Proc. Int. Conf. on Multimedia & Expo (ICME)*, pp. 365–368, 2001.

[4] J-R Ohm, "Advances in scalable video coding," *Proc. of The IEEE*, vol. 93, no. 1, pp. 42–56, Jan. 2005.

[5] X. M. Zhang, A. Vetro, Y. Q. Shi, and H. Sun, "Constant quality constrained rate allocation for FGS-coded video," *IEEE Trans. Circuits Syst. Video Technol.*, p. 121C130, Feb. 2003.

[6] J. Cai, Z. He, and C. W. Chen, "A novel frame-level bit allocation based on two-pass video encoding for low bit rate video streaming applications," *accepted by Journal of Visual Communications and Image Representation*.

[7] P. Chen and J. W. Woods, "Bidirectional MC-EZBC with lifting implementation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 10, pp. 1183–1194, Oct. 2004.

[8] K. Hanke, J. R. Ohm, and T. Rusert, "Adaptation of filters and quantization in spatio-temporal wavelet coding with motion compensation," *Proc. Picture Coding Symposium (PCS)*, pp. 49–54, 2003.

[9] T. Rusert, K. Hanke, and J. R. Ohm, "Transition filtering and optimized quantization in interframe wavelet video coding," *Proc. Visual Communication and Image Processing (VCIP)*, pp. 682–393, 2003.

[10] A. Mavlankar, S. E Han, C. L. Chang, and B. Girod, "A new update step for reduction of PSNR fluctuations in motion-compensated lifted wavelet video coding," *Proc. IEEE Int. Workshop on Multimedia Signal Processing (MMSP)*, 2005.

[11] R. Xiong, J. Xu, F. Wu, S. Li, and Y.-Q. Zhang, "Optimal subband rate allocation for spatial scalability in 3D wavelet video coding with motion aligned temporal filtering," *Proc. Visual Commun. Image Processing (VCIP)*, pp. 381–392, 2005.