A TEMPORAL QUALITY SMOOTH METHOD FOR WAVELET VIDE CODERS

Jiann-Jone Chen and Kun-Wei Wang

Dept. of Electrical Engineering, National Taiwan University of Science & Technology,
43 Keelung Rd., Sec. 4, Taipei Taiwan 10673
Email:jjchen@mail.ntust.edu.tw

ABSTRACT

The wavelet-based video codec is capable of providing scalable bitstreams such that quality of service (QoS) among heterogeneous networks can be manipulated in a unified framework. The three-dimensional wavelet video coder 3D WVC had been developed to provide this fully scalable bitstream. However, the quality control or rate allocation was performed on the scale of one GOP such that the reconstructed pictures suffer temporal quality fluctuations. The theoretical coding model provided ratio parameters among subbands to smooth temporal picture quality. However, it needs further adaptation for dealing with practical WVC. We propose to investigate the rate-distortion relation, between the decomposed subbands ($\mathbf{\Psi}$) and the reconstructed pictures in one GOP ($\mathbf{P}$), under which the rate allocation for constant quality (RACQ) can be operated. Experiments show that the temporal quality fluctuations can be reduced to 10-30 times smaller as compared to 3D-SPIHT and 5-10 times smaller to the ratio parameter approach.

1. INTRODUCTION

The wavelet-based scalable video coder has been developed by many [1] [2] as an alternative to predictive coding approaches, e.g., MPEG-4 or H.264. By investigating the temporal correlation in addition to 2D wavelet coding, the 2D+t (or 3D) wavelet video coders demonstrate fine granularity properties in temporal, spatial or quality displaying dimensions. Previous works on 3D wavelet video coders provide frameworks [1] [2] to assemble the embedded scalable bitstream. These frameworks aim to achieve high compression efficiency for scalable videos in which a group of pictures (GOP) is the basic processing unit to demonstrate scalable capabilities. Since the embedded bitstream can be truncated at any point to fit the rate constraints, the QoS problem can be simplified to how to allocate bits among subband images. In JPEG2000, EBCOT [3] uses the Lagrangian technique to construct the embedded bitstream such that any truncation points are optimal. The EBCOT coding approach can be extended to 3D wavelet video coder [4] to provide temporal and spatial scalable bitstreams. Note that global optimal bit allocation for one GOP, i.e., 3D WVC, may not present good subjective performances due to temporal quality fluctuations [5]. If the temporal correlation was not exploited, such as that in the motion JPEG2000, the temporal quality fluctuation can be reduced by controlling the picture bitstream under the buffer-channel constraints [6]. This rate control algorithm is similar to EBCOT, operated under independent R-D function, but is proceeded with larger video units. Another [7] proposed to find the rate-distortion ratio among temporal subbands to smooth the temporal video quality. A set of ratio parameters was derived based on the filtering process of Daubechies 9-7 wavelets. Notwithstanding the assumed $i.i.d.$ property among subbands and high bit rate coding model may not be satisfactory in practical coding, the ratio parameters provide a good initial bit allocation for RACQ.

We propose to investigate the relation from the rate allocation in Ψ to the error distribution in reconstructed pictures $\tilde{P}$ for performing the RACQ algorithm. The analysis/synthesis framework of the scalable 3D WVC [2] is utilized to formulate this rate-distortion mapping relation, denoted as $\mathfrak{m}_S$. A rate allocation algorithm which targets constant picture quality can then be developed based on this $\mathfrak{m}_s$ function.

This paper is organized as follows. Section 2 discuss the theoretical ratio parameters for constant quality video. The rate-distortion relation among decomposed subbands Ψ and reconstructed pictures $\tilde{P}$ of the scalable 3D WVC and the RACQ algorithm are presented in Section 3. Section 4 is the experimental study. Section 5 concludes this paper.

2. THEORETICAL CODING CONTROL MODEL

For 3D wavelet video coders, potential temporal PSNR fluctuations in the reconstructed picture can be eliminated by adjusting the slopes of rate-distortion operation points among subbands. In high bit rate coding, the distortion relationship between wavelet filter coefficients and reconstructed data

This work was partially supported by the National Science Council of Taiwan with grant no: NSC 93-2213-E011-070.

0-7803-9332-5/05/$20.00 ©2005 IEEE

is [3]:

$$\tilde{d}(2n) = d(L)\sum_i f'^2_{2i} + d(H)\sum_j g'^2_{2j+1} \quad (1)$$

$$\tilde{d}(2n+1) = d(L)\sum_i f'^2_{2i+1} + d(H)\sum_j g'^2_{2j},$$

where g' and f' are synthesis high-pass and low-pass filter coefficients respectively and $d(H)$ and $d(L)$ are the coding errors, in mean square sense, of high and low subband respectively. To make reconstructed data with constant error, i.e., $\tilde{d}_{2n} = \tilde{d}_{2n+1} = \tilde{d}$, it can be derived that [7]

$$\frac{d(L)}{d(H)} = \frac{\sum_i g'^2_{2i+1} - g'^2_{2i}}{\sum_i f'^2_{2i+1} - f'^2_{2i}}. \quad (2)$$

For k level 1D wavelet decomposition, there will be at most $M = 2^k$ subbands. Let $S = \{s^0 s^1 \cdots s^{M-1}\}$ denote the corresponding synthesis basis matrix that s^i is the basis for coefficients in position $Mk+i$. Based on the $i.i.d.$ assumption, the relation between distortion of the reconstructed data and the wavelet coefficients is

$$\tilde{D} = S \cdot D, \quad (3)$$

where $\tilde{D} = \{\tilde{d}_0 \tilde{d}_1 \cdots \tilde{d}_{M-1}\}$, $D = \{d_0 d_1 \cdots d_{M-1}\}$ and $S = \{\sum_k s^{i\,2}_{j-i}\}$ [7] respectively. In high rate coding, the rate distortion operation slopes can be assumed to be proportional to subband distortions, i.e.,

$$\lambda^0 : \lambda^1 : \cdots : \lambda^{M-1} = d(0) : d(1) : \cdots : d(m-1). \quad (4)$$

By controlling R-D operation slopes (and hence D) according to the specified filters S, it can lead to smooth temporal video quality, i.e., $\tilde{D} = \mathbf{I} = \{11 \cdots 1\}$. The ratio parameter in eq. 4 is derived under high-bit rate assumptions and is analyzed in statistical approaches. Further adjustments are required when performing bit allocation in GOP-based coding framework. Experiments show that the ratio parameter approach (**STD-EWVC**) did provide good initial bit allocations for constant quality video. However, the adaptation capability should be exploited to accommodate the non-ideal properties of video signals.

3. SCALABLE WAVELET VIDEO CODER

3.1. Rate-Distortion Relation in 3D WVC

The temporal analysis, for example, of an eight picture GOP in 3D WVC is shown in Fig. 1. Solid rectangular boxes without shading denote the source pictures of the *k*-th GOP, $P_k = \{p_k(i) = p(i + (k-1) \cdot N)\}_{i=1,\cdots,N}$. Let $l_{i,j}$ and $h_{i,j}$ be the *j*-th pair of the approximate and detail pictures on the *i*-th decomposition level. Final subband pictures,

$$\Psi_k = \{l_{log_2 N,1}, h_{log_2 N,1}\} \cup$$

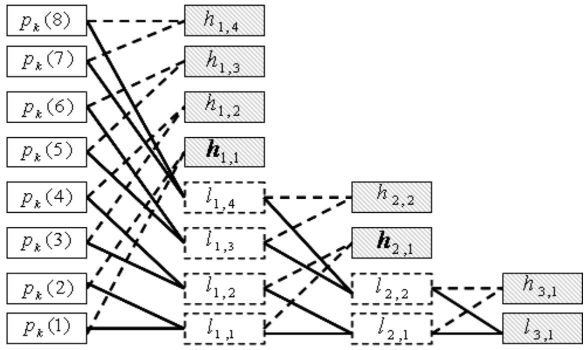

Fig. 1. The temporal analysis/synthesis of one GOP in wavelet vide coder.

$$\{h_{i,j} | i = log_2 N - 1, \cdots, 2, 1, j = 1, 2, \cdots, \lceil \frac{N}{2^i} \rceil\} \quad (5)$$

under which the RACQ would be operated, are marked as shaded boxes. Intermediate subband pictures are marked as dashed boxes. Solid and dotted lines between subband pictures denote the low- and high-pass analysis filtering in the temporal decomposition process respectively. Note that motion compensated temporal filtering (MCTF [1]) could be utilized before performing the decomposition. It needs to transmit the motion vector information but expects a higher compression ratio.

In this 3D WVC, the control target to smooth quality among reconstructed pictures in $\tilde{P}_k$ is achieved by performing proper rate allocation in Ψ_k. The mapping between $\tilde{P}_k$ and Ψ_k in the synthesis (reconstruction) process of the 3D WVC (Fig. 1) is found to be

$$\mathfrak{M}_S(\psi_{i,j}) = \{\tilde{p}_k(m) | m = m_b + 1, m_b + 2, \cdots, m_b + 2^i\}, \quad (6)$$

where $m_b = 2^i \cdot (j-1)$, $\psi_{i,j} \in \Psi_k$ and $\mathfrak{M}_S(\psi_{i,j}) \subset \tilde{P}_k$ would return the set of pictures in which the rate adjustment in $\psi_{i,j}$ would distribute. For example, if rates allocated to $h_{2,1}$ were adjusted, then the resultant error distribution would include pictures specified by $\mathfrak{M}_S(h_{2,1}) = \{\tilde{p}_k(1), \tilde{p}_k(2), \tilde{p}_k(3), \tilde{p}_k(4)\}$.

3.2. Rate Allocation for Constant Quality

In JPEG2000 EBCOT [3], the Lagrange multiplier optimization process is used to solve the following rate-distortion optimization problem:

$$\min_{R(i), n \le i \le n+N-1} \sum_{i=n}^{n+N-1} D(i),$$

$$subject \quad to \quad \sum_{i=n}^{n+N-1} R(i) \le R_b, \quad (7)$$

where $R(i)$ and $D(i)$ are the rate-distortion (in mean square error) pair for the *i*-th codeblock of subband images. Since the rate-distortion functions $R(D)$s among codeblocks are independent, the Lagrange multiplier optimization procedure can be operated on these $R(D)$s directly. The total distortion in subband codeblocks, $\sum_i D(i)$ equals to that in the spatial domain, i.e.,$\sum_{x,y}(p(x,y) - \tilde{p}(x,y))^2$. However, the optimal bit allocation in one GOP of WVC may lead to unpleasant quality fluctuations [5] among reconstructed pictures in $\tilde{P}$. In addition, the quality in the spatial domain $\tilde{P}$ does not correspond to the quality modifications in subband pictures Ψ. The STD-EWVC [7] provides ratio parameters $\alpha_{ij}s$, i.e., $\lambda_i = \alpha_{ij} \cdot \lambda_j$ for $i \neq j$, to utilize the unconstrained Lagrangian optimization process. However, the theoretical approach to yield constant temporal quality may need to be modified due to non-ideal signal properties. We propose to perform rate adjustment, based on the initial bit allocation, among subband images Ψ_k to provide practical smooth video quality in $\tilde{P}$. This RACQ algorithm is described below.

1. For each N-picture GOP, say P_k, it performs the temporal decomposition with the framework in Fig. 1 and returns N subband pictures specified by Ψ_k in eq. (5).

2. With the bit budget for one GOP, it performs the unconstrained Lagrange multiplier optimization based on R-D relation of pictures in Ψ_k. Outputs of this step are denoted as:

$$\{r(l_{log_2 N,1}), r(h_{i,j})|_{i=log_2 N,\cdots,2,1,j=1,2,\cdots,\lceil\frac{N}{2}\rceil})\}, \tag{8}$$

   where $r(\psi_{i,j})$ denotes the rate allocated for subband picture $\psi_{i,j}$.

3. For $i = log_2 N - 1, \cdots, 2, 1$,
   for $j = 1, 3, \cdots, \lceil\frac{N}{2^i}\rceil - 1$, {
   $\Delta_Q = Q_{ave}\{\mathfrak{M}_S(h_{i,j})\} - Q_{ave}\{\mathfrak{M}_S(h_{i,j+1})\}$;
   while ($|\Delta_Q| > T_Q$) {
   if ($\Delta_Q > 0$) move δ_b bits from $r(h_{i,j})$ to $r(h_{i,j+1})$;
   else move δ_b bits from $r(h_{i,j+1})$ to $r(h_{i,j})$;
   }
   },
   where T_Q is the threshold of acceptable quality fluctuations and $Q_{ave}(\mathfrak{M}_s(\psi_{i,j}))$ denotes the average PSNR of the reconstructed pictures in $\mathfrak{M}_s(\psi_{i,j})$.

Note that the initial bit allocation may dominate the Q_{ave} but it would not affect much in temporal quality smoothness performed by RACQ. In addition to this initial bit allocation, the scale of other control factors, δ_b and T_Q, would also have impacts on the PSNR variations in $\tilde{P}_k$. From our experiments, the ratios $\lambda_{LLL} : \lambda_{LLH} : \lambda_{LH} : \lambda_H = 2.827 : 1.999 : 1.414 : 1$ provides satisfactory performances either in Q_{ave} or PSNR variances.

4. EXPERIMENTAL STUDY

The Haar wavelet basis is used to temporally decompose pictures in P. To provide scalable video bitstreams, the 2D SPIHT [8] wavelet coder is used to encode subband pictures in Ψ (eq. (5)). In evaluating the performance of the proposed RACQ algorithm, the PSNR performance of the high bit rate modeling approach STD-EWVC is provided for comparisons.

The test sequences are QCIF (176×120) "football" (high activity) and "table tennis" (medium activity) at rate 30 picture/sec. The compression ratios are selected to be 0.6 bpp and 0.3 bpp (bits per pixel) respectively. As shown in Fig. 2(a), the PSNR variances are largely reduced by the RACQ algorithm (thick gray line) as compared to the 3D-SPIHT (solid line). For one GOP, the dynamic ranges are reduced from 3-4 dB to be within 0.5 dB. As compared to the high bit rate approach STD-EWVC (dotted line), the RACQ also demonstrates much smoother temporal picture quality. Fig. 2(b) demonstrates the PSNR performance at 0.3 bpp. As shown, the proposed RACQ algorithm still provides stable and smooth temporal quality as compared to the other two algorithms. Numerical analyses show that the temporal picture quality variations can be reduced to 10-30 times smaller as compared to 3D-SPIHT and 5-10 times smaller to STD-EWVC. To further inspect coding control performances, the MCTF process [1] was performed prior to utilizing RACQ algorithm. As shown in Fig. 3, the MCTF-RACQ did improve the average PSNR and the temporal quality smoothness. Further simulations demonstrate that the MCTF does not improve much in the overall PSNR or temporal quality smoothness since the operational functions of MCTF and temporal video decomposition are almost the same.

Note that in performing the RACQ algorithm under the 3D WVC system, any rate adjustment among subband pictures would distribute errors with units of pairwise pictures so that quality variations between the pairwise pictures in $\tilde{P}$ can not be eliminated further. In addition, the RACQ may sometimes be operated in the rate-distortion boundary such that no further rate adjustment toward constant quality is possible. These two factors, hindering further rate adjustment toward constat quality, are unavoidable by nature in the 3D WVC. As shown in the 3rd, 4-th and 5-th GOPs (16 pictures) of Fig. 3, the temporal picture quality could not be further smoothed even large dynamic quality ranges remain.

5. CONCLUSIONS

We had proposed a quality control method for the fully scalable 3D WVC. For one GOP, the rate-distortion relation $\mathfrak{M}_S$ between temporally decomposed subbands and reconstructed pictures had been exploited. By utilizing this re-

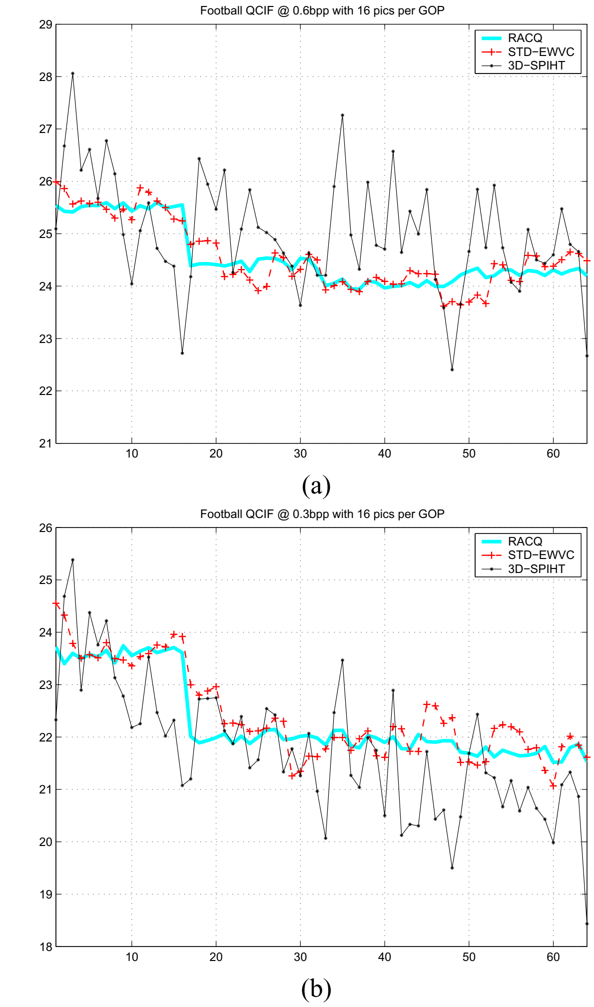

Fig. 2. Temporal PSNR performance among RACQ, STD-EWV & 3D-SPIHT on the “football” seq.

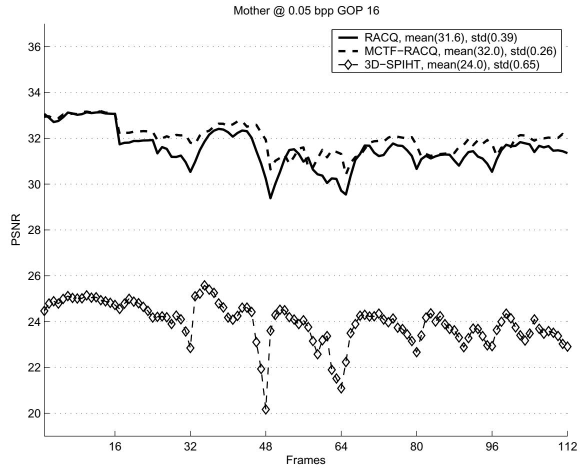

Fig. 3. Temporal PSNR performances among RACQ, MCTF-RACQ and 3D-SPIHT on the “mother and daughter” seq.

lation $\mathfrak{m}_S$, the qualities among reconstructed pictures in $\tilde{P}$ can be controlled by performing the rate control in subband pictures. The control algorithm begins with initial bit allocation under the high bit rate modeling (STD-EWVC) and then performs the quality adjustment in the hierarchical approach. Experiments show that the proposed algorithm largely reduces the temporal PSNR fluctuations down to 10-30 times smaller than that in the 3D-SPIHT and 5-10 times smaller than the STD-EWVC. The most distinguished feature of RACQ is that it provides a robust coding control method for fully scalable wavelet video codecs.

6. REFERENCES

[1] J.-R. Ohm, “Three-dimensional subband coding with motion compensation,” *IEEE Trans. Image Processing*, vol. 3, no. 5, pp. 559-571, 1994.

[2] B.-J. Kim, Z. Xiong and W. A. Pearlman, “Low bit-rate scalable video coding with 3-D set partitioning in hierarchical trees (3-D SPIHT),” *IEEE Trans. CSVT.*, vol. 10, no. 8, Dec. 2000.

[3] D. S. Taubman et al, “JPEG2000: Image Compression Fundamentals, Standards, and Practice,” Kluwer International Series in Engineering and Computer Science, 642.

[4] L. Luo et al, “Compression of concentric mosaic scenery with alignment and 3D wavelet transform,” *SPIE VCIP.*, vol. 3974-10, Jan. 2000.

[5] P.-Y. Cheng, “Rate-distortion optimized bit rate control scheme for a wavelet video coder,” *IEEE Trans. CSVT.*, vol.7, no. 4, pp. 696-702, 1997.

[6] J. C. Dagher et al, “Resource-constrained rate control for motion JPEG2000,” *IEEE Trans. Image Processing*, vol. 12, no. 12, Dec. 2003.

[7] L. Luo, et al, “Rate control with smoothed temporal distortion for a 3D embedded wavelet video coder,” *IEEE Conf. Information, Communications & Signal Processing*, Oct. 2001.

[8] A. Said and W. A. Pearlman, “A new, fast, and efficient image codec based on set partitioning in hierarchical trees,” *IEEE Trans. CSVT*, vol. 6, no. 3, pp. 243-250, 1996.