NEW SCALING COEFFICIENTS FOR BIORTHOGONAL FILTER TO CONTROL DISTORTION VARIATION IN 3D WAVELET BASED VIDEO CODING

Vidhya Seran, Lisimachos P. Kondi

332 Bonner Hall, Dept. of Electrical Engineering
State University of New York at Buffalo, Buffalo, NY 14260
Tel: (716) 645-2422 ext. 1147 Fax: (716) 645-3656
Email: { vseran,lkondi }@eng.buffalo.edu

ABSTRACT

In 3D based wavelet video coders that use longer filters, the PSNR of the reconstructed frames within a GOF varies significantly. This distortion variation is not directly related to the motion in the video sequence as in the case of hybrid video coding schemes. The paper focuses on the problem of controlling erratic variation of distortion in 3D coders. The wavelet filter properties are explored and scaling coefficients for the wavelet filter is calculated with the objective of smoothening the temporal PSNR. The biorthogonal 5/3 wavelet filter is considered in this paper and experimental results are presented for 2D+t and t+2D wavelet coders.

Index Terms— Wavelet Transforms, Video Coding

1. INTRODUCTION

Wavelet based coding has received considerable attention in the past few years and has the very best coding efficiency in image coders. Wavelet based coding has now become a powerful coding option for video both in traditional predictive feedback (closed loop) schemes and in three-dimensional (open loop) methods. The efficient closed loop wavelet video coders perform motion estimation/compensation in the Overcomplete Discrete Wavelet Transform (ODWT) domain. The open loop or the 3D methods treat video sequence as a three dimensional signal. The temporal redundancy is exploited using temporal wavelet filters. Unlike closed loop methods, 3D coders provide drift free SNR scalability. The multi-resolution nature of the wavelet coefficients provides temporal and resolution scalability.

Introducing motion estimation/compensation in 3D wavelet coders increases the coding efficiency when there is moderate or high motion in the video sequence. Lifting operation for temporal wavelet transform allows the incorporation of motion compensation with perfect reconstruction. The motion compensated three-dimensional wavelet analysis/synthesis can be implemented in two ways: two- dimensional spatial filtering followed by temporal filtering (2D+t) [1, 2, 3] or, temporal filtering followed by two-dimensional spatial filtering (t+2D)[4, 5].

3D wavelet-based video codecs that utilize longer temporal filters improve the coding efficiency. But they increase the delay in the coding system and introduces significant variation in the PSNR of the reconstructed frames within a GOF. This distortion fluctuation is unrelated to the motion in the sequence and can be in the order of 0.5-4 dB [2, 5]. This may lead to annoying flickering effects and poor visual quality. The fluctuation in the image quality across the group of frames (GOF) should be addressed while optimizing the 3D wavelet coder performance.

The temporal wavelet filter properties are known to be a major factor contributing to distortion fluctuation. The problem of controlling the temporal distortion fluctuations has been addressed in a few designs [6, 7]. In [6], distortion fluctuation control is designed for the bi-directional unconstrained motion compensated temporal filtering and the distortion in the decoded frame is expressed as a function of the distortions in the reference frames at the same temporal level. In [7], the relationship between the distortion in temporal wavelet subbands and the reconstructed frames are examined for the modified 5/3 filter (ignoring $\sqrt{2}$). Based on the relationship, a distortion ratio model is theoretically developed and a simple rate control algorithm is used to set priorities for the temporal subbands according to the distortion ratio.

Our work aims at exploring the MCTF filter properties and we present a complete analysis of the filter and mathematical derivations. In this work, the relationship between the distortion in the reconstructed frames and the filter coefficients is examined. On this basis, scaling coefficients for the filter are calculated to control the distortion fluctuation. We considered the most popular biorthogonal 5/3 filter in this paper. However, the method proposed here can be directly extended for other longer filters.

The rest of the paper is organized as follows: In Section 2, we examine the filter properties and in Section 3, the scaling coefficients for reducing distortion fluctuation is derived. Finally, in Section 4, we present the simulation results for different video sequences and in Section 5, we present our conclusions.

2. THREE DIMENSIONAL FILTER ANALYSIS

The distortion fluctuation in the temporal filters can be better understood by analyzing the filter properties. We selected the most popular biorthogonal 5/3 wavelet transform using lifting steps in this work. The analysis and synthesis equations are given below:

$$\begin{aligned} h_k &= [x_{2k+1} - \frac{1}{2}[x_{2k} + x_{2k+2}]]/\sqrt{2} \\ l_k &= \sqrt{2}[x_{2k} + \frac{1}{4}[\sqrt{2}h_k + \sqrt{2}h_{k-1}]]. \end{aligned} \quad (1)$$

$$\begin{aligned} x_{2k} &= \frac{l_k}{\sqrt{2}} - \frac{1}{4}[\sqrt{2}h_{k-1} + \sqrt{2}h_k] \\ x_{2k+1} &= \sqrt{2}h_k + \frac{1}{2}[x_{2k} + x_{2k+2}]. \end{aligned} \quad (2)$$

where l_k and h_k are the low-pass and highpass temporal subbands. x_{2k} and x_{2k+1} represent the even and odd frames respectively. Let $D_{x_{2k}}$ and $D_{x_{2k+1}}$ be the Mean Square Error (MSE) distortion corresponding to the even and odd frames. D_{l_k} and D_{h_k} be the MSE distortion of the low-pass and high-pass temporal subbands respectively. If we assume that all the temporal subbands are uncorrelated with zero mean [8], the distortion equations for even and odd frames in terms of the distortions of the low-pass and high-pass temporal subbands are given by:

$$\begin{aligned} D_{x_{2k}} &= \frac{D_{l_k}}{2} + \frac{D_{h_k}}{8} + \frac{D_{h_{k-1}}}{8} \\ D_{x_{2k+1}} &= \frac{1}{8}[D_{l_k} + D_{l_{k+1}}] + \frac{9}{8}D_{h_k} \\ &\quad + \frac{1}{32}[D_{h_{k+1}} + D_{h_{k-1}}]. \end{aligned} \quad (3)$$

We can also write the distortion equations for odd and even frames in terms of filter coefficients. Let SH_i be the low pass synthesis coefficients and SG_i be the high pass synthesis coefficients. Now, the distortion equations are:

$$\begin{aligned} D_{x_{2k}} &= D_l \sum_i SH_{2i}^2 + D_h \sum_j SG_{2j+1}^2 \\ D_{x_{2k+1}} &= D_l \sum_i SH_{2i+1}^2 + D_h \sum_j SG_{2j}^2. \end{aligned} \quad (4)$$

If we assume that the distortions in different temporal subbands are equal, i.e. $D_l = D_h = D$, the ratio of distortions of the even and odd frames is

$$\frac{D_{x_{2k}}}{D_{x_{2k+1}}} = \frac{\sum_i SH_{2i}^2 + \sum_j SG_{2j+1}^2}{\sum_i SH_{2i+1}^2 \sum_j SG_{2j}^2} \quad (5)$$

By substituting the filter coefficients in equation (5), the difference between odd and even frames will be 2.8182 dB. In [8], it is shown that when the $\frac{D_l}{D_h}$ is equal to equation (5), the average or the overall distortion can be minimized. Under this condition, the difference between odd and even frames will be 1.4 dB. When the number of temporal decomposition level increases, the distortion fluctuations become even more severe.

Let us consider a case where $\sqrt{2}$ is ignored in the analysis and synthesis equations. The synthesis equation (2) can be rewritten as

$$\begin{aligned} x_{2k} &= l_{2k} - \frac{1}{4}[h_{k-1} + h_k] \\ x_{2k+1} &= h_k + \frac{1}{2}[x_{2k} + x_{2k+1}]. \end{aligned} \quad (6)$$

The distortion equations (3) will become

$$\begin{aligned} D_{x_{2k}} &= D_{l_k} + \frac{D_{h_k}}{16} + \frac{D_{h_{k-1}}}{16} \\ D_{x_{2k+1}} &= \frac{1}{4}[D_{l_k} + D_{l_{k+1}}] + \frac{9}{16}D_{h_k} \\ &\quad + \frac{1}{64}[D_{h_{k+1}} + D_{h_{k-1}}]. \end{aligned} \quad (7)$$

Following the same steps as in the previous case, solving for the difference in PSNR between odd and even frames will result in 0.122 dB. The distortion ratio is given by

$$\frac{D_{x_{2k}}}{D_{x_{2k+1}}} = \frac{1.125}{1.09375} \quad (8)$$

The distortion fluctuation is reduced when $\sqrt{2}$ is omitted. However, the overall distortion is decreased considerably in this case. In [7], a distortion ratio model for temporal subbands is theoretically derived for the filters with no $\sqrt{2}$ and a simple rate control algorithm is used. The rates for different temporal subbands are selected based on the distortion ratio to reduce distortion fluctuation and to improve the average PSNR of the sequence.

So far, a brief overview of the wavelet filter properties and distortion variation between even and odd frames are discussed for 5/3 filter.

3. DISTORTION FLUCTUATION CONTROL

The distortion fluctuation exhibited by the 3D wavelet coders is due to the filter properties and implicit rate control operation. In order to control the temporal PSNR fluctuation, the rate control can be performed in a controlled manner or filter properties can be modified. In this section, we derive new scaling coefficients for the filter to eliminate distortion fluctuation. The new filter coefficients are designed with the objective of making the odd and even frame distortions equal. We considered a special case of making the odd and even frames equal at every temporal decomposition level.

3.1. Filter Coefficient Design

In order to control the fluctuation in the temporal direction, a scaling factor for the filter coefficients is derived. Let α_1 and β_1 be the scaling coefficients for SH_i and SG_i respectively. For a one level temporal decomposition, we solve for the ratio of α_1 and β_1 to arrive at $D_{x_{2k}} = D_{x_{2k+1}}$.

Then, from equation (5), we have

$$\alpha^2 \sum_i SH_{2i}^2 + \beta^2 \sum_j SG_{2j+1}^2 = \alpha^2 \sum_i SH_{2i+1}^2 + \beta^2 \sum_j SG_{2j}^2 \quad (9)$$

For a 5/3 filter, if we solve equation (9) for the relationship between α_1 and β_1, we get

$$\frac{\alpha_1}{\beta_1} = \sqrt{\frac{15}{4}}. \quad (10)$$

If we assume α_1 to be equal to 1, then β_1 will be equal to $\sqrt{\frac{4}{15}}$. By using these scaling coefficients for the synthesis high and low pass filters, the distortion for odd and even frames will be equal.

For a three level temporal decomposition, we find three sets of scaling coefficients such that the distortion for odd and even frame at every stage is equal. The third level reconstructed frame distortion for x_{2k} and x_{2k+1} is given by,

$$\begin{aligned} D_{x_{2k}} &= ([\alpha_1^2 \sum_i SH_{2i}^2 + \beta_1^2 \sum_j SG_{2j+1}^2] \quad (11) \\ &\quad *\alpha_2^2 \sum_i SH_{2i}^2 + \beta_2^2 \sum_j SG_{2j+1}^2) \\ &\quad *\alpha_3^2 \sum_i SH_{2i}^2 + \beta_3^2 \sum_j SG_{2j+1}^2 \\ D_{x_{2k+1}} &= \{[\alpha_1^2 \sum_i SH_{2i}^2 + \beta_1^2 \sum_j SG_{2j+1}^2] \quad (12) \\ &\quad *\alpha_2^2 \sum_i SH_{2i+1}^2 + \beta_2^2 \sum_j SG_{2j}^2\} \\ &\quad *\alpha_3^2 \sum_i SH_{2i+1}^2 + \beta_3^2 \sum_j SG_{2j}^2. \end{aligned}$$

endeqnarray The "$*$" used in the above equation represents convolution operation. The equations for the reconstructed frames are used to solve for α and β at various level, to eliminate quality variations. The relationship between α and β for a three level temporal decomposition is given below:

$$\frac{\alpha_3}{\beta_3} = 1.9365; \quad \frac{\alpha_2}{\beta_2} = 2.5725; \quad \frac{\alpha_1}{\beta_1} = 3.4173; \quad (13)$$

The derived values in equation (13), are used as scaling coeffiecients for the filter.

4. EXPERIMENTAL RESULTS

The results are presented for both types of motion compensated 3D wavelet coders (2D+t and t+2D methods). We considered the standard "Football" and "Flower Garden" test sequences in SIF (352×240) resolution for the 2D+t method and the "Foreman" and "Susie" test sequences in QCIF (176×144) resolution for the t+2D method. A Daube- chies $(9, 7)$ filter with a three level decomposition is used to compute the wavelet coefficients. The motion estimation is performed using the block matching technique for integer pixel accuracy for both the methods. A 16×16 wavelet block is matched in a search window of $[-16, 16]$ in the case for 2D+t method. 3D-SPIHT [9] is used to encode the wavelet coefficients after performing motion estimation/compensation. The scaling coefficients derived in section 3 are used. No explicit rate control is selected for all the cases discussed.

The PSNR of each reconstructed frame of the test sequences for the 5/3 filter are plotted in Figure 1 to Figure 4. The "Proposed Distortion Control" case in the figure uses the scaling coefficients for the 5/3 filter. The 'No Distortion Control' is the original 5/3 filter set coded using 3D-SPIHT. Table 1 gives the average PSNR values of the Y component for the three cases discussed. From the results, it can be seen that, with the distortion control scheme, the PSNR variation is greatly reduced. The average PSNR for the proposed case is slightly less than the original (no distortion control) case. The distortion controlled video will not have any flickering effects. The average PSNR for the proposed case can be further improved by including explicit rate control methods and it is not explored in this paper.

5. CONCLUSION

The wavelet filter properties are studied to understand the variation in distortion inside a group of frames. The distortion relationship for the reconstructed frame and temporal filter coefficients at various temporal levels are explored and a ratio for the scaling coefficients to control the fluctuation is derived. The modified 5/3 filter with the derived scaling coefficients reduces the distortion fluctuation. The proposed method can be applied to any filter to obtain the scaling coefficients to control distortion variation.

6. REFERENCES

[1] Y. Andreopoulos, A. Munteanu, J. Barbarien, M. van der Schaar, J. Cornelis, and P. Schelkens, "In-band motion compensated temporal filtering," *Signal Processing: Image Communication (special issue on "Subband/Wavelet Interframe Video Coding")*, vol. 19, pp. 653–673, August 2004.

[2] Y. Wang, S. Cui, and J. E. Fowler, "3D video cod-

Table 1. Average PSNR values of Y component

Sequence	Rate	Proposed Distortion Control	No Distortion Control	3D Method
Football	1.5 Mbps	30.44 dB	30.66 dB	2D+t
Garden	1.5 Mbps	29.67 dB	29.94 dB	2D+t
Susie	250 Kbps	40.12 dB	40.31 dB	t+2D
Foreman	250 Kbps	35.49 dB	35.85 dB	t+2D

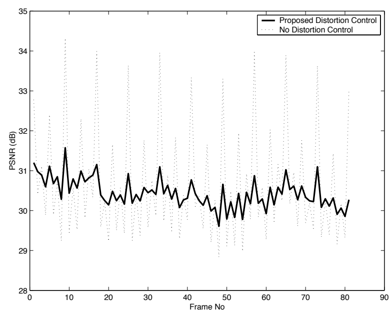

Fig. 1. Distortion Control for Football sequence

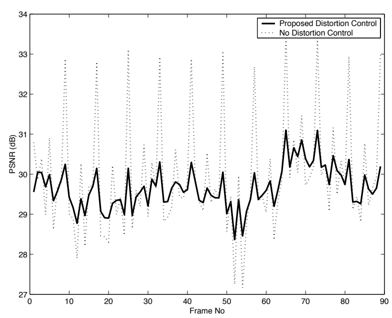

Fig. 2. Distortion Control for Garden sequence

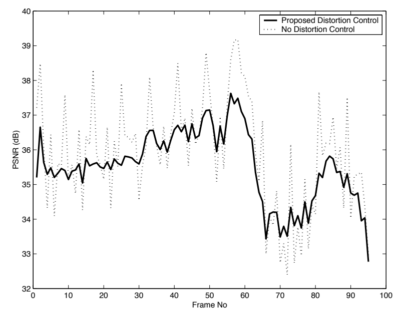

Fig. 3. Distortion Control for Foreman sequence

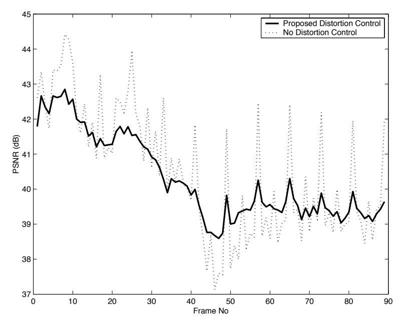

Fig. 4. Distortion Control for Susie sequence

ing using redundant-wavelet multihypothesis and motion-compensated temporal filtering," in *Proceedings of the ICIP*, Barcelona, Spain, 2003, vol. 2, pp. 755–758.

[3] Xin Li, "Scalable video compression via overcomplete motion compensated wavelet coding," *Signal Processing: Image Communication (special issue on "Subband/Wavelet Interframe Video Coding")*, vol. 19, pp. 637–651, August 2004.

[4] A. Secker and D. Taubman, "Lifting based invertible motion adaptive transform (LIMAT) framework for highly scalable video compression," *IEEE Transactions on Image Processing*, pp. 1530–1542, Dec 2003.

[5] A. Golwelkar and J. Woods, "Scalable video compression using longer motion compensated temporal filters," in *Proceedings of SPIE VCIP*, 2003, vol. 5150, pp. 1406–1416.

[6] A. Munteanu, Y. Andreopoulos, M. van der Schaar, and J. Cornelis, "Control of the distortion variation in video coding systems based on motion compensated temporal filtering," *Proceedings of the ICIP*, vol. 2, pp. 62–64, 2003.

[7] V. Seran and L. P. Kondi, "Distortion fluctuation control for 3D wavelet based video coding," *Proceedings of SPIE VCIP*, Jan. 2006.

[8] D. S. Taubman and M. W. Marcellin, *JPEG2000, Image compression fundamentals, standards and practice*, Kluwer academic Publishers Group, 2002.

[9] B. J. Kim, Z. Xiong, and W. A. Pearlman, "Low bit-rate scalable video coding with 3D set partitioning in hierarchical trees," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 10, pp. 1374–1387, Dec 2000.