

## Research Article

# Quality Variation Control for Three-Dimensional Wavelet-Based Video Coders

Vidhya Seran and Lisimachos P. Kondi

*Department of Electrical Engineering, State University of New York at Buffalo, 332 Bonner Hall, Buffalo, NY 14260, USA*

Received 15 August 2006; Revised 8 January 2007; Accepted 9 January 2007

Recommended by James E. Fowler

The fluctuation of quality in time is a problem that exists in motion-compensated-temporal-filtering (MCTF-) based video coding. The goal of this paper is to design a solution for overcoming the distortion fluctuation challenges faced by wavelet-based video coders. We propose a new technique for determining the number of bits to be allocated to each temporal subband in order to minimize the fluctuation in the quality of the reconstructed video. Also, the wavelet filter properties are explored to design suitable scaling coefficients 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 MCTF wavelet coders.

Copyright © 2007 V. Seran and L. P. Kondi. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## 1. INTRODUCTION

Research in image sequence compression or video coding is a natural extension of research in image compression/coding. Beyond the removal of spatial and spectral redundancy in response to our human visual system (HVS), video coding exploits further temporal correlation between consecutive frames. Owing to high similarity between adjacent frames, efficient video coding significantly relies on effective removal of temporal redundancy in the source video. Wavelet-based image coding has enabled not only good compression but also efficient scalability. Image compression algorithms like set partitioning in hierarchical trees (SPIHT) [1], embedded zerotree wavelet (EZW) [2], and JPEG2000 [3] are wavelet-based and they are known to outperform the discrete-cosine-transform- (DCT-) based compression techniques for image coding. As a result, recent research efforts on video coding were targeted on wavelet-based techniques.

With the increase in demand for video over the Internet, scalability has become an important issue. A conventional hybrid coder with closed-loop prediction is not a very efficient method for deriving a scalable codec. For an encoder to provide scalable bitstream, it must operate without any prior knowledge about the rate, resolution or temporal level at which the video sequence will be reconstructed. Hence the feedback structure that is present in the current hybrid

coders (which is optimal only for one particular rate) makes scalable compression inefficient. Hence a new method for exploiting temporal redundancy that eliminates the feedback loop is required. On the other hand, the 3D transforms provide a better way of deriving an efficient scalable codec because no such feedback loop is required. The coder operates on a current block of frames for temporal and spatial decomposition. Since the 3D system forms an open-loop system, the disadvantages associated with traditional hybrid coders can be avoided. The open-loop coding scheme is currently an ongoing research problem and wavelet-based coding has now become a powerful coding option for video in three-dimensional (open-loop) methods. The main theoretical development that promises efficient 3D wavelet-based video codecs with perfect invertibility is motion compensated temporal filtering (MCTF) using lifting. The MCTF using lifting can be performed in two ways.

- (1) Two-dimensional spatial filtering followed by temporal filtering (2D+t) [4–7].
- (2) Temporal filtering followed by two-dimensional spatial filtering (t+2D) [8–11].

All current wavelet-based video codecs that employ temporal filtering exhibit a fluctuation in the PSNR of the reconstructed frames within a group of frames (GOF). This is true for both t+2D and 2D+t schemes. The distortion fluctuation

is more pronounced with longer filters and is undesirable at low-bit rates. Most of the coders aim at optimizing the average PSNR, disregarding the fluctuation in the image quality across the GOF. The distortion fluctuation inside a GOF can be in the order of 0.5–4 dB. This may lead to annoying flickering effects and poor visual quality. It is well known that the average PSNR for the whole video sequence alone is not an adequate indicator of subjective video quality. Hence the fluctuation in the image quality across the GOF should be addressed while optimizing the 3D wavelet coder performance. The distortion fluctuation considered in this paper is due to the temporal filter characteristics and is present even if the temporal filtering is not motion compensated. For MCTF, the distortion fluctuation also depends on the motion model [12–15].

The problem of significant variation in the quality of the reconstructed video has been identified by few designs [16, 17] for the motion compensated temporal prediction case. In [16], a design for controlling the distortion variation is proposed for the unconstrained motion compensated temporal prediction [18]. The distortion of each decoded frame is expressed as a function of the distortions of the decoded reference frames at the same temporal level. A control parameter is set and by varying the control parameter, tradeoffs between the average PSNR within the GOF and the decoded PSNR fluctuation are achieved. In [17], the quality fluctuation control is treated as a quadratic programming problem based on the distortion analysis for MCTF-based video coding.

Our work aims at exploring the MCTF filter properties and we present a complete analysis of the filter and mathematical derivations. Based on the mathematical derivations and the experimental results, the solution for controlling the quality variation is achieved. The proposed methods are applicable to any motion model and can be directly extended to any temporal filter. The reduction in the average PSNR is also very small.

The temporal wavelet filter properties are known to be a major factor contributing to distortion fluctuation. The temporal distortion fluctuation is due to different filter synthesis gains for even and odd frames [19]. In this paper, we propose two novel methods to control the distortion fluctuation. In the first method, the relationship between the distortion in temporal wavelet subbands and the reconstructed frames is examined for the modified 5/3 filter (ignoring the factor  $\sqrt{2}$ ). Based on the relationship, a distortion ratio model is theoretically developed and a rate control algorithm is proposed to set priorities for the temporal subbands according to the distortion ratio. In the second method, based on the relationship between the distortion in the reconstructed frames and the filter coefficients, new scaling coefficients for the filter are calculated. We consider the popular biorthogonal 5/3 filter in our work. Some preliminary results of our work have appeared in [7, 20, 21].

The rest of the paper is organized as follows: in Section 2, we examine the filter properties and in Section 3, the two methods for controlling the distortion fluctuation are discussed. 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.

### 2.1. Biorthogonal 5/3 filter

The analysis and synthesis equations are given below:

$$\begin{aligned} h_k(x, y) &= \frac{[f_{2k+1}(x, y) - 1/2[f_{2k}(x, y) + f_{2k+2}(x, y)]]}{\sqrt{2}}, \\ l_k(x, y) &= \sqrt{2} \left[ f_{2k}(x, y) + \frac{1}{4} [\sqrt{2}h_k(x, y) + \sqrt{2}h_{k-1}(x, y)] \right], \\ f_{2k}(x, y) &= \frac{l_k(x, y)}{\sqrt{2}} - \frac{1}{4} [\sqrt{2}h_{k-1}(x, y) + \sqrt{2}h_k(x, y)], \\ f_{2k+1}(x, y) &= \sqrt{2}h_k(x, y) + \frac{1}{2} [f_{2k}(x, y) + f_{2k+2}(x, y)], \end{aligned} \quad (1)$$

where  $l_k$  and  $h_k$  are the low-pass and high-pass temporal subbands and  $f_{2k}$  and  $f_{2k+1}$  represent the even and odd frames, respectively. To make notation simpler, the motion mappings are not explicitly included in the filter equations, but the experimental results use the block-based motion model to calculate the temporal subbands. Let  $D_{f_{2k}}$  and  $D_{f_{2k+1}}$  be the mean square error (MSE) distortion corresponding to the even and odd frames.  $D_{l_k}$  and  $D_{h_k}$  are 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 [22], the distortion equations for the even and odd frames in terms of the distortions for the low-pass and high-pass temporal subbands are given by

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

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. Also, let us assume that the distortions of all low-pass temporal subbands are equal to  $D_l$  and the distortions of all high-pass temporal subbands are equal to  $D_h$ . Now, the distortion equations are

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

If we assume that the distortions in different temporal subbands are equal, that is,  $D_l = D_h = D$ , the ratio of distortions

for the even and odd frames is

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

By substituting the filter coefficients in (5), the difference between odd and even frames will be 2.8182 dB. In other words, the ratio of distortion in even and odd frames is

$$\frac{D_{f_{2k}}}{D_{f_{2k+1}}} = \frac{0.75}{1.4375}. \quad (6)$$

In [20], it is shown that when the ratio of temporal subbands,  $D_l/D_h$ , is made equal to 0.75/1.4375, the average distortion can be minimized for the considered group of frames. If we force the temporal subbands ratio to be equal to 0.75/1.4375 or make  $D_l = (0.75/1.4375)D_h$ , there will still be a difference of 1.4 dB between odd and even frames. This can be verified by substituting for  $D_l$  in (5) or (3). When the number of temporal decomposition levels increases, the distortion fluctuation becomes even more severe.

Let us consider a case where the factor  $\sqrt{2}$  is ignored in the analysis and synthesis equations. The analysis equations (1) can be rewritten as

$$\begin{aligned} h_k(x, y) &= f_{2k+1}(x, y) - \frac{1}{2}[f_{2k}(x, y) + f_{2k+2}(x, y)], \\ l_k(x, y) &= f_{2k}(x, y) + \frac{1}{4}[h_k(x, y) + h_{k-1}(x, y)]. \end{aligned} \quad (7)$$

Then, distortion equations (3) will become

$$\begin{aligned} D_{f_{2k}} &= D_{l_k} + \frac{D_{h_k}}{16} + \frac{D_{h_{k-1}}}{16}, \\ D_{f_{2k+1}} &= \frac{1}{4}[D_{l_k} + D_{l_{k+1}}] + \frac{9}{16}D_{h_k} + \frac{1}{64}[D_{h_{k+1}} + D_{h_{k-1}}]. \end{aligned} \quad (8)$$

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

$$\frac{D_{f_{2k}}}{D_{f_{2k+1}}} = \frac{1.125}{1.09375}. \quad (9)$$

The distortion fluctuation is reduced when the factor  $\sqrt{2}$  is omitted. However, the overall distortion is increased, thereby decreasing the average PSNR. However, this analysis provides an insight for the distortion fluctuation control problem.

## 2.2. Biorthogonal 5/3 filter without update step

Consider the analysis of the lifting steps discussed in (8). If the high-pass temporal subbands are not used for low-pass filtering [18], then the equations can be rewritten as

$$\begin{aligned} h_k(x, y) &= f_{2k+1}(x, y) - \frac{1}{2}[f_{2k}(x, y) + f_{2k+2}(x, y)], \\ l_k(x, y) &= f_{2k}(x, y). \end{aligned} \quad (10)$$

This filter is commonly referred to as 1/3 filter. When compared to the 5/3 filter, the distortion fluctuation is even

more pronounced in 1/3 filter. This is an effect of ignoring the update step. Though inclusion of an update step increases the encoding and decoding delay, the compression efficiency is higher. If we derive the temporal subband distortion relationship as in Section 2.1 for the 1/3 filter, the distortion ratio is

$$\frac{D_{f_{2k}}}{D_{f_{2k+1}}} = \frac{1.0}{1.5}. \quad (11)$$

If the ratio  $D_l/D_h$  is made equal to 1.0/1.5, the difference between odd and even frames will be 3 dB. Including the update step may reduce the quality variation to some extent, but it introduces additional delay [7]. Hence under delay constraints, we might opt for the 1/3 filter where the distortion variation is even more pronounced. Hence it is important to control the quality variation in both the 5/3 and the 1/3 filter.

So far, the wavelet filter properties were examined and the distortion variation between even and odd frames was studied for 5/3 and 1/3 filter. Assumptions made here will assist in understanding the relationship between temporal subbands.

## 3. DISTORTION FLUCTUATION CONTROL

### 3.1. Fluctuation reduction through rate control: the distortion ratio method

We propose a novel technique for assigning priorities to temporal subbands at different levels in order to control distortion fluctuation inside a GOF. The priorities for the temporal subbands can be set according to their distortion relationship. A new distortion ratio model is developed based on the distortion relationship, which will serve as a reference for the rate control algorithm.

#### 3.1.1. Distortion ratio model

In order to control the fluctuation in the temporal direction, the ratio  $D_l/D_h$  is derived. For a one-level temporal decomposition, we solve for the ratio  $D_l/D_h$  to arrive at  $D_{f_{2k}} = D_{f_{2k+1}}$ .

From (8), we have

$$D_l + \frac{1}{8}D_h = \frac{1}{2}D_l + \frac{19}{32}D_h, \quad (12)$$

then the ratio of  $D_l$  to  $D_h$  will be

$$\frac{D_l}{D_h} = \frac{15}{16}. \quad (13)$$

If the distortions of low- and high-pass temporal subbands are made to follow (13), the fluctuation will be reduced. For a three level temporal decomposition of the 5/3 filter, we get eight temporal subbands (one  $l^3$  and  $h^3$ , two  $h^2$ , and four  $h^1$ ). The distortion equations for eight reconstructed frames can be derived in terms of the distortions of the eight temporal subbands. For simplicity, let us assume  $D_h^1$  to be the distortion of the first-level temporal high-pass subbands  $h^1$  and  $D_h^2$  to be the distortion of  $h^2$ . Let  $D_l^3$  be the

third-level low-pass temporal subband distortion and let  $D_h^3$  be the temporal highpass distortion at third level.

The distortion of the frames inside a GOF can be denoted in terms of the distortion of the temporal subbands. For a modified 5/3 filter (no  $\sqrt{2}$  factor) with three-level temporal decomposition, the reconstructed frame distortions for frames  $f_{2k}$  to  $f_{2k+4}$  are given by

$$\begin{aligned} D_{f_{2k}} &= D_l^3 + 0.125D_h^3 + 0.125D_h^2 + 0.125D_h^1, \\ D_{f_{2k+1}} &= 0.78D_l^3 + 0.048D_h^3 + 0.102D_h^2 + 0.594D_h^1, \\ D_{f_{2k+2}} &= 0.625D_l^3 + 0.102D_h^3 + 0.594D_h^2 + 0.125D_h^1, \\ D_{f_{2k+3}} &= 0.5D_l^3 + 0.283D_h^3 + 0.289D_h^2 + 0.594D_h^1, \\ D_{f_{2k+4}} &= 0.5D_l^3 + 0.594D_h^3 + 0.125D_h^2 + 0.125D_h^1. \end{aligned} \quad (14)$$

The equations for the reconstructed frames are used to solve for the temporal subband distortion ratios in order to eliminate quality variations. The relationship between various temporal subbands for a three-level temporal decomposition is given below:

$$\frac{D_l^3}{D_h^3} = \frac{15}{16}, \quad \frac{D_h^3}{D_h^2} = \frac{15}{12}, \quad \frac{D_h^2}{D_h^1} = \frac{15}{12}. \quad (15)$$

Similarly, if we solve for the 1/3 filter set, we get the following ratio set:

$$\frac{D_l^3}{D_h^3} = 2, \quad \frac{D_h^3}{D_h^2} = 2, \quad \frac{D_h^2}{D_h^1} = 2. \quad (16)$$

The derived ratios in (15) are used to design the reference model for our rate control algorithm.

### 3.1.2. Rate allocation

The rate control problem for a video coder can be roughly stated as the determination of proper coding parameters so that the decoded video quality is optimized with respect to a certain fixed rate. For an embedded coder, the bit rate of each subband can be directly controlled to achieve the required distortion. Let  $N$  be the number of frames within a group of frames (GOF) and let  $R_N$  be the rate assigned to the GOF. The rate control problem can be formulated as: given the rate  $R_N$  for the GOF, we want to allocate the rate such that the overall distortion is minimized. For example, if we consider a three-level temporal decomposition and the GOF length  $N=8$ ,

$$\begin{aligned} R_l^3 + R_h^3 + R_{h_1}^2 + R_{h_2}^2 + R_{h_1}^1 + R_{h_2}^1 + R_{h_3}^1 + R_{h_4}^1 \\ = R_N \min (D_l^3 + D_h^3 + D_{h_1}^2 + D_{h_2}^2 + D_{h_1}^1 + D_{h_2}^1 + D_{h_3}^1 + D_{h_4}^1). \end{aligned} \quad (17)$$

The superscripts denote the level of decomposition and the subscripts denote subband type and number. In this work, a search algorithm described in Section 3.1.3 is used to select the rates, such that the distortion criterion is met. For the search algorithm, the temporal subband distortion has to be modeled first. We choose the exponential rate-distortion model [22, 23] for the temporal subband distortion. Then, the temporal subband distortion is given by

$$D_n = \sigma_n^2 2^{-\gamma_n R_n}, \quad (18)$$

where  $\sigma_n^2$  is the source variance and  $\gamma_n$  is the coding efficiency parameter. For each temporal subband  $n$ , the coding efficiency parameter  $\gamma_n$  and the variance  $\sigma_n^2$  have to be determined.

### 3.1.3. Rate control algorithm

The algorithm to choose the rate to minimize distortion fluctuation is given below.

- (1) For each wavelet temporal subband in the GOF calculate  $\sigma_n^2$ ,  $\gamma_n$ , and  $q$  R-D points.
- (2) Get the total rate  $R_N$  assigned for the GOF of size  $N$ .
- (3) Initially, let  $R_l^3 = c \cdot R_N/N$ , where  $c$  is a multiplication constant. The corresponding distortion  $D_l^3$  is found.
- (4) Using the distortion ratios for temporal subbands, select  $D_h^3$ ,  $D_h^2$ , and  $D_h^1$  from the  $q$  points and get the corresponding rates  $R_h^3$ ,  $R_h^2$ , and  $R_h^1$ .
- (5) Check if the sum of the rates of temporal subbands is equal to  $R_N$ ; if equal, then go to next GOF.
- (6) If the sum is greater than  $R_N$ , decrease the value for  $c$ . Else, increase  $c$  and go to Step (3).

The accuracy of the assumed exponential model for temporal subband is very important to get optimal rates.

## 3.2. Fluctuation reduction through scaling of transform coefficients: the filter coefficient method

In order to control the temporal PSNR fluctuation, the rate control can be performed in a controlled manner or the filter properties could 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 consider a special case of making the odd and even frames equal at every temporal decomposition level. Hence at any temporal level, the distortion fluctuation is minimized.

Let  $\alpha_1$  and  $\beta_1$  be the scaling coefficients for  $SH_i$  and  $SG_j$ , respectively. For a one-level temporal decomposition, we solve for the ratio of  $\alpha_1$  and  $\beta_1$  to arrive at  $D_{f_{2k}} = D_{f_{2k+1}}$ .

Then, from (5), we have

$$\alpha_1^2 \sum_i SH_{2i}^2 + \beta_1^2 \sum_j SG_{2j+1}^2 = \alpha_1^2 \sum_i SH_{2i+1}^2 + \beta_1^2 \sum_j SG_{2j}^2. \quad (19)$$

For a 5/3 filter, if we solve (19) for the relationship between  $\alpha_1$  and  $\beta_1$ , we get

$$\frac{\alpha_1}{\beta_1} = \sqrt{\frac{15}{4}}. \quad (20)$$

If we assume  $\alpha_1$  to be equal to 1, then  $\beta_1$  will be equal to  $\sqrt{4/15}$ . 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 distortions for odd

and even frames at every stage are equal. The third-level reconstructed frame distortion for frames  $f_{2k}$  and  $f_{2k+1}$  is given by

$$\begin{aligned}
 D_{f_{2k}} &= \left( \left[ \alpha_1^2 \sum_i SH_{2i}^2 + \beta_1^2 \sum_j SG_{2j+1}^2 \right] \right. \\
 &\quad * \alpha_2^2 \sum_i SH_{2i}^2 + \beta_2^2 \sum_j SG_{2j+1}^2 \left. \right) \\
 &\quad * \alpha_3^2 \sum_i SH_{2i}^2 + \beta_3^2 \sum_j SG_{2j+1}^2, \\
 D_{f_{2k+1}} &= \left\{ \left[ \alpha_1^2 \sum_i SH_{2i}^2 + \beta_1^2 \sum_j SG_{2j+1}^2 \right] \right. \\
 &\quad * \alpha_2^2 \sum_i SH_{2i+1}^2 + \beta_2^2 \sum_j SG_{2j}^2 \left. \right\} \\
 &\quad * \alpha_3^2 \sum_i SH_{2i+1}^2 + \beta_3^2 \sum_j SG_{2j}^2.
 \end{aligned} \tag{21}$$

The “\*” used in the above equations represents convolution operation. The equations for the reconstructed frames are used to solve for  $\alpha$  and  $\beta$  at various level, to eliminate quality variations. The relationship between  $\alpha$  and  $\beta$  for a three-level temporal decomposition at various levels is given below:

$$\frac{\alpha_3}{\beta_3} = 1.9365, \quad \frac{\alpha_2}{\beta_2} = 2.5725, \quad \frac{\alpha_1}{\beta_1} = 3.4173. \tag{22}$$

The derived values in (22) are used as scaling coefficients for the filter.

#### 4. EXPERIMENTAL RESULTS

We implemented the two types of wavelet-based video codecs described and the results are presented for both types of motion compensated 3D wavelet coders (2D+t and t+2D methods). A Daubechies (9,7) filter with a three-level spatial decomposition is used to compute the wavelet coefficients in all the cases considered. The motion estimation is performed using the block matching technique for integer pixel accuracy for both methods. The wavelet block matching technique in the overcomplete transform domain [24] is used in 2D+t schemes and spatial block method is used in t+2D schemes. A  $16 \times 16$  wavelet block is matched in a search window of  $[-16, 16]$  in the case for 2D+t method.

We considered the standard “Football” and “Flower Garden” test sequences in SIF ( $352 \times 240$ ) resolution for the 2D+t method and the “Foreman” and “Susie” test sequences in QCIF ( $176 \times 144$ ) resolution for t+2D method.

##### 4.1. Distortion ratio method

The SPIHT image coder was used to encode each temporal subband independently so that we could easily select the number of bits to match the distortion ratio derived in Section 3.1. The algorithm described in Section 3.1.2 is used for the rate selection. Since it is very difficult to exactly achieve the distortions to follow, the derived ratios from  $q$  points, a room for 2% error in distortion was allowed.

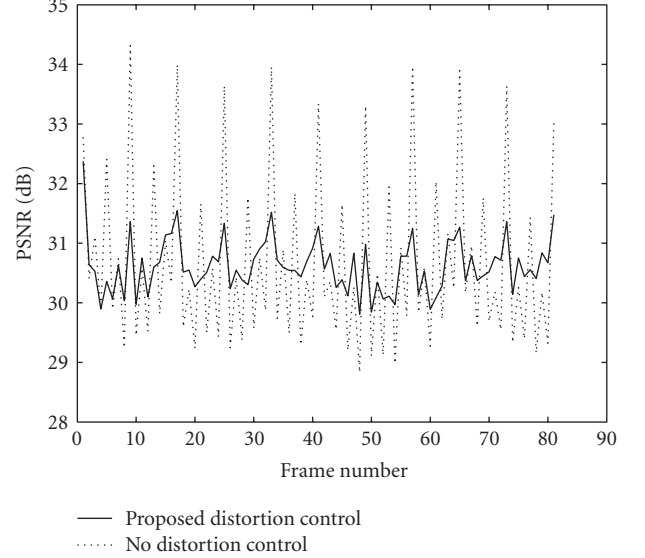


FIGURE 1: Football sequence: distortion control for 5/3 filter using ratio method.

TABLE 1: Distortion ratio method: average PSNR values of Y component.

| Sequence | Rate     | Proposed distortion control | No distortion control | No root 2 |
|----------|----------|-----------------------------|-----------------------|-----------|
| Football | 1.5 Mbps | 30.62 dB                    | 30.66 dB              | 29.82 dB  |
| Garden   | 1.2 Mbps | 29.72 dB                    | 29.74 dB              | 28.97 dB  |
| Susie    | 220 Kbps | 40.77 dB                    | 40.63 dB              | 40.01 dB  |
| Foreman  | 228 Kbps | 35.65 dB                    | 35.57 dB              | 34.96 dB  |

The PSNRs of each reconstructed frame of test sequences for the 5/3 filter are plotted in Figures 1–4. The 1/3 filter case for “Football” sequence is plotted in Figure 5 at 1.4 Mbps.

The “proposed distortion control” case in the figures follows the rate control algorithm. The “No root 2” case is coded using 3D-SPIHT and no explicit rate control is used. Both the cases use the modified 5/3 filter set without including the factor  $\sqrt{2}$ . The “No distortion control” is the 5/3 filter set coded using 3D-SPIHT [25]. Table 1 gives the average PSNR values of the Y component for the three cases discussed. From the results, it can be seen, with the distortion control scheme, the PSNR variation is greatly reduced and the average PSNR is also close to the implicit rate allocation “No distortion control” case.

##### 4.2. Filter coefficient method

3D-SPIHT [25] is used to encode the wavelet coefficients after performing motion estimation/compensation. The scaling coefficients derived in Section 3.2 are used. No explicit rate control is selected for all the cases discussed.

The peak signal-to-noise ratios of each reconstructed frame of the test sequences for the 5/3 filter are plotted in



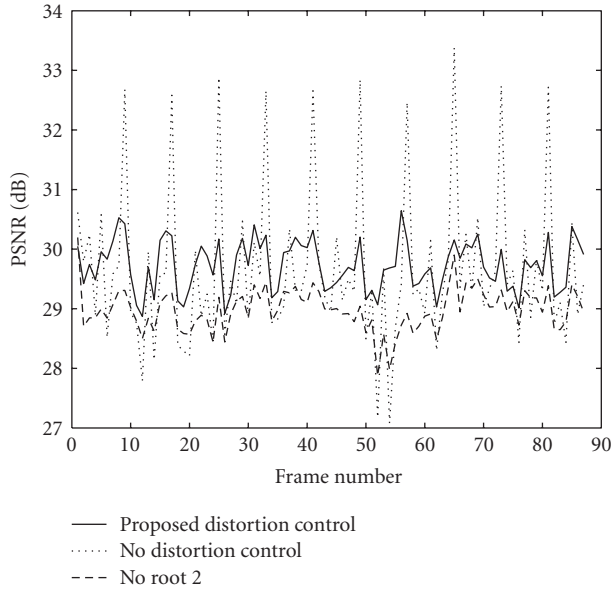


FIGURE 2: Garden sequence: distortion control for 5/3 filter using ratio method.

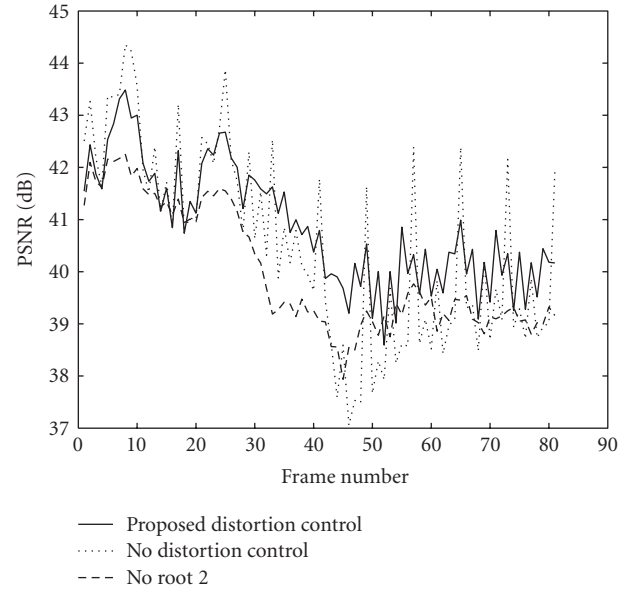


FIGURE 4: Susie sequence: distortion control for 5/3 filter using ratio method.

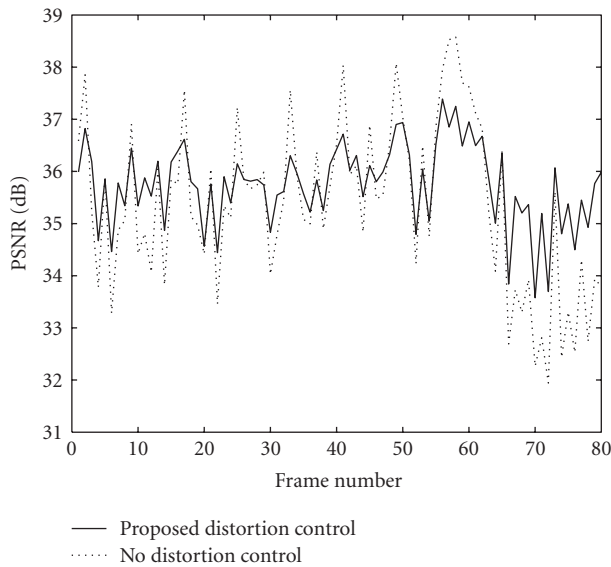


FIGURE 3: Foreman sequence: distortion control for 5/3 filter using ratio method.

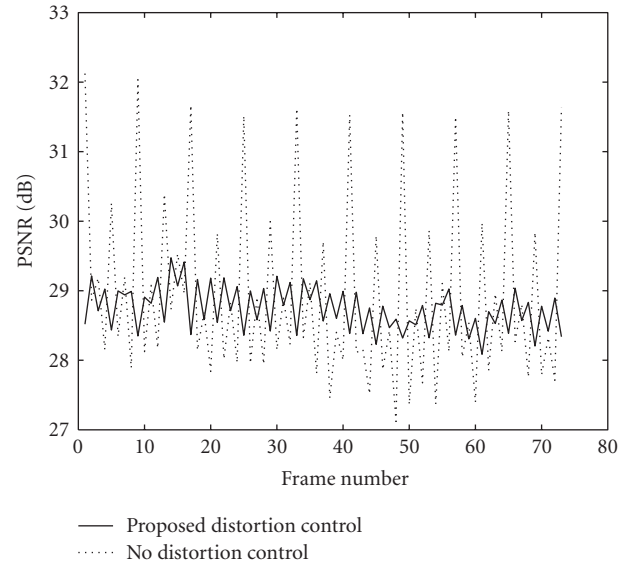


FIGURE 5: Football sequence: distortion fluctuation control for 1/3 filter using ratio method.

Figures 6–9. 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 2 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 but the distortion controlled video will not have any flickering effects. The ratio method performs

better in terms of average PSNR than the filter coefficient case, but the computation cost involved in the search algorithm is high.

## 5. CONCLUSION

The wavelet filter properties are studied to understand the variation in distortion of image quality inside a group of frames. The modified 5/3 filter without including the factor  $\sqrt{2}$  reduces distortion fluctuation at the cost of reducing

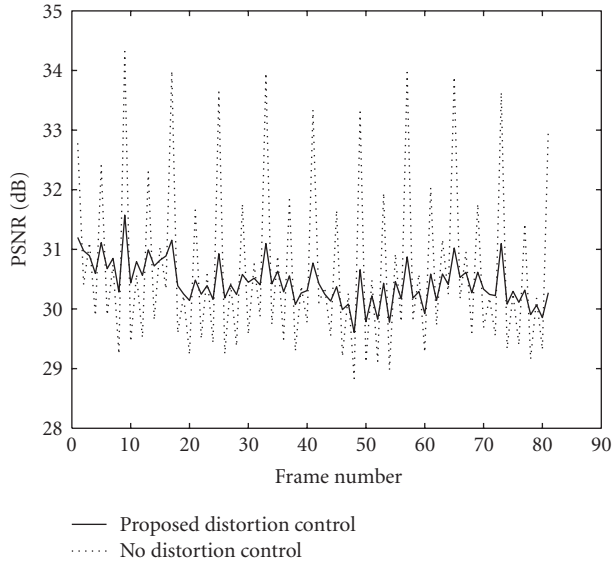


FIGURE 6: Football sequence: distortion control using filter coefficient method for 5/3 filter.

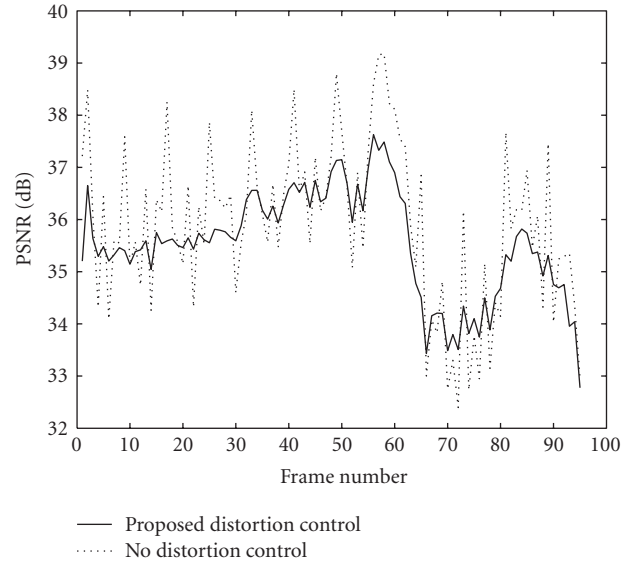


FIGURE 8: Foreman sequence: distortion control using filter coefficient method for 5/3 filter.

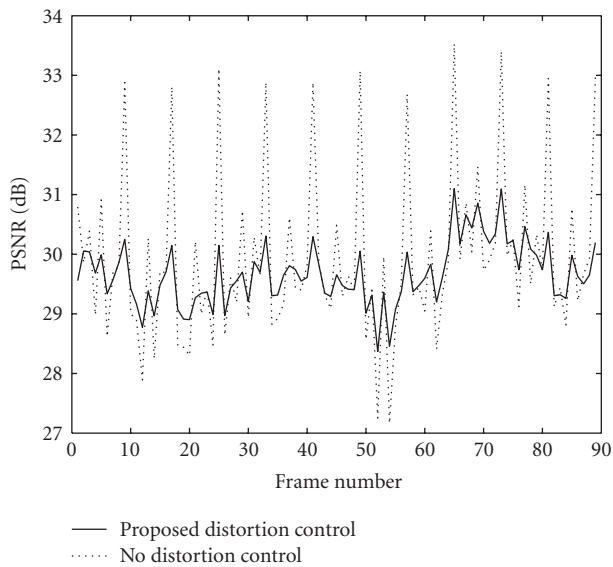


FIGURE 7: Garden sequence: distortion control using filter coefficient method for 5/3 filter.

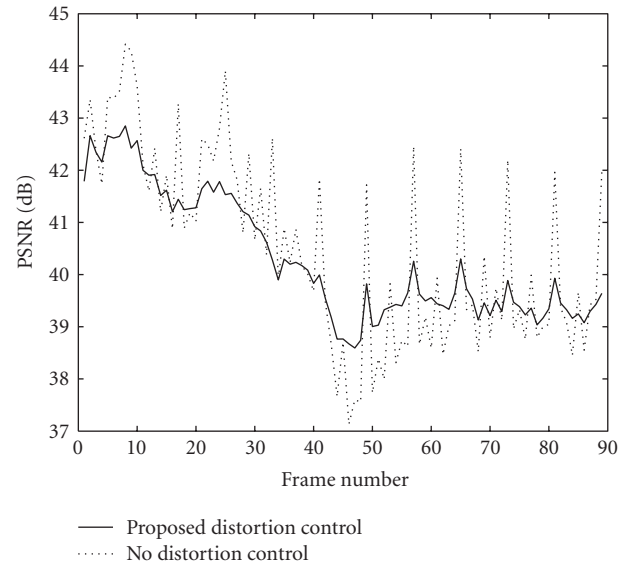


FIGURE 9: Susie sequence: distortion control using filter coefficient method for 5/3 filter.

TABLE 2: Filter coefficient method: 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   | 1.2 Mbps | 29.67 dB                    | 29.74 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      |

the overall PSNR. The distortion relationship of the temporal subbands at various temporal levels are explored and a ratio for controlling the fluctuation is derived. A rate control algorithm is used to control the quality variation. Also, 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 methods can be applied to any filter to obtain the scaling coefficients to control distortion variation. The distortion ratio method gives a better average PSNR for the considered sequences

compared to the filter coefficient method at the expense of a higher computational complexity. Our experimental results show that the reduction in the average PSNR is very small.

## REFERENCES

- [1] A. Said and W. A. Pearlman, "A new, fast, and efficient image codec based on set partitioning in hierarchical trees," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 6, no. 3, pp. 243–250, 1996.
- [2] J. M. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3445–3462, 1993.
- [3] C. Christopoulos, A. Skodras, and T. Ebrahimi, "The JPEG-2000 still image coding system: an overview," *IEEE Transactions on Consumer Electronics*, vol. 46, no. 4, pp. 1103–1127, 2000.
- [4] 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*, vol. 19, no. 7, pp. 653–673, 2004.
- [5] Y. Wang, S. Cui, and J. E. Fowler, "3D video coding using redundant-wavelet multihypothesis and motion-compensated temporal filtering," in *Proceedings of IEEE International Conference on Image Processing (ICIP '03)*, vol. 2, pp. 755–758, Barcelona, Spain, September 2003.
- [6] X. Li, "Scalable video compression via overcomplete motion compensated wavelet coding," *Signal Processing: Image Communication*, vol. 19, no. 7, pp. 637–651, 2004.
- [7] V. Seran and L. P. Kondi, "3D based video coding in the overcomplete discrete wavelet transform domain with reduced delay requirements," in *Proceedings of IEEE International Conference on Image Processing (ICIP '05)*, vol. 3, pp. 233–236, Genova, Italy, September 2005.
- [8] A. Secker and D. Taubman, "Lifting-based invertible motion adaptive transform (LIMAT) framework for highly scalable video compression," *IEEE Transactions on Image Processing*, vol. 12, no. 12, pp. 1530–1542, 2003.
- [9] S. T. Hsiang and J. W. Woods, "Embedded video coding using motion compensated 3-D subband/wavelet filter bank," in *Proceedings of the Packet Video Workshop*, Sardinia, Italy, May 2000.
- [10] A. Golwelkar and J. W. Woods, "Scalable video compression using longer motion compensated temporal filters," in *Visual Communications and Image Processing*, vol. 5150 of *Proceedings of SPIE*, pp. 1406–1416, Lugano, Switzerland, July 2003.
- [11] G. Pau, C. Tillier, B. Pesquet-Popescu, and H. Heijmans, "Motion compensation and scalability in lifting-based video coding," *Signal Processing: Image Communication*, vol. 19, no. 7, pp. 577–600, 2004.
- [12] K. Hanke, J.-R. Ohm, and T. Ruser, "Adaptation of filters and quantization in spatio-temporal wavelet coding with motion compensation," in *Proceedings of the IEEE International Picture Coding Symposium (PCS '03)*, pp. 49–54, Saint Malo, France, April 2003.
- [13] C.-L. Chang, A. Mavlankar, and B. Girod, "Analysis on quantization error propagation for motion-compensated lifted wavelet video coding," in *Proceedings of the 7th IEEE International Workshop on Multimedia Signal Processing (MMSP '05)*, Shanghai, China, October–November 2005.
- [14] A. Mavlankar and E. Steinbach, "Distortion prediction for motion-compensated lifted Haar wavelet transform and its application to rate allocation," in *Proceedings of the IEEE International Picture Coding Symposium (PCS '04)*, pp. 533–538, San Francisco, Calif, USA, December 2004.
- [15] 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," in *Proceedings of the 7th IEEE International Workshop on Multimedia Signal Processing (MMSP '05)*, Shanghai, China, October–November 2005.
- [16] A. Munteanu, Y. Andreopoulos, M. van der Schaar, P. Schelkens, and J. Cornelis, "Control of the distortion variation in video coding systems based on motion compensated temporal filtering," in *Proceedings of IEEE International Conference on Image Processing (ICIP '03)*, vol. 2, pp. 61–64, Barcelona, Spain, September 2003.
- [17] Y. Chen, J. Xu, F. Wu, and H. Xiong, "Quality-fluctuation-constrained rate allocation for MCTF-based video coding," in *Visual Communications and Image Processing*, vol. 6077 of *Proceedings of SPIE*, San Jose, Calif, USA, January 2006.
- [18] M. van der Schaar and D. S. Turaga, "Unconstrained motion compensated temporal filtering (UMCTF) framework for wavelet video coding," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '03)*, vol. 3, pp. 81–84, Hong Kong, April 2003.
- [19] N. Mehrseresht and D. Taubman, "An efficient content-adaptive MC 3D-DWT with enhanced spatial and temporal scalability," in *Proceedings of the IEEE International Conference on Image Processing (ICIP '04)*, vol. 2, pp. 1329–1332, Singapore, October 2004.
- [20] V. Seran and L. P. Kondi, "Distortion fluctuation control for 3D wavelet based video coding," in *Visual Communications and Image Processing*, vol. 6077 of *Proceedings of SPIE*, San Jose, Calif, USA, January 2006.
- [21] V. Seran and L. P. Kondi, "New scaling coefficients for biorthogonal filter to control distortion variation in 3D wavelet based video coding," in *Proceedings of the IEEE International Conference on Image Processing (ICIP '06)*, Atlanta, Ga, USA, October 2006.
- [22] D. S. Taubman and M. W. Marcellin, *JPEG2000, Image Compression Fundamentals, Standards and Practice*, Kluwer Academic, Boston, Mass, USA, 2002.
- [23] P.-Y. Cheng, J. Li, and C.-C. J. Kuo, "Rate control for an embedded wavelet video coder," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 7, no. 4, pp. 696–702, 1997.
- [24] H.-W. Park and H.-S. Kim, "Motion estimation using low-band-shift method for wavelet-based moving-picture coding," *IEEE Transactions on Image Processing*, vol. 9, no. 4, pp. 577–587, 2000.
- [25] 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 Transactions on Circuits and Systems for Video Technology*, vol. 10, no. 8, pp. 1374–1387, 2000.