[7]

Distortion Fluctuation Control for 3D Wavelet Based Video Coding

Vidhya Seran[a] and Lisimachos P. Kondi[a]

[a] 332 Bonner Hall, Dept. of Electrical Engineering,
State University of New York at Buffalo,
Buffalo, NY 14260

ABSTRACT

In this work, we propose a method to control distortion fluctuation in 3D wavelet based video coding. A novel technique of assigning priorities to temporal subbands at different levels are used to smoothen the temporal PSNR. The wavelet filter properties are explored and an optimum criterion is derived to reduce the distortion fluctuation.

Keywords: 3D Wavelet Video, Distortion fluctuation control

1. INTRODUCTION

As an alternative approach to traditional predictive coding for video, wavelet based coding has emerged as a viable option. In 3D based wavelet coding, a video sequence is treated as a three dimensional signal. A three dimensional wavelet transform is applied to the video source and the temporal redundancy is exploited using temporal filtering. In the presence of motion, the temporal correlation moves away from the temporal axis. In the popular 3D-SPIHT implementation,[1] a three dimensional wavelet transform is applied to a group of frames and the inter-band coefficients are arranged as 3D spatial orientation trees. In addition to providing high coding efficiency, 3D coding schemes has good spatial, temporal and drift free scalability. There are two main theoretical developments that promise efficient wavelet-based video codecs: Temporal filtering using lifting and motion compensation in the Overcomplete Discrete Wavelet Transform (ODWT). The three-dimensional wavelet decomposition can be performed in two ways:

1. Two-dimensional spatial filtering followed by temporal filtering (2D+t)[2–4] or,
2. Temporal filtering followed by two-dimensional spatial filtering (t+2D).[5–8]

All current wavelet-based video codecs that employ temporal filtering exhibit a significant fluctuation in the PSNR of the reconstructed frames within a GOF. This distortion variation within a GOF is not directly related to the motion in the sequence as in the case of hybrid video coding schemes. 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 bitrates.[3,5] Most of the coders aim at optimizing the average PSNR, disregarding the fluctuation in the image quality across the group of frames (GOF). The distortion fluctuation inside a GOF can be in the order of 0.5-4 dB.[3,8] 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 and the PSNR fluctuation should be taken into account. Hence, the PSNR variation inside a GOF should be controlled.

The temporal wavelet filter properties play a vital role in distortion fluctuation. The temporal distortion fluctuation is due to different filter synthesis gains for even and odd frames.[9] The problem of reducing the temporal distortion fluctuations has been addressed in a few designs.[8, 10–12] Filter based scaling coefficients are used to reduce the quality variation in Ref. 8. Though the scaling coefficients reduce the distortion variation this does not completely eliminate the fluctuation. In Ref. 10 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 this work, the

Visual Communications and Image Processing 2006, edited by John G. Apostolopoulos, Amir Said,
Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 6077, 607721, © 2006 SPIE-IS&T · 0277-786X/06/$15

relationship between the distortion in temporal wavelet subbands and the reconstructed frames are examined. Based on this relationship, an optimal condition for controlling distortion fluctuation is derived.

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

2. THREE DIMENSIONAL FILTER ANALYSIS

The distortion fluctuation in the temporal filters can be better understood by exploring the filter properties.

Let us consider an example of biorthogonal 5/3 wavelet transform using lifting steps. The analysis and synthesis equation are given below:

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

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

where l_k and h_k are the low-pass and highpass temporal subbands. x_{2k} and x_{2k+1} represents the even and odd frames respectively.

Let $D_{x_{2k}}$ and $D_{x_{2k+1}}$ be the MSE distortion corresponding to the even and odd frames. D_{l_k} and D_{h_k} are the MSE distortion of low-pass and high-pass temporal subbands respectively. If we assume that all the temporal subbands are uncorrelated,[13] then the distortion equations for even and odd frames in terms of distortions of low-pass and high-pass temporal subbands are given by:

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

If we assume that the distortions in different temporal subbands are equal, i.e. $D_{l_k} = D_{l_{k+1}} = D_l$, $D_{h_k} = D_{h_{k-1}} = D_h$ and $D_l = D_h = D$, then the differences between odd and even frames will be 2.8182 dB. In other words, the ratio of distortions of even and odd frames is

$$\frac{D_{x_{2k}}}{D_{x_{2k+1}}} = \frac{0.75}{1.4375}. \tag{4}$$

When the distortion in each subband is inversely proportional to the synthesis gain factor for that subband, the overall distortion is minimized.[9,13] This can be achieved when the ratio $\frac{D_l}{D_h}$ is made equal to equation (4). Even if we force the ratios to be equal to $\frac{0.75}{1.4375}$, there will still be a difference of 1.4 dB between odd and even frames.

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

$$\begin{aligned} x_{2k} &= l_{2k} - \frac{1}{4}[h_{k-1} + h_k] \\ x_{2k+1} &= h_k + \frac{1}{2}[x_{2k} + x_{2k+1}]. \end{aligned} \tag{5}$$

The distortion equation (3) will be

$$
\begin{aligned}
D_{x_{2k}} &= D_{l_k} + \frac{D_{h_k}}{16} + \frac{D_{h_{k-1}}}{16} \\
D_{x_{2k+1}} &= \frac{1}{4}[D_{l_k} + D_{l_{k+1}}] + \frac{9}{16}D_{h_k} + \frac{1}{64}[D_{h_{k+1}} + D_{h_{k-1}}].
\end{aligned}
\quad (6
$$

Under the condition that all D_l and D_h are equal, the difference in PSNR between odd and even fram drops to 0.122 dB. The distortion ratio is then given by

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

Though the fluctuation of PSNR is controlled by ignoring $\sqrt{2}$, it must be noted that the average PSNR is reduc considerably.

2.1. Biorthogonal 5/3 Filter Without Update Step

Consider the analysis of the lifting steps discussed in equation (1). If the high pass temporal subbands are r used for lowpass filtering,[14] then the equations can be rewritten as

$$
\begin{aligned}
h_k &= x_{2k+1} - \frac{1}{2}[x_{2k} + x_{2k+2}] \\
l_k &= x_{2k}.
\end{aligned}
$$

This filter is commonly referred as 1/3 filter. When compared to the 5/3 filter, the distortion fluctuatio even more pronounced in the 1/3 filter case. This is an effect of ignoring the update step. Though inclusio an update step increases the encoding and decoding delay, the compression efficiency will be high. If we de the temporal subband distortion relationship as in Section 2 for 1/3 filter, the distortion ratios will be

$$\frac{D_{x_{2k}}}{D_{x_{2k+1}}} = \frac{1.0}{1.5}.$$

If the ratio $\frac{D_l}{D_h}$ is made equal to equation (9), the difference between odd and even frames will be 3 dB.

In this section, the wavelet filter properties were examined and distortion variation between even and frames are studied for 5/3 and 1/3 filter. Assumptions made here will assist in the understanding of relationship between temporal subbands.

3. DISTORTION FLUCTUATION CONTROL

The distortion fluctuation exhibited by the 3D wavelet coders can be attributed to the implicit rate co operation. In order to control the temporal PSNR fluctuation, the rate allocation must be performed controlled manner.

A novel technique of assigning priorities for temporal subbands at different levels in order to control disto fluctuation inside a GOF is proposed. The priorities for the temporal subbands can be set according to distortion relationship. A distortion ratio model will serve as a reference for the rate control algorithm.

Frame No	0	1	2	3	4	5	6	7
Level 1	L	H	L	H	L	H	L	H
Level 2	LL		LH		LL		LH	
Level 3	LLL				LLH			

Figure 1. 5/3 Filter

3.1. Distortion Ratio Model

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

From equation (6), we have

$$D_l + \frac{1}{8}D_h = \frac{1}{2}D_l + \frac{19}{32}D_h. \tag{10}$$

Then, the ratio D_l to D_h of will be

$$\frac{D_l}{D_h} = \frac{15}{16}. \tag{11}$$

If the distortion of low and high pass temporal subbands are made to follow equation (11), the fluctuation will be reduced. A three level temporal decomposition for 5/3 filter is shown is Figure 1 and 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 for the first level temporal highpass subbands h^1 and D_h^2 be the distortion for h^2. Let D_l^3 be the third level lowpass temporal subband distortion and D_h^3 be the temporal highpass distortion at the third level.

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

$$\begin{aligned}
D_{x_{2k}} &= D_l^3 + 0.125D_h^3 + 0.125D_h^2 + 0.125D_h^1 \\
D_{x_{2k+1}} &= 0.78D_l^3 + 0.048D_h^3 + 0.102D_h^2 + 0.594D_h^1 \\
D_{x_{2k+2}} &= 0.625D_l^3 + 0.102D_h^3 + 0.594D_h^2 + 0.125D_h^1 \\
D_{x_{2k+3}} &= 0.5D_l^3 + 0.283D_h^3 + 0.289D_h^2 + 0.594D_h^1 \\
D_{x_{2k+4}} &= 0.5D_l^3 + 0.594D_h^3 + 0.125D_h^2 + 0.125D_h^1.
\end{aligned} \tag{12}$$

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

$$\frac{D_l^3}{D_h^3} = \frac{15}{16}; \quad \frac{D_h^3}{D_h^2} = \frac{15}{12}; \quad \frac{D_h^2}{D_h^1} = \frac{15}{12}. \tag{13}$$

Similarly, for the 1/3 filter set, the ratios are

$$\frac{D_l^3}{D_h^3} = 2; \quad \frac{D_h^3}{D_h^2} = 2; \quad \frac{D_h^2}{D_h^1} = 2. \tag{14}$$

The derived ratios in equation (13), are used to design the reference model for our rate control algorithm.

3.2. Rate Control Algorithm

The coding bitrate R can be directly controlled to achieve the required distortion D in a embedded wavelet coder. Let N be the number of frames within a group of frames (GOF) and R_N be the rate assigned for the GOF. The rate control problem is formulated as follows: Select the rates for each temporal subband such that the sum of all rates in the GOF is equal to R_N and the temporal subband distortions follow the ratios derived in section 3.1. For example if we consider a three level temporal decomposition for the 5/3 filter with GOF length N=8, the constraints are

$$\begin{aligned} R_l^3 + R_h^3 + R_{h_1}^2 + R_{h_2}^2 + R_{h_1}^1 + R_{h_2}^1 + R_{h_3}^1 + R_{h_3}^1 &= R_N \\ \text{and } \{ \frac{D_l^3}{D_h^3} = \frac{15}{16}; \quad \frac{D_h^3}{D_h^2} = \frac{15}{12}; \quad \frac{D_h^2}{D_h^1} &= \frac{15}{12} \}. \end{aligned} \tag{15}$$

The empirical formula for rate distortion for wavelet embedded coder[15] and the temporal subband distortion relationship can be combined to solve the optimal rate control problem. In this paper, a simple search algorithm is used to decide the rates to meet the distortion criteria. The algorithm to choose the rate to minimize distortion fluctuation is given below:

1. For each wavelet temporal subband in the GOF calculate q R-D points.
2. The total rate R_N assigned for the GOF of size N is selected.
3. Initially, let $R_l^3 = c.R_N/N$, where c is a multiplication constant. The corresponding distortion D_l^3 is chosen.
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 and if true Goto Step 7
6. If the sum is greater than R_N, decrease the value for c. Else increase c.
7. Goto step 3. For the next GOF use c as the initial value.

4. EXPERIMENTAL RESULTS

We considered the standard "Football", "Flower Garden" test sequences in SIF (352×240) resolution and "Foreman", "Susie" sequence in QCIF (176×144) resolution. A Daubechies (9,7) filter with a three level decomposition is used to compute the wavelet coefficients. The motion estimation is performed in the overcomplete wavelet domain using the block matching technique for integer pixel accuracy. A 16×16 wavelet block is matched in a search window of $[-16, 16]$. 3D-SPIHT and 2D-SPIHT[16] are used to encode the coefficients after performing motion estimation/compensation in ODWT domain.

The 2D-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. The algorithm described in Section 3.2 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.

The PSNR of each reconstructed frame of test sequences for 5/3 filter are plotted in Figures 2-5 . The 1/3 filter case for "Football" sequence is plotted in Figure 6 at 1.4 Mbps. The "Proposed Distortion Control" Case

Table 1. 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

in the figure 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 $\sqrt{2}$. The "No Distortion Control" is the 5/3 filter set coded using 3D-SPIHT. Table 1 gives the average PSNR values of the Y component for the three cases discussed. Form 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.

5. CONCLUSION

The wavelet filter properties are studied to understand the variation in distortion inside a group of frames. The modified 5/3 filter without including $\sqrt{2}$ eliminates distortion fluctuation at the cost of reducing the overall PSNR. The distortion relationships of temporal subbands at various temporal levels are explored and a ratio for controlling fluctuation is derived. A simple rate control algorithm is used to control the quality variation. More sophisticated models can be designed to achieve further improvement.

REFERENCES

1. 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* **10**, pp. 1374–1387, Dec 2000.
2. 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")* **19**, pp. 653–673, August 2004.
3. Y. Wang, S. Cui, and J. E. Fowler, "3D video coding using redundant-wavelet multihypothesis and motion-compensated temporal filtering," in *Proceedings of the IEEE International Conference on Image Processing*, **2**, pp. 755–758, (Barcelona, Spain), 2003.
4. X. Li, "Scalable video compression via overcomplete motion compensated wavelet coding," *Signal Processing: Image Communication (special issue on "Subband/Wavelet Interframe Video Coding")* **19**, pp. 637–651, August 2004.
5. S. Choi and J. Woods, "Motion compensated 3-D subband coding of video," *IEEE Transactions on Image Processing* **8**, pp. 155–167, Feb 1999.
6. A. Secker and D. Taubman, "Lifting based invertible motion adaptive transform (LIMAT) framework for highly scalable video compression," *IEEE Transactions on Image Processing* **12**, pp. 1530–1542, Dec 2003.
7. 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.
8. A. Golwelkar and J. Woods, "Scalable video compression using longer motion compensated temporal filters," in *Proceedings of SPIE VCIP*, **5150**, pp. 1406–1416, 2003.
9. N. Mehrseresht and D. Taubman, "An efficient content adaptive motion compensation 3D-DWT with enhanced spatial and temporal scalability," in *Proceedings of the IEEE International Conference on Image Processing*, 2004.

10. 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 IEEE International Conference on Image Processing* **2**, pp. 62–64, 2003.
11. V. Seran and L. P. Kondi, "3D based video coding in the overcomplete discrete wavelet transform domain with reduced delay requirements," *Proceedings of the IEEE International Conference on Image Processing* , Sept. 2005.
12. V. Seran and L. P. Kondi, "Improved temporal filtering scheme to reduce delay and distortion fluctuation in 3d wavelet based video coding," *Proceedings of the IEEE Western New York Image Processing Workshop* , Sept. 2005.
13. D. S. Taubman and M. W. Marcellin, *JPEG2000, Image compression fundamentals, standards and practice*, Kluwer academic Publishers Group, 2002.
14. M. V. der Schaar and D. 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)*, 2003.
15. P. Y. Cheng, J. Li, and C. C. J. Kuo, "Rate control for embedded wavelet coder," *IEEE Transactions on Circuits and Systems for Video Technology* **7**, pp. 692–701, August 1997.
16. A. Said and W. Pearlman, "A new, fast, and efficient image codec based on set partitioning in hierarchical trees," *IEEE Transactions on Circuits and Systems for Video Technology* **6**, pp. 243–250, June 1996.

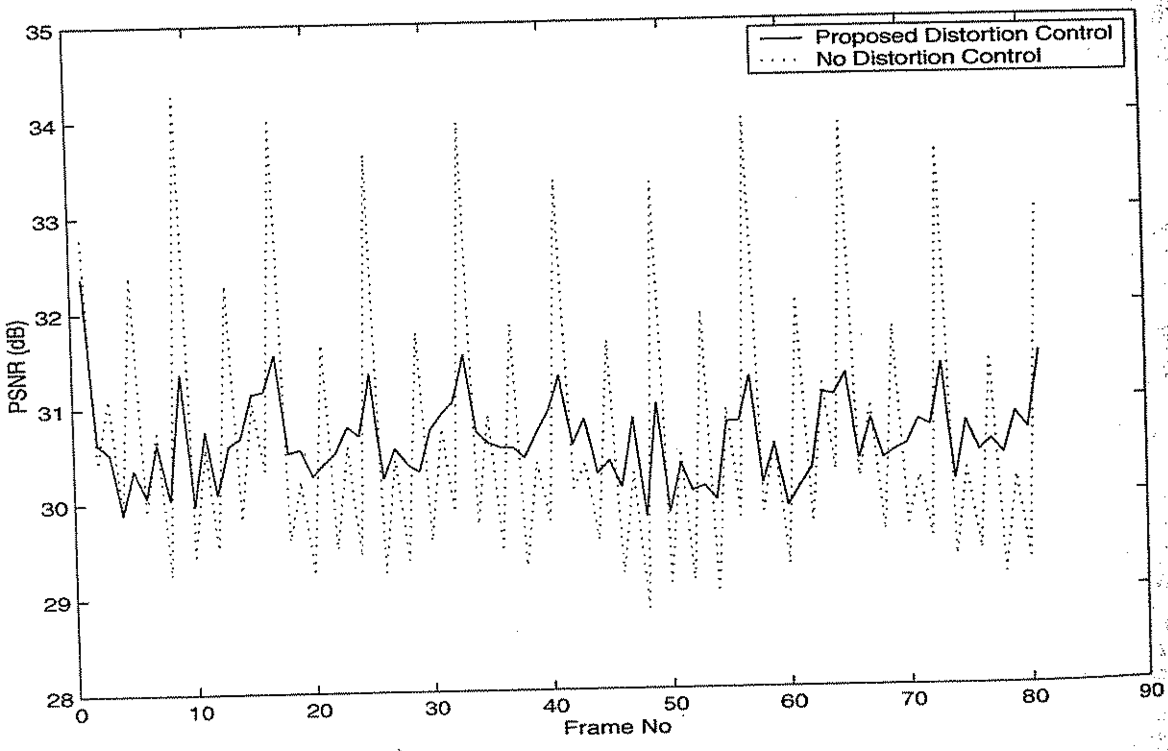

Figure 2. Distortion Control for Football sequence

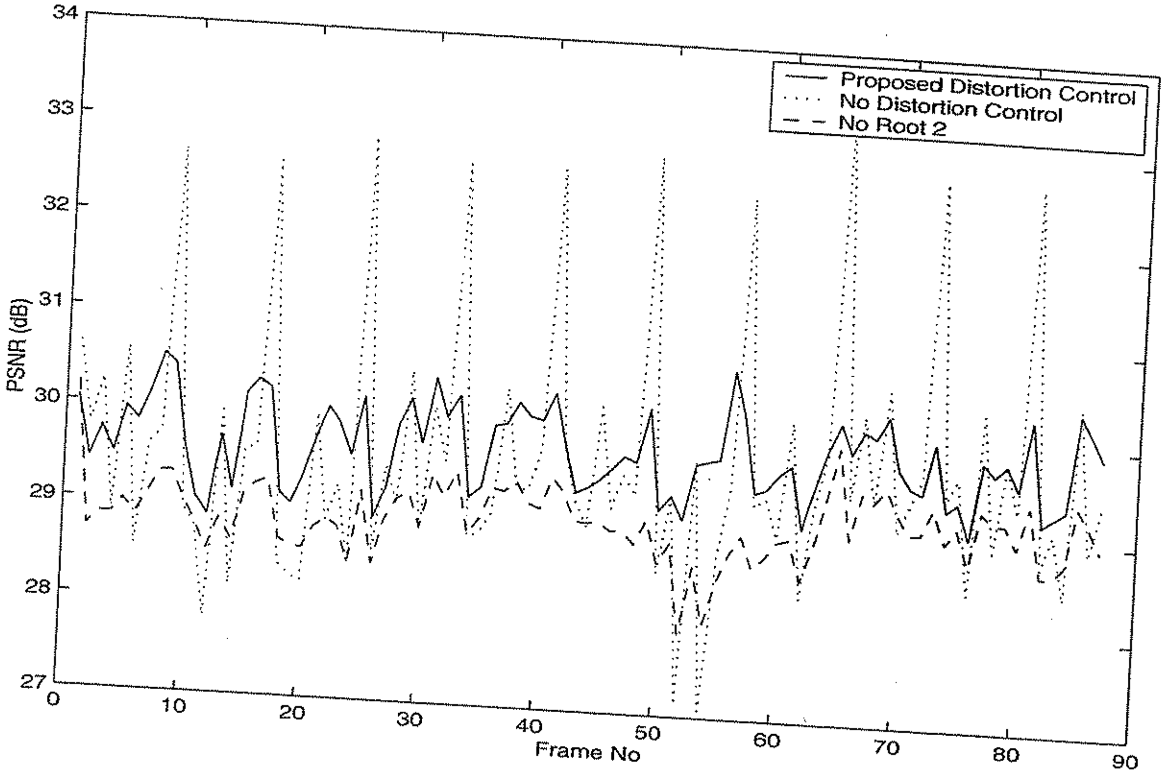

Figure 3. Distortion Control for Garden sequence

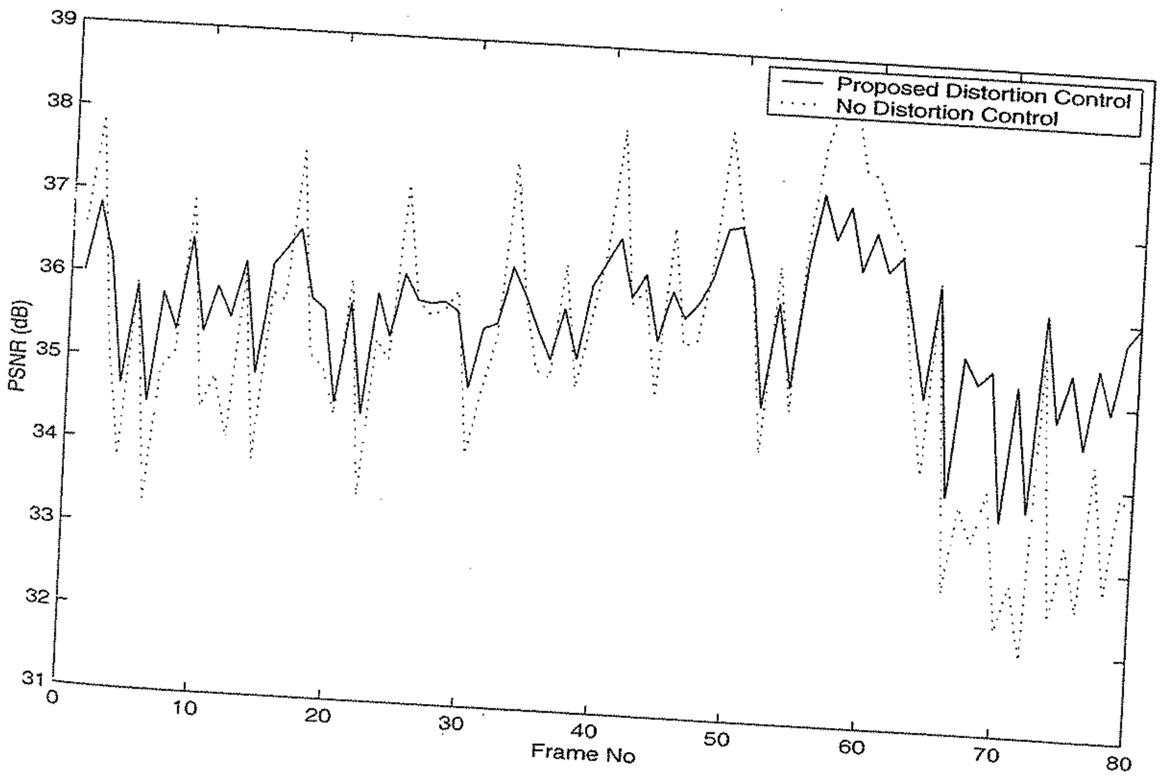

Figure 4. Distortion Control for Foreman sequence

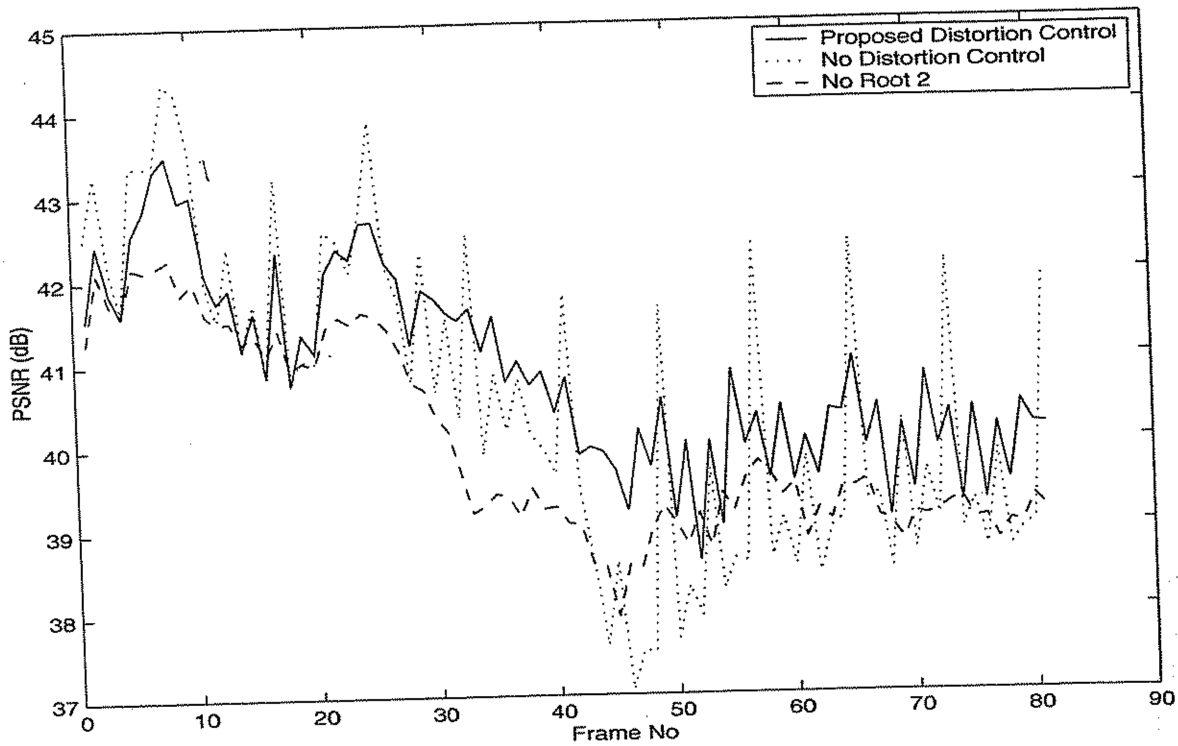

Figure 5. Distortion Control for Susie sequence

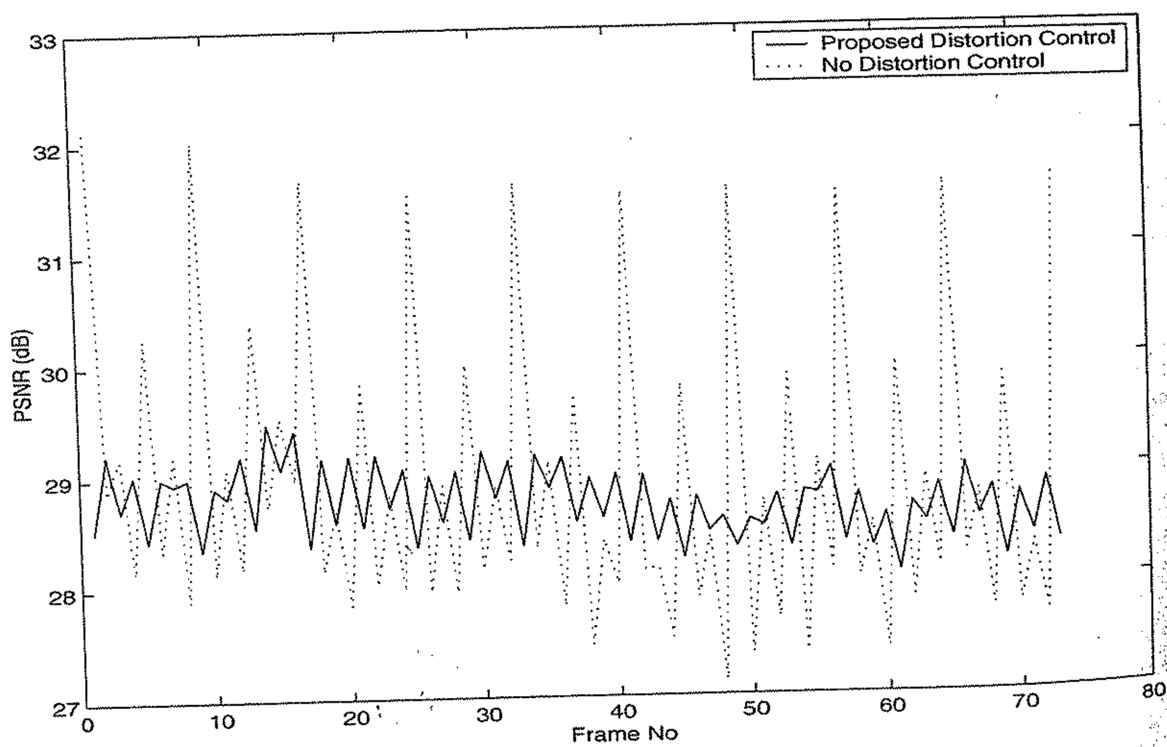

Figure 6. Distortion Fluctuation Control for 1/3 Filter - Football sequence