DRIFT CONTROL IN VARIABLE BITRATE WIRELESS CHANNELS FOR SCALABLE WAVELET BASED VIDEO CODING IN THE OVERCOMPLETE DISCRETE WAVELET TRANSFORM DOMAIN

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. 2133 Fax: (716) 645-3656
Email: { vseran,lkondi }@eng.buffalo.edu

ABSTRACT

In this work, we propose a novel scheme to minimize drift in scalable wavelet based video coding, which gives a balanced performance between compression efficiency and reconstructed quality with less drift. Our drift control mechanism maintains two frame buffers in the encoder and decoder; one that is based on the base layer and the one that is based on the base plus enhancement layers. Drift control is achieved by switching between these two buffers for motion estimation and compensation and base layer information is used for predicting enhancement layers. Our coder is designed for variable bit rate wireless channels and offers high compression efficiency and sustained video quality.

1. INTRODUCTION

Wavelet transform of a signal provides a multi-resolution/multi-frequency representation in both the spatial and frequency domains. Wavelet based image coding has achieved tremendous success both in coding efficiency and in scalability. Many researchers are trying to exploit the advantages of wavelet transforms in video coding. Several works have been recently proposed for motion estimation and compensation in the wavelet domain [1][2].

Layered coding together with error protection techniques offers high error resilience to channel induced errors [3][4]. In traditional coders, motion estimation/motion compensation (ME/MC) is only based on the base layer and it ignores all the enhancement layer information, which may not be always available at the receiver. By neglecting the enhancement layers in the prediction, traditional coders will lose in compression efficiency. Including the enhancement layer in the ME/MC loop introduces drift, defined as the propagation of errors due to partial reception of enhancement information. Base layer prediction using enhancement layers is of particular importance because it offers very good compression efficiency though it suffers from drift in a lossy network. Recent works have shown that drift need not be completely eliminated, but it can be managed in Discrete Cosine Transform (DCT) based video coders [5][6]. Our proposed coder uses both base and enhancement layers for ME/MC in a wavelet based video coder that allows some drift but has a mechanism of managing it.

Another way to totally eliminate drift is to use 3-D subband coding methods. In 3-D subband coding, a group of frames is processed at a time to compress the video sequence [7][8]. Motion compensation in the temporal domain (either 2D+t or t+2D) and with lifting techniques offers high compression and scalability [9][10]. However, 3-D techniques require the availability of future frames for ME/MC. This introduces a delay, which makes them unsuitable for real time video applications.

Our work is an attempt to control drift in wavelet based video coders where enhancement layers are used to predict the base layer information. The proposed coder eliminates the need to transmit an intra frame at regular intervals to completely eliminate drift. The focus of this work is to manage drift in a wireless transmission system with Unequal Error Protection (UEP)/variable bit rate channels that have a known loss rate.

The rest of the paper is organized as follows: In Section 2, we explain our coder architecture and drift control mechanism. In Section 3, we deal with the channel modeling. Finally, in Section 4, we present the simulation results for different packet loss probabilities.

2. DRIFT CONTROLLED CODER

The proposed encoder and decoder are shown in Figure 1 and Figure 2 respectively. The first frame is treated as an intra frame and transformed in the discrete wavelet domain. The Discrete Wavelet Transform (DWT) coefficients are coded using Set Partitioning in Hierarchical Trees (SPIHT) coder [11] to produce a bitstream. The bitstream is partitioned into blocks of data such that there is one block per layer per frame. A block of the embedded bitstream that contains the most significant information forms the base layer. The remaining blocks are used to form one or more enhancement layers. The SPIHT decoded intra frame will be the reference frame for ME of the current (second) frame. The reference frame is transformed in the overcomplete discrete wavelet transform (ODWT) domain and the current frame is transformed using DWT. Motion vectors are estimated using the low band shift method [1]. The prediction error is encoded using the SPIHT coder.

The encoder maintains two frame buffers: Base Buffer Enc_BB for base layer only prediction and Enhancement Buffer Enc_EB for prediction based on base plus enhancement layers. The decoder also maintains two buffers: decoder base buffer (Dec_BB) and decoder enhancement buffer (Dec_EB). The Enc_BB and Dec_BB will maintain a predicted frame using only the base layer, though the ME is done with the Enc_EB information. If the base layer is assumed to be received in full, Enc_BB and Dec_BB have identical data and hence can be used to control drift in the Dec_EB. Given the approximate channel conditions and the rate, the encoder computes a measure of drift (MD) for the Enc_EB output. The MD is

0-7803-9134-9/05/$20.00 ©2005 IEEE

compared with a threshold, which we call as enhancement threshold (ET) and if MD exceeds the preset ET, the drift is significant in the encoder output. Switching the prediction to be based on the Enc_BB for the next frame eliminates the drift introduced in the system. For subsequent frames prediction is again based on Enc_EB.

The use of Enc_EB motion estimation information in Enc_BB for MC will progressively decrease the quality of the Enc_BB. When switching to the base buffer is done, a poor quality Enc_BB would yield poor prediction. Thus, although drift will still be eliminated, compression efficiency will be low. Hence, it is a precautionary measure to arrest any considerable drop in the quality of Enc_BB. Initiating a switching action when the difference between the quality of Enc_BB and Enc_EB (DB) rises above a base threshold (BT) will maintain the Enc_EB quality.

The switching decisions are made in the drift control box based on the two threshold settings and the measure of drift. The switching instances are conveyed to the decoder as control information. The drift control box in the decoder examines the received control information and does the switching between the buffers at exactly the same instance as in the encoder.

2.1. Drift Estimation and Drift Control

We consider the channel model explained in Section 3. The embedded coder generates a bitstream with $Renc_f$ bits per frame. The encoder assumes that all bits are received by the decoder for given $Renc_f$ and computes three peak to signal noise ratio (PSNR) values to calculate the measure of drift (MD) and difference between the quality of Enc_EB and Enc_BB buffers (DB). The three values are PSNR of Enc_EB at rate $Renc_f$ (P_{EB}), PSNR of Enc_BB at base rate R_b (P_{BB}) and the PSNR at the average rate of $Renc_f$ and R_b (P_{Av}). Then MD and DB are calculated as follows:

$$MD = P_{EB} - P_{Av}, \tag{1}$$

$$DB = P_{EB} - P_{BB}, \tag{2}$$

The switching between Enc_EB and Enc_BB occurs when the following conditions are met:

$$MD > ET \;\; or \;\; DB > BT. \tag{3}$$

If this switching is done very often or less frequently, it affects the compression efficiency.

3. CHANNEL MODELLING

The wireless video transmission system consists of an encoder, wireless channel and a decoder. Transmission over a wireless channel is subject to loss and hence the received bits will be fewer than the encoded bits. We assume a wireless medium with a known a priori probabilistic model, which has a feedback channel for error detection and re-transmission [12]. The encoded video frame is partitioned into packets of fixed size of C bits. Packets are indexed sequentially and transmitted at regular time intervals τ_l. These link layer packets are transmitted with a certain success probability p. Packet losses are statistically independent and the lost packet is retransmitted at the next time instant (instantaneous feedback is assumed).

Let the frame rate be $1/t_f$ frames/sec. Then, the number of packets per frame, $N = t_f/\tau_l$. Let X_i be a random variable where $i = 1, 2, 3, \ldots, N$. $X_i = 1$ denotes a successful packet transmission with probability p and $X_i = 0$ denotes a lost packet with probability $1 - p$. The random variable T_i defines the number of successfully received packets after i transmission attempts,

$$T_i = \sum_{j=1}^{i} X_j \quad \text{where } i \leq N. \tag{4}$$

T_i is binomially distributed since we assumed statistically independent packets.

$$Pr(T_i = j) = \begin{pmatrix} i \\ j \end{pmatrix} p^i (1-p)^{i-j}. \tag{5}$$

Therefore, the received number of bits per frame is given by, $R_f = CT_i$. At the encoder, the bits per frame will be always equal to, $Renc_f = CN$.

3.1. Selection of Base Rate

In layered coding, the channel should at least guarantee the lossless delivery of the base layer to the receiver. With the knowledge of success probability and the number of fixed size packets required per frame, we can estimate the base rate for lossless delivery. For our channel model discussed, the base rate is selected as $R_b = Cj^*_{base}$, where j^*_{base} is the minimum value of j_{base} that satisfies the inequality,

$$\sum_{i=j_{base}}^{N} \begin{pmatrix} N \\ i \end{pmatrix} p^i (1-p)^{N-i} \geq Pr_{base}, \tag{6}$$

where Pr_{base} is the probability of successfully receiving the base layer packets.

In the ideal case, Pr_{base} would be equal to 1, i.e., absolutely no base layer packet loss. However, when Pr_{base} is equal to 1, it results in a very low value for the base rate, which is undesirable. In practice, Pr_{base} can be selected to be less than but very close to 1. The selection of Pr_{base} is discussed in Section 4.2.

4. EXPERIMENTAL RESULTS

A wavelet based video coder is implemented using the low band shift method [1]. 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 domain using the block matching technique. The residues are encoded using the SPIHT coder. We use the “Foreman”, “Susie” and “Carphone” video sequences to analyze the performance of the proposed coder. A frame rate of 10 frames/sec is maintained for all sequences and only one intra frame is used in all the simulations.

As discussed in Section 3, the bitstream is broken down into link layer packets of length $C = 320$ bits and time interval $\tau_l = 5$ ms. Simulations are performed for five different values of p $=0.85, 0.87, 0.9, 0.93$ and 0.95. The number of packets and the bits per frame are calculated as, $N = t_f/\tau_l = 20$ and $Renc_f = CN = 6400$ bits.

The encoder assumes that the transmission channel is capable of delivering 6400 bits per frame under lossless conditions. In our implementation, we have one base layer and one enhancement layer. The partitioning of the two layers is done according to base layer rate selection. The results presented for each experiment were averaged over 50 simulations.

4.1. Drift, Optimum and Base Case

Drift is introduced in a single layer coder due to prevailing variations in the channel rate. To gauge the performance of our proposed coder, we compared with three different encoder setups. Drift Case: The video sequence is encoded at 6400 bits per frame and all the bits are used for prediction. Due to lossy channel condition, the drift is introduced at the decoder. Optimum Case: We assumed that each frame is encoded using the number of bits that are actually received by the decoder. Though this not practically achievable, this would serve as the upper bound for the proposed coder. Base Case: We also evaluated the traditional layered coding concept, where only the base layer information is used for prediction and enhancement layers are added only to improve the overall quality. The base layer rate is the same rate used by the proposed coder cases.

4.2. Base layer Selection

As explained in Section 3.1, the base layer is a function of Pr_{base}. Hence, we ran simulations for different values of base probability $Pr_{base} = \{0.9, 0.99, 0.999, 0.9999, 0.99999\}$ that would yield different base layer rates. For each Pr_{base}, we calculate a base layer rate for different p. The results from our simulations for the "Foreman" sequence to identify the base rate is plotted in Figure 3. For any p, it is observed that when $Pr_{base} = 0.999$, we get the best quality. When $Pr_{base}>0.999$, it results in a lower value for base rate, which will reduce the base quality. So when switching is performed, it reduces the overall quality of the decoder output. When $Pr_{base}<0.999$, the base rates will be higher, but the base layer is also subjected to higher loss probability. Hence we obtain a relatively better performance when $Pr_{base} = 0.999$.

4.3. Threshold Setting

Threshold setting is a crucial parameter and decides the switching instances. The range of the enhancement threshold ET is approximately identified by calculating the qualities (PSNR) of the received frames for $p = 1$, $p = 0.95$ and $p = 0.85$ in the "Drift case". The lower and upper limits are set from the difference between $p = 1$ and $p = 0.95$ and $p = 1$ and $p = 0.85$ respectively. From our experimental results for different sequences, we found these limits to be between 1.1 dB and 2.3 dB. The base threshold BT is used to monitor the Enc_BB quality, which is obviously lesser than the quality of Enc_EB. Our experiments were performed with five different sets of ET and BT, $\{TS_1 = (2, 2.5), TS_2 = (1.6, 1.9), TS_3 = (1.2, 1.5), TS_4 = (0.8, 1.0), TS_5 = (0.4, 0.6)\}$ in dB. Figure 4 show PSNR as a function of p for different threshold sets that operate using the optimum base layer rate. From our results, we observed that the threshold set TS_3 gives the best performance. This threshold set is used for all the sequences reported.

4.4. Results

The average PSNR of the received frames for the "Foreman" and "Susie" are plotted in Figure 5 and Figure 6 for different values of p. The threshold set is selected as $(1.2, 1.5)$ dB and base rate corresponding to $Pr_{base} = 0.999$ is used for the proposed coder. Traditional "Base case" does not suffer from drift but the PSNR is less than the optimum by 0.5-0.7 dB. The results also show that the proposed coder is very close to the ideal case (unrealizable) and outperforms the "Base case" by 0.4-0.5 dB. From the plots, we can infer that the drift has been regulated without compromising on coding efficiency and quality.

The base layer rate and the thresholds that were used with the "Foreman" and "Susie" sequences were also tested on the "Carphone" sequence. Figure shows the performance of PSNR vs. frame number for the "Carphone" sequence decoded at $p = 0.85$. When there is no drift management, we can see that the quality degrades with each successive frame. This is because the reference frame at the decoder is not exactly the same as in the encoder. With the proposed buffer switching action, we control drift without using an intra frame. This confirms that our selection of the thresholds and the base rates suits the "Carphone" sequence as well.

5. CONCLUSION

The drift problem in traditional motion compensated predictive coders can be completely eliminated by using the base layer prediction only as in the MPEG4-FGS case. Also, a periodic introduction of intra frames will erase drift. But in both the cases, we need more bits to eliminate drift. In wavelet based video coders using 3-D subband coding methods, drift is eliminated and it also achieves high compression efficiency. But, the 3-D scheme has to process a group of frames to take wavelet transforms, it introduces unacceptable coding delays in transmission. We proposed a novel scheme that gives a performance better than the traditional ME/MC coders and without any delays in transmission. Our proposed coder controls drift without significant loss in compression efficiency. We optimized the coder performance for a wireless variable bit rate channel.

6. REFERENCES

[1] 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, pp. 577–587, April 2000.

[2] Y. Andreopoulos, M. Van der Schaar, A. Munteau J. Barbarien, and P.Schelkens, "Complete-to-overcomplete discrete wavelet transforms for fully scalable video coding with MCTF," in *Proc. SPIE VCIP*, 2003, vol. 5150, pp. 719–731.

[3] R. Aravind, M. R. Civanlar, and A. R. Reibman, "Packet loss resilience of MPEG-2 scalable video coding algorithms," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 6, no. 5, pp. 426–435, Oct 1996.

[4] L. P. Kondi, D. Srinivasan, D. A. Pados, and S. N. Batalama, "Layered video transmission over multi-rate DS-CDMA wireless systems," in *Proc. of SPIE Conference on Image and Video Communications and Processing*, San Jose, CA, 2003, pp. 289–300.

[5] A. R. Reibman and L. Bottou, "Managing drift in a DCT-based scalable video encoder," in *IEEE Data Compression Conference*, March 2001, pp. 351–360.

[6] J. F. Arnold et al, "Efficient drift-free signal-to-noise ratio scalability," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 1, pp. 70–82, Feb 2000.

[7] 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.

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

[9] 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, pp. 1530–1542, Dec 2003.

[10] S. Choi and J. Woods, "Motion compensated 3-D subband coding of video," *IEEE Transactions on Image Processing*, vol. 8, pp. 155–167, Feb 1999.

[11] 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*, vol. 6, pp. 243–250, June 1996.

[12] H. Jenkac, T. Stockhammer, and G. Kuhn, "On video streaming over variable bit rate and wireless channels," in *Packet Video*, France, April 2003.

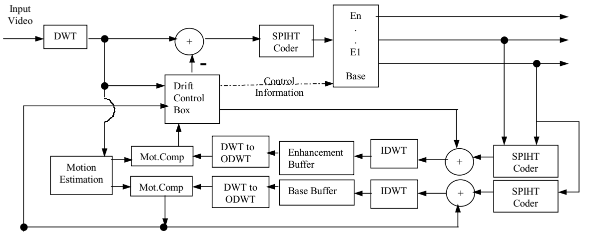

Fig. 1. Proposed Encoder Block Diagram.

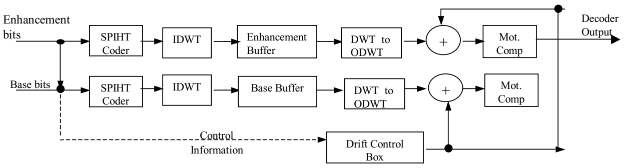

Fig. 2. Proposed Decoder Block Diagram.

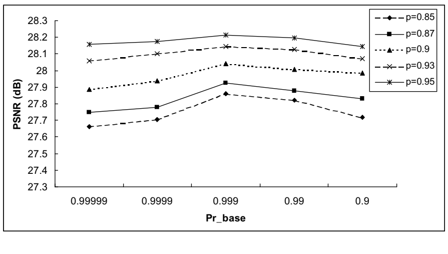

Fig. 3. Base rate selection for "Foreman" sequence in PSNR (dB) vs base rate probabilities Pr_{base} for five p values.

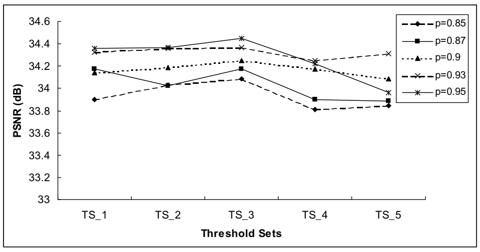

Fig. 4. "Susie" sequence threshold selection: PSNR vs five threshold sets for different p values.

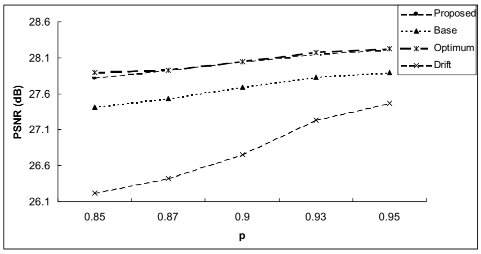

Fig. 5. Comparison of the proposed technique for different p values "Foreman"sequence.

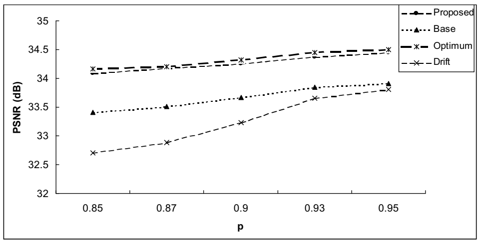

Fig. 6. Comparison of the proposed technique for different p values "Susie"sequence.

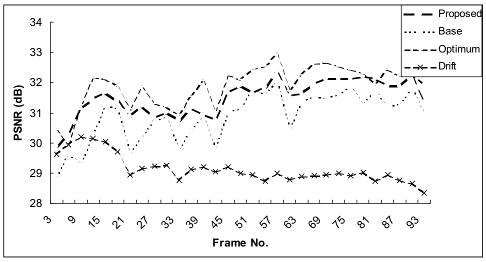

Fig. 7. Frame No. vs PSNR plot for "Carphone"sequence.