DRIFT-CONTROLLED SCALABLE VIDEO CODING IN OVER-COMPLETE WAVELET DOMAIN

Vidhya Seran, Lisimachos P. Kondi
Dept. of Electrical Engineering, 332 Bonner Hall,
SUNY at Buffalo, Buffalo, NY 14260
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 quality. Our drift control mechanism maintains two frame buffers in the encoder and decoder; one for the base layer and the other for the enhancement layer. Drift control is achieved by switching between these two buffers for motion compensation and prediction. In the encoder, the residues are coded using the embedded zerotree wavelet (EZW) algorithm. Our prediction is based on the enhancement layer, which inherently introduces drift in the system, if part of the enhancement layer is not available at the receiver. A measure of drift is computed based on channel information, and a threshold is set. When the measure exceeds the threshold, i.e. when drift becomes significant, we switch the prediction to be based on the base layer, which is always available to the receiver.

The proposed coder determines the threshold and the buffer switching instance based on given channel conditions and the rate. Our coder offers high compression efficiency and acceptable video quality for different bit rate conditions. This proves that we need not completely eliminate drift and decrease compression efficiency to get better video quality.

Keywords: *Wavelet-based video coding, scalability, drift*

1. INTRODUCTION

The popularity of multimedia applications demands support for different receivers that operate at different bit rates, resolution and complexity. This mandates the need for a scalable video coder with high compression efficiency. Wavelet based image coding has the very best coding efficiency and provides SNR scalability, besides resolution scalability. This has kindled the minds of many researches to explore the possibilities to use wavelets in video coding. Initial research in this area had to face challenges like aliasing caused by decimation in wavelet decomposition. This led to the use of over-complete wavelet decomposition to overcome the aliasing problem. Several works have been recently proposed for motion estimation and compensation in the over-complete wavelet domain [1]-[3]. Using an embedded coder [4] [5], the residues are compressed to produce a bitstream.

The embedded bitstream can be partitioned into layers. The base layer also includes the motion vectors. One or more enhancement layers are used to achieve further refinements to the base layer. Layered coding together with error protection techniques, offers high error resilience to channel induced errors.

Traditional coders base motion estimation/motion compensation (ME/MC) on the base layer, thus ignoring all enhancement layer information. This is done because the enhancement layers are not always available at the receiver. Hence they lose in compression efficiency and a loss of 0.5-1.5 dB for every layer [6] [7]. Basing ME/MC on the enhancement layer introduces drift, defined as the propagation of errors due to partial reception of enhancement information. Drift when left unchecked results in poor video quality. Base layer prediction using enhancement layers in the prediction loop is of particular importance because it offers very good compression efficiency though it suffers from drift in a lossy network. Fine granularity Scalability (FGS) coders are designed in such a way it offers zero tolerance to drift [8]. Recent work has shown that drift need not be completely eliminated, but it can be controlled [6] [9] [10].

Another way to completely eliminate drift is to use three dimensional (3-D) subband coding methods [11]. ME/MC introduces a predictive feedback loop, which hinders in achieving a high degree of scalability. This is due to the fact that scalable coding introduces drift in the predictive coders. In 3-D subband coding, a group of frames is processed at a time and predictive feedback loop is not required, thus eliminating drift. The encoder and decoder must buffer the required number of frames before they can apply wavelet transform. This introduces

CCECE 2004- CCGEI 2004, Niagara Falls, May/mai 2004
0-7803-8253-6/04/$17.00 ©2004 IEEE

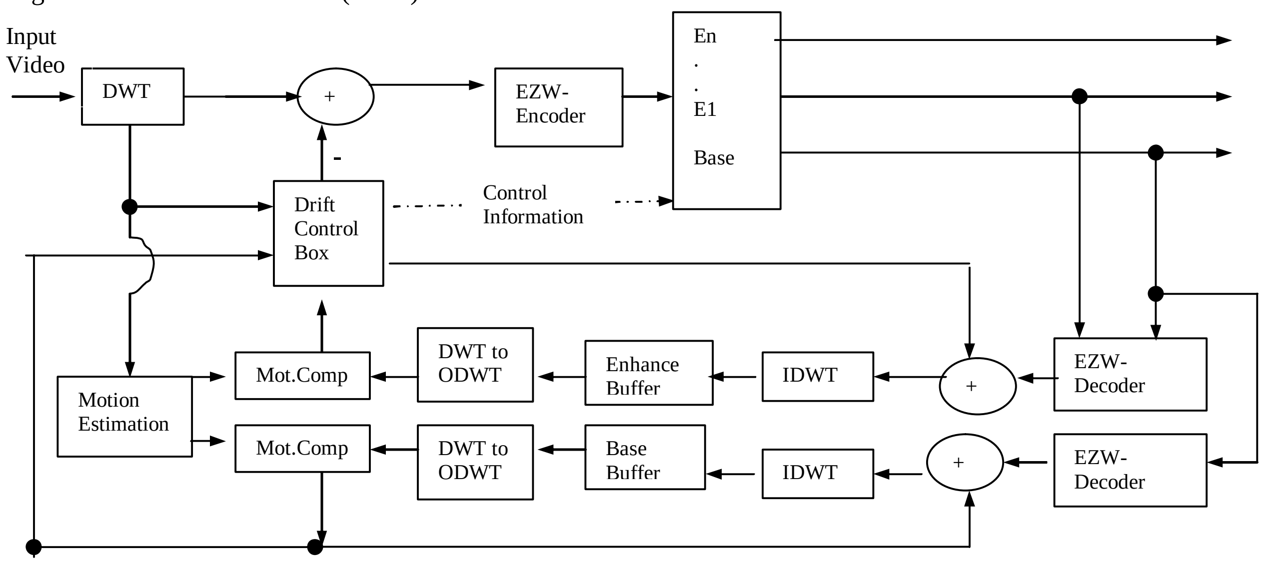
a delay, which makes it unsuitable for real time video applications like tele- conferencing.

Our work is an attempt to control drift in ME/MC wavelet based video coders where enhancement layer information is 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. Section 2 explains our coder architecture and Section 3 deals with the drift estimation and channel modeling. In Section 4, we present the simulation results for different bit rate conditions.

2. DRIFT CONTROLLED SCALABLE VIDEO CODER

The proposed encoder is shown in Figure 1. The previous frame is transformed in the over-complete wavelet domain and the current frame is transformed using discrete wavelet transform (DWT). Motion vectors are estimated using the low band shift method [1]. The prediction error is encoded using the embedded zerotree wavelet coder (EZW) [4]. The bitstream is partitioned into chunks of data such that there is one chunk per layer per frame. A portion of the embedded bitstream that contains the most significant information together with the motion vectors forms the base layer. The remaining bitstream is used to form one or more enhancement layers. Our prediction is normally based on the enhancement layer, which inherently introduces drift in the system if part of the enhancement layer is not received by the system.

The encoder maintains two frame buffers: one for base layer prediction and other for base plus enhancement layer prediction. Without loss of generality, we assume that the base layer is always available at the decoder. Hence we will be using this base layer information to control drift. Given the approximate channel conditions and the rate, the encoder computes measure of drift for enhancement (MDE). When MDE exceeds an enhancement threshold (ET), the prediction is switched to use base buffer information. It is to be noted that the use of enhancement layer in the base layer prediction introduces a drift in the base buffer. Thus the quality of the base buffer will decrease with time. However both the encoder and decoder have exactly the same base layer buffer frames. Therefore, the base layer buffer can be used to eliminate drift.



DWT- Discrete Wavelet Transform
EZW-Embedded Zerotree Wavelet
ODWT-Over-complete Wavelet Transform
IDWT- Inverse Discrete Wavelet Transform
En … E1 – Enhancement layers
Mot. Comp – Motion Compensation

Fig.1. Proposed encoder

The drift in the base buffer is due to use of the enhancement layer prediction information in the base buffer motion compensation. The base buffer drift should also be maintained below a base threshold (BT). This is a precautionary measure to reduce the amount of bits required to control drift in the enhancement layer. The switching decisions are done in the drift control box based on the threshold settings and measure of drift both in base and enhancement buffers. The switching instance is conveyed to the decoder as control information.

The decoder also maintains two buffers and a drift control box. The prediction is based on either the enhancement or the base buffer. The drift control box in the decoder examines the received information and does the switching between the buffers at exactly the same instance as in the encoder.

3. DRIFT ESTIMATION AND CHANNEL MODELLING

In a typical video transmission scenario, the bit rate R of the channel will not be constant during transmission, but the approximate minimum R_{min} and a maximum R_{max} bit rate of the channel will be known. If the encoder is optimized to perform over this range, $R_{min} \leq R \leq R_{max}$, the decoder can reconstruct the video with acceptable quality. In [12] [13], different optimization criteria are used to minimize distortion for channel rates. We consider the following rate constraints for our channel: $R_{base} \leq R_{min}$ and $R_{base+enh} \leq R_{max}$ where R_{base} is the base layer bit rate and $R_{base+ enh}$ is the base plus enhancement bit rate. R_{base} is chosen to be a constant satisfying the rate criterion. Encoding is done at a rate R, so as to produce a base layer with rate R_{base} and one or more enhancement layers. At the decoder a fixed portion of the enhancement layer is truncated to match the channel rate.

The encoder assumes all bits are received by the decoder for given rate R and it computes the measure of drift. The encoder calculates the mean square error for base buffer, enhancement buffer, measure of drift for base (MDB) and measure of drift for enhancement (MDE) buffer cases. The switching between base and enhancement buffers occurs when the following conditions are met:

$$MDE > ET \text{ and } MDB > BT \quad (1)$$

If this switching is done very often or less frequently, we need more bits to encode. If we choose a large BT, the drift in the base layer will become significant and we require more bits to maintain the quality for the successive frames. If ET is small, then switching occurs frequently and is also not desirable. Therefore setting the threshold value is very important. Here, threshold *ET* and *BT* are chosen based on simple heuristic methods.

4. EXPERIMENTAL RESULTS

A wavelet based video coder is implemented using low band shift method [1] for our simulation purpose. A Daubechies (9,7) filter with a three level decomposition is used to compute wavelet coefficients. The motion estimation is performed in the over complete domain using block matching techniques. A 16 X16 block is matched in a search window of [-16, 16]. The residues are encoded using the EZW coder [4]. A frame rate of 30 frame/sec is maintained. Only one intra frame is used in all the simulations. The base layer bit rate is set to 75 Kbps. The bitstream is encoded for a channel rate 500 Kbps for all cases and the same bitstream is used decode at various bit rate conditions. We use “Foreman”, “Susie” and “Carphone” video sequences to analyze the performance of the proposed coder.

Two bitstreams are generated, one with the prediction based on base layer (base bitstream) and the other in which the prediction is based on the enhancement layer (enhancement bit stream) for comparison studies. The average PSNR of the received frames at different bit rate is plotted in Figure 3 and 4. When the enhance bitstream is decoded at a rate lower than the encoded rate, the quality at the decoder drops drastically which is denoted by the line ‘Drift’. The upper bound case denoted by ‘Enhancement’ line is obtained, when encoding and decoding are done at the same rate and the prediction is based on the enhancement layer. The decoding of the base bitstream at different rates eliminates drift but the quality is lower by 2-2.5 dB when compared to the upper bound case. With the proposed coder we get a performance improvement over the base prediction case of the order of 1.5 dB at higher bit rates. At low bit rates, we get a marginal improvement of 0.1-0.15dB when compared to the no drift base case. The switching decision will have different effects on the performance, for different sequences. Figure 5 shows the performance of PSNR vs. frame number for “Carphone” sequence decoded at 300 Kbps. When there is no drift control, 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.

5. CONCLUSIONS

Drift problem in traditional motion compensated predictive coders can be completely eliminated by using the base layer prediction only as in 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. So 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. More sophisticated models can be created for the proposed scheme and this will further help in increasing the performance.

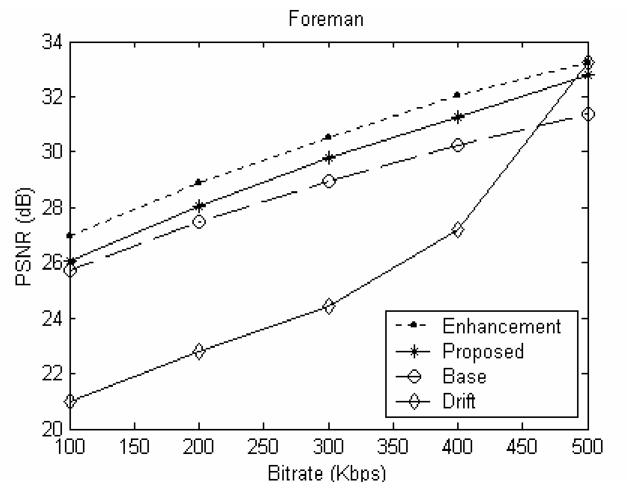


Figure 3: Foreman 30 frames/s for 100frames

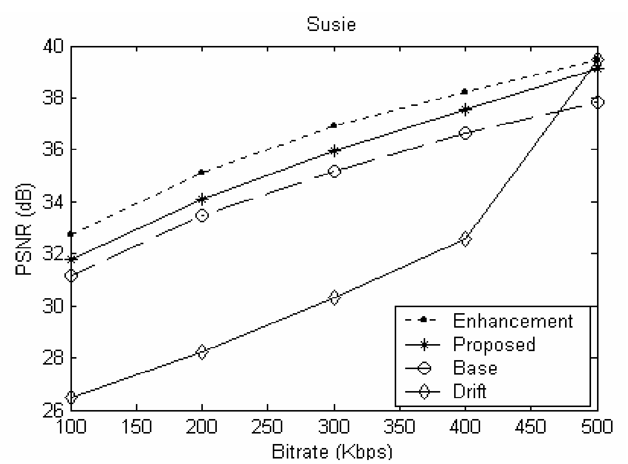


Figure 4: Susie 30 frames/s for 100frames

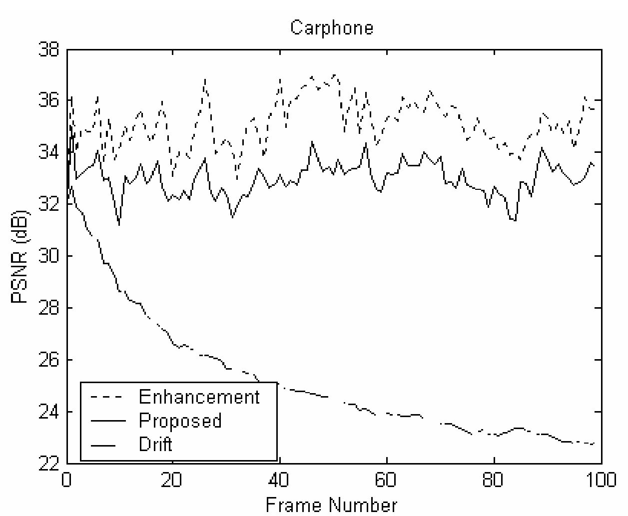


Figure 5: Carphone 30 frames/s for 100 frames

References

[1] H. W. Park and H. S. Kim, "Motion estimation using low-band-shift method for wavelet-based moving-picture coding," *IEEE Trans. Image Processing* 9, pp. 577–587, Apr. 2000.

[2] X. Li, L. Kerofski, and S. Lei, "All-phase motion compensated prediction in the wavelet domain for high performance video coding," in *Proc. IEEE Int. Conf. Image Processing ICIP '01*, Thessaloniki, GR, 2001, vol. 3, pp. 538–541.

[3] Y. Andreopoulos, A. Munteanu, G. Van der Auwera, P. Schelkens, and J. Cornelis, "Wavelet-based fully scalable video coding with in-band prediction," in *Benelux Signal Processing Symposium*, Leuven, BE, Mar. 2002, IEEE.

[4] J. Shapiro, "Embedded image coding using zerotrees of wavelets coefficients," in *IEEE Trans. Signal Processing*, vol. 41, pp. 3445–3462, Dec. 1993.

[5] A. Said and W. Pearlman, "A new, fast, and efficient image codec based on set partitioning in hierarchical trees," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, pp. 243–250, June 1996.

[6] A.R. Reibman, L. Bottou, "Managing drift in a DCT-based scalable video encoder," *IEEE Data Compression Conf.*, pp.351-360, Mar. 2001.

[7] A.R. Reibman, L. Bottou, A. Basso, "DCT-based scalable video coding with drift," *IEEE ICIP 2001*, pp 989-992, Oct. 2001.

[8] H. Radha *et al.*, "Fine-granular-scalable video for packet networks," in *Packet Video Workshop*, New York, Apr. 1999.

[9] C. Mayer *et al.*, "Bit plane quantization for scalable video coding," *Proc. SPIE Visual Communications and Image Processing*, Jan. 2002.

[10] J. F. Arnold *et al.*, "Efficient drift-free signal-to-noise ratio scalability," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, pp. 70–82, Feb. 2000.

[11] B-J. Kim, Z. Xiong, and W. A. Pearlman, "Low Bit-Rate Scalable Video Coding with 3D Set Partitioning in Hierarchical Trees," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, pp. 1374-1387, Dec.2000.

[12] A.R.Reibman, L.Bottou and A.Basso, "Scalable video coding with managed drift," in *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, pp. 131–140, Feb 2003.

[13] R. Zhang *et al.*, "Video coding with optimal inter/intra-mode switching for packet loss resilience," *IEEE J. Select. Areas Commun.*, vol. 18, pp. 966–976, June 2000.