# ADAPTIVE LIFTING CODING SCHEME FOR VIDEO SCENE CHANGES

**Wade Trappe, Haitao Zheng, and K.J. Ray Liu**
Institute for Systems Research,
University of Maryland,College Park, MD 20742
e-mail: wxt, haitaoz, kjrliu@eng.umd.edu; phone: 301-405-6619

**Abstract - The lifting scheme is used as a motivational tool for designing a better temporal subband scheme in video coding. The new subband coding scheme replaces the typical Haar filterbank. A simple criterion is used for the detection of scene changes in video sequences, and a modification to the Haar lifting scheme is developed that allows for adjusting the filterbank near a scene change. The new edge adaptive lifting scheme is applied to a scene change in video data and achieves improved performance at scene change frames when compared with the lifting Haar filterbank.**

## INTRODUCTION

The problem of designing filterbanks for signal processing applications has typically involved using techniques based on wavelet analysis. These methods, and their corresponding filterbanks, once designed are fixed systems. In video applications, the temporal filterbank that is most often employed is the Haar wavelet system. This filterbank consists of taking frame averages as lowpass data and frame differences for highpass data.

Lifting theory provides a manner by which we can take a basic filterbank and add new components to achieve better results. In this paper we use the lifting framework to improve the design of the temporal subband decomposition in 3D subband coding. Our goal is to design a nonlinear temporal filterbank that codes video sequences with scene changes better than the traditional Haar system.

## LIFTING SCHEME

The lifing scheme was originally presented by Wim Sweldens for designing new biorthogonal wavelet bases from old wavelet bases[1]. In a paper by Sweldens and Daubechies[3], they show that all perfect reconstruction filterbanks may be factored into a sequence of lifting steps. The lifting scheme may be viewed either in the Z domain or in the temporal domain. The temporal domain interpretation is the basis upon which we develop our approach. We now describe the temporal interpretation of lifting.

0-7803-5610-1/99/$10.00 © 1999

1. Split: This step takes a signal $S(n)$ and splits it into two distinct sets $L(n)$ and $H(n)$. It is required that we have some method to join $L(n)$ and $H(n)$ together to form the original $S(n)$. In this paper we use the polyphase transform as the splitting step.

2. Predict: We have broken the signal $S(n)$ into two halves. We would like to use the values of $L(n)$ to predict the values of $H(n)$. The goal of the predict stage is to reduce the dynamic range of $H(n)$ by using $L(n)$ to predict $H(n)$ and updating as:

   $$H^{new}(n) = H(n) - P(L(n))$$

   We will drop the superscript in $H^{new}(n)$ as is convention.

3. Update: The purpose of the update stage is to use information in the highpass signal to update and reduce the effects of aliasing in the low-pass signal. Aliasing is reduced by preserving a scalar quantity $Q()$, like the mean, i.e.

   $$Q(L(n)) = Q(S(n))$$

   To do this, we would like to reuse the work done in stage 1 and 2. Therefore one uses $H(n)$ to update the $L(n)$ so that we preserve $Q()$:

   $$L^{new}(n) = L(n) + U(H(n))$$

   Again, we will drop the use of the superscript in the rest of the paper.

## EDGE ADAPTIVE LIFTING SCHEME

One of the key motivations for using the lifting scheme is that it allows one to make a new filterbank from an old filterbank. What has made the lifting scheme so attractive is that it offers a completely spatial (temporal in the one-dimensional case) interpretation of the perfect reconstruction problem. It is this interpretation that allows us to adapt the filterbank locally near edges.

Originally we were inspired by the work of [4], but our initial investigations found that their use of polynomial predictors didn't achieve good performance on real video data with scene changes. It was observed that if one followed the magnitude of single pixel as a function of time, then the resulting data was best approximated as a piecewise constant time series. Higher order predictors don't achieve improvement over a linear predictor on constant data. At the same time, a 3D subband coding scheme using a lifting version of the Haar filterbank in the time domain exhibited degraded performance near a scene change.

In order to overcome this quality reduction near scene changes, it is necessary to improve the predictor performance near these temporal edges. As [4] observed, any edge detection and predictor selection criterion must be entirely based on $L(n)$, and in order to keep the analysis and synthesis stages synchronized, one must use the dequantized $\hat{L}(n)$, see figure 1.

A preliminary criterion has been developed for detecting scene changes and adapting the predictor near scene changes. The resulting edge adaptive lifting scheme is now described. First, the splitting step yields two signals $L(n) = S(2n)$ and $H(n) = S(2n+1)$. We switch update and predict, and update the lowpass signal as

$$L(n) = (S(2n) + S(2n+1))/2.$$

The next step is to predict $H(n)$ using $L(n)$. In an update-predict version of Haar, this would mean $H(n) = H(n) - L(n)$. In order to adapt near edges we need to first determine where significant scene changes occur in a video sequence. This must be determined using the dequantized lowpass data. Scene change is determined by calculating the average magnitude of pixel differences between successive frames of $\hat{L}(n)$ and if the average is above a threshold then a flag is toggled to indicate a scene change at that particular frame. The highpass processing then uses a modified Haar prediction scheme that we refer to as the left-right Haar predict near scene changes, and the normal Haar predict away from scene changes.

In order to explore the left-right Haar, we must examine the possible scene change scenarios. There are three basic scene change scenarios that we try to exploit, see figure 2. If one draws some piecewise constant figures according to the cases, one can observe that in Case I it is better to estimate $S(2n+1)$ by $L(n+1)$ than $L(n)$. In the remaining two cases the choice of $L(n)$ is appropriate. Therefore a criterion was needed to determine when Case I occurs so that $L(n+1)$ may be used in the prediction.

Ignoring issues of quantization, define the following difference operator

$$\tilde{\nabla}(n) = L(n) - L(n-1)$$

with the boundary case $\tilde{\nabla}(0) = 0$. The three cases correspond to when $|\tilde{\nabla}(n)|$ is large. Then for Case I we get that $\tilde{\nabla}(n+1) \simeq (S(2n+1) - S(2n))/2$ where the $\simeq$ denotes that approximations were made using the piecewise constant assumption. In Case II we get that $\tilde{\nabla}(n+1) \simeq 0$, and Case III also gives $\tilde{\nabla}(n+1) \simeq 0$. In Case II and Case III we want to use $L(n)$ to predict $H(n)$, which corresponds to looking at the frames to the left. In Case I it is desirable to use $L(n+1)$ which corresponds to using frames to the right. We therefore propose to use $\tilde{\nabla}(n+1)$ as a criterion for choosing whether a pixel in a scene change frame is a left case or a right case. To do this, we simply set a threshold $\epsilon$ and if $|\tilde{\nabla}(n+1)| > \epsilon$ then we have Case I and must use a right predictor, else we simply use the normal Haar predictor.

## SIMULATION RESULTS

In this section we present results for the edge adaptive lifting filterbank of the last section. The goal behind our simulation was to demonstrate the potential for coding gain through improved reduction of a highpass signal's energy near a scene change.

In order to simulate a scene change we took two different video sequences and grafted them together. In particular we chose 9 frames from one video sequence and 11 frames from a second video sequence to form a block of video with 20 frames. The frames are referenced from 0 to 19 and so the scene change occurs between frame 8 and 9.

Figure 3 shows the comparison of edge adaptive Haar and a direct implementation of the Haar system. The proposed adaptive algorithm achieves significant performance improvement at frames where a scene change occurs. In both cases, the spatial processing employed was two levels of a Daubechies wavelet decomposition. The 1.0bpp/0.5bpp corresponds to allocating 1bpp to the lowpass subband and 0.5bpp to the highpass subband. As a comparison between the two reconstructions, we show the results of frame 9, see figure 4. One can see the improved visual quality of the edge adaptive Haar algorithm. The jagged nature of the PSNR performance as a function of frame number for the proposed scheme is due to a feedback of quantization noise that is inherent in the lifting scheme. Even indexed frames, corresponding to the lowpass subband, experience an increase in noise stemming from the inverse update routine. We are investigating ways of mitigating this undesirable problem.

## SUMMARY

A method has been presented for improving the temporal subband decomposition in 3D subband coding schemes using the lifting framework. The method involved replacing the standard Haar prediction with a modified Haar prediction scheme based on scene change criterion.

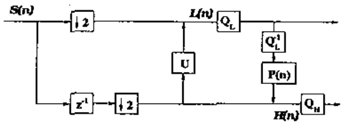
A simulation of a scene change in a video sequence was created and used to examine the performance of the new method and compared with the direct implementation of the Haar system. An improvement near the scene change was observed.

The presented work is only partially completed. In particular, we plan to further investigate the scene change criterion, as well as an automatic method for determining an appropriate threshold. We do not feel we have exploited all the information in a scene change, and would also like to examine any ways to improve the robustness of our case selection criteria. Finally, we are investigating methods for decreasing the feedback of quantization noise that is associated with using the lifting scheme.

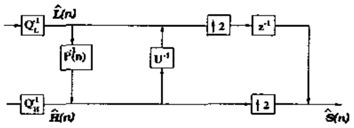
## References

[1] W. Sweldens and P. Schroder. Building your own wavelets at home. In *Wavelets in Computer Graphics*, pgs 15-87, ACM SIGGRAPH Course Notes, 1996.

[2] P.P. Vaidyanathan, *Multirate Systems and Filter Banks* Prentice-Hall 1993.



(a) Analysis



(b) Synthesis

Figure 1: Synchronized lifting diagram

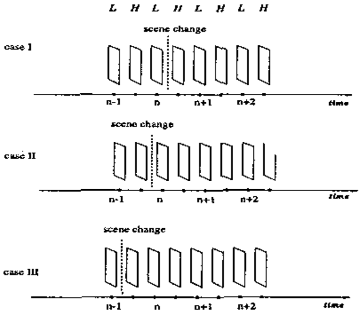
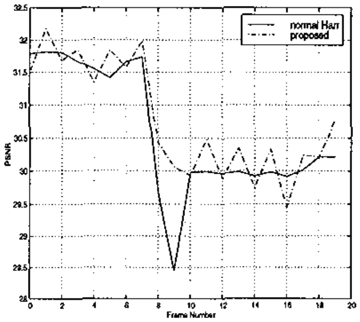


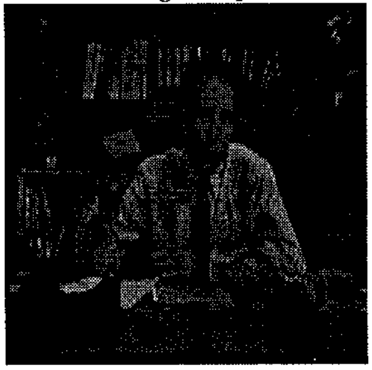
Figure 2: Scene Change Scenarios



1.0/0.5bpp

Figure 3: Performance comparison between Haar and edge adaptive lifting Haar

frame 9: Edge Adaptive Haar



frame 9: Conventional Haar

Figure 4: Comparison between Edge Adaptive Haar and Conventional Haar recontruction of frame 9, using 1.0bpp for low pass and 0.5 bpp for highpass

[3] Ingrid Daubechies and Wim Sweldens. Factoring wavelet transforms into lifting steps. preprint from: http://cm.bell-labs.com/who/wim/

[4] R. Claypoole, G. Davis, W. Sweldens, and R. Baraniuk. Nonlinear wavelet transform for image coding. Proceedings of 31st Asilomar Conference on Signals, Systems, and Computers, Volume 1, pp 662-667, 1997