ANTIALIASING SCALABLE VIDEO WITH A MODULATED LIFTING STRUCTURE

Jonathan Gan and David Taubman

The University of New South Wales

ABSTRACT

A key problem existing in scalable video based on wavelet transforms is the incidence of aliasing at reduced spatial resolutions. The problem exists because of a fundamental limit in the design of wavelet filters. However, typical video frames are observed to have a strong rolloff in the high frequencies before the Nyquist frequency, which we term the "deadzone". In this paper, we propose a transform that not only exploits this property, but replicates the property in successive DWT levels. The transform is created by modifying the lifting implementation of any regular discrete wavelet transform with two additional modulated lifting steps. The two steps act to transfer aliased content from the low pass band into the deadzone of the high pass band. The proposed transform still retains the property of perfect reconstruction. Visual confirmation shows that the transform performs well in antialiasing at half resolution. However, currently the modulated lifting transform exacts a significant coding penalty relative to baseline DWT compression.

Index Terms— Antialiasing, scalable video, modulated lifting

1. INTRODUCTION

In recent times there has been considerable research into scalable video compression. This has been partly reignited by the discovery of motion compensated lifting as a means to incorporate motion models within a temporal wavelet transform structure without losing perfect reconstruction. This raises the possibility of efficient 3D motion compensated wavelet transforms, with multi-resolution attributes in both space and time. One problem which has become apparent with such approaches is the visible impact of aliasing distortion at reduced spatial resolutions. This is caused by the slow rolloff of low pass wavelet analysis filters. While the levels of aliasing produced by common wavelet kernels, such as the Daubechies 9/7, are acceptable for multi-resolution image compression, the same is not true for video; the reason is that the eye is much more sensitive to aliasing in the presence of motion.

The fundamental problem is well known: given any 2 channel subband transform with low and high pass analysis filters $\hat{h}_0(\omega)$ and $\hat{h}_1(\omega)$, perfect reconstruction requires

$$\hat{h}_0(\omega)\hat{h}_1(\pi-\omega)+\hat{h}_0(\pi-\omega)\hat{h}_1(\omega)=1. \tag{1}$$

In particular, this implies that at $\hat{h}_0(\frac{\pi}{2})\hat{h}_1(\frac{\pi}{2}) = \frac{1}{2}$, meaning that substantial attenuation of the aliasing contributions (frequency components beyond $\omega = \frac{\pi}{2}$) can only be obtained at the expense of large amplification of the frequencies around $\frac{\pi}{2}$ in the high-pass channel.

Various researchers (e.g. [1] and [2]) have proposed wavelet transforms with low pass filters designed to roll off faster than the 9/7. [3] proposes an adaptive approach, in which the low pass filter is affected by the presence of edges. Ultimately, however, all such approaches are constrained by equation (1), which limits the degree to which aliasing can be suppressed.

An alternative approach is to construct a multi-resolution Laplacian spatial pyramid. This allows each successive reduced spatial resolution to be generated using arbitrarily selected low pass anti-aliasing filters. The price paid for this flexibility is a loss of critical sampling – dyadic pyramids oversample each frame by $4:3$. Recent extensions to the H.264 video compression standard to enable scalability [4] are based on this principle. Santa Cruz et al. [5] propose a substantially improved spatial pyramid in which update steps are used to orthogonalise the quantisation errors produced at different levels of the pyramid.

In this paper, we take a third approach, which preserves critical sampling. Starting with a conventional wavelet transform, our approach moves aliasing left behind in the low-pass (L) band *after* DWT analysis, into the high-pass (H) band. The method is completely reversible, ensuring perfect reconstruction; it is achieved by 2 additional steps in the lifting structure, with the important twist that these operate on modulated versions of the subband sequences. This process can be iterated over multiple DWT levels, so that antialiasing is achieved at each resolution.

In Section 2 we introduce the proposed lifting structure, and show how it reduces aliasing. Section 3 provides design guidelines for the low pass filters within the modulated lifting steps. In Section 4 we discuss the coding penalty incurred by this method, and a possible strategy for improving this. We then compare the scheme with baseline DWT analysis in terms of compression efficiency.

2. FREQUENCY MODULATED LIFTING

The key observation behind our proposed approach is that the power spectrum of typical video frames decays rapidly before the Nyquist frequency is reached. In fact, this is by design. Video cameras incorporate optical anti-aliasing filters and video resampling algorithms employ digital filters which strongly attenuate high frequency content so as to minimise the production of aliasing effects. The actual portion of the spectrum with negligible energy content depends on the camera or processing filters employed; however, we take the "deadzone" region to have some known width, Δ.

The basic idea is to follow a conventional wavelet/subband transform with two modulated lifting steps, which serve to move aliasing components from the upper edge of the low-pass subband into the deadzone within the high-pass. This is illustrated in Figure 1. The modulated lifting steps are shown in Figure 2. In the first step, we multiply the L band by $(-1)^n$ and add a low-pass filtered copy of this modulated signal to the H band. We refer to this as the "transfer" step, since it transfers aliasing content from the L band to the deadzone in the H band. Note that modulation reverses the spectrum of the L band so that the low-pass filter h_t extracts the aliasing components. In the second ("cancellation") step, another low pass filter h_c selects the same aliased content from the H band and subtracts it from the L band.

Evidently, we are exploiting the existence of the deadzone; with-

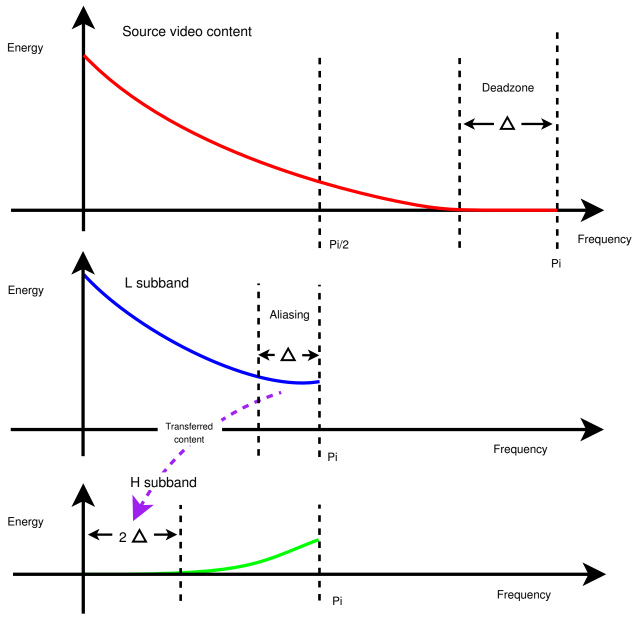

Fig. 1. Notional spectral content of video

out this, the cancellation step would leave non-negligible aliasing contributions from the H band in the L band. In practice, of course, there will always be some spectral content in the deadzone. Moreover, non-ideality of h_c also leads to some leakage between the H and L bands. Nevertheless, we can expect such leakage terms to have much lower power than the original aliasing content which has been cancelled.

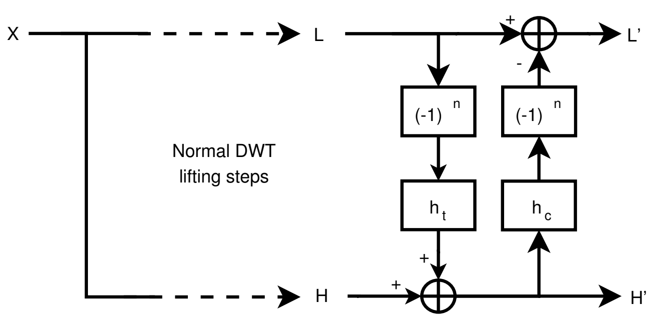

Fig. 2. Modulated lifting analysis structure

2.1. Analysis of the transform

Let $\hat{x}(\omega)$ denote the Fourier transform of a one dimensional signal $x[n]$. Write $\hat{h}_0(\omega)$ and $\hat{h}_1(\omega)$ for the frequency responses of the low and high pass DWT analysis filters, respectively. Prior to modulated lifting, the L and H subband are given by

$$\hat{f}_{\mathrm{L}}(\omega) = \hat{h}_0(\frac{\omega}{2})\hat{x}(\frac{\omega}{2}) + \hat{h}_0(\pi - \frac{\omega}{2})\hat{x}(\pi - \frac{\omega}{2}) \tag{2}$$

$$\hat{f}_{\mathrm{H}}(\omega) = \hat{h}_1(\pi - \frac{\omega}{2})\hat{x}(\pi - \frac{\omega}{2}) + \hat{h}_1(\frac{\omega}{2})\hat{x}(\frac{\omega}{2}) \tag{3}$$

where the first term in each equation is the desired signal, and the second term represents aliasing.

The modulated lifting steps shown in Figure 2 produce modified L$'$ and H$'$ bands, satisfying

$$\hat{f}_{\mathrm{H}'}(\omega) = \hat{f}_{\mathrm{H}}(\omega) + \hat{h}_t(\omega)\hat{f}_{\mathrm{L}}(\pi - \omega) \tag{4}$$

$$\hat{f}_{\mathrm{L}'}(\omega) = \hat{f}_{\mathrm{L}}(\omega) - \hat{h}_c(\pi - \omega)\hat{f}_{\mathrm{H}'}(\pi - \omega) \tag{5}$$

Substituting (4) into (5), we obtain

$$\hat{f}_{\mathrm{L}'} = \left[1 - \hat{h}_c(\pi - \omega)\hat{h}_t(\pi - \omega)\right]\hat{f}_{\mathrm{L}}(\omega) - \hat{h}_c(\pi - \omega)\hat{f}_{\mathrm{H}}(\pi - \omega) \tag{6}$$

Inspecting the first term, we see the L band is effectively antialiased by the filter:

$$\hat{h}_{\mathrm{aa}}(\omega) = 1 - \hat{h}_t(\pi - \omega)\hat{h}_c(\pi - \omega) \tag{7}$$

The second term expresses the leakage from the H band, and is negligible when h_c is constrained to have zero response outside of the deadzone, which has width 2Δ.

If h_t and h_c are designed to have passband width Δ, h_{aa} will correspondingly have stopband width Δ. We shall soon see why this constraint is useful. Substituting (2) and (3) into (6), we take the squares of both sides to arrive at a relationship between the power spectra of the source and L$'$. Cross products between different frequency components are neglected as they are assumed to be uncorrelated. The result is an equation relating the power spectral densities of L$'$ and the source x.

$$\begin{aligned}\Gamma_{\mathrm{L}'}(\omega) = & h_{\mathrm{aa}}^2(\omega)\hat{h}_0^2(\frac{\omega}{2})\Gamma_x(\frac{\omega}{2}) + h_{\mathrm{aa}}^2(\omega)\hat{h}_0^2(\pi - \frac{\omega}{2})\Gamma_x(\pi - \frac{\omega}{2}) \\ & + \hat{h}_c^2(\pi - \omega)\hat{h}_1^2(\frac{\pi + \omega}{2})\Gamma_x(\frac{\pi + \omega}{2}) \\ & + \hat{h}_c^2(\pi - \omega)\hat{h}_1^2(\frac{\pi - \omega}{2})\Gamma_x(\frac{\pi - \omega}{2})\end{aligned} \tag{8}$$

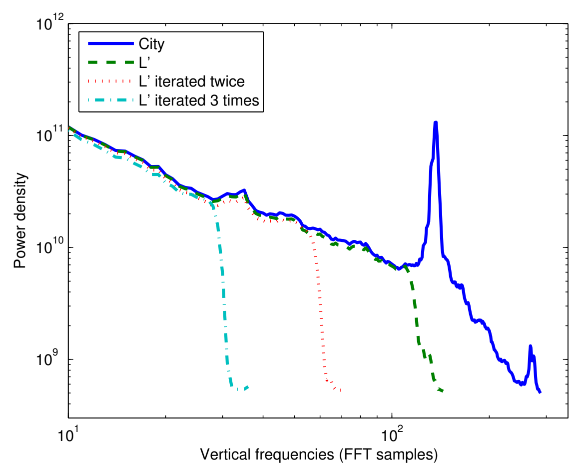

Fig. 3. Iterations of low pass analysis with frequency modulation

To demonstrate the impact of our proposed transform, we apply the above relations to a vertical power spectral density estimate, taken from the MPEG 4CIF video sequence "City". The 9/7 transform is used for the base DWT analysis filters, and suitable filters are designed for the modulated lifting steps, based on $\Delta = \frac{\pi}{8}$.[1] We

[1] $h_t[n]$ = -0.0149, 0.0102, 0.0162, 0.0217, 0.0220, 0.0136, -0.0037, -0.0263, -0.0470, -0.0573, -0.0502, -0.0228, 0.0223, 0.0767, 0.1293, 0.1684, 0.1862, 0.1802, 0.1538, 0.1150, 0.0735, 0.0455
$h_c[n] = h_t[-n]$

use (8) to compute the spectrum for each successive L$'$ band in an iterative tree-based decomposition. The L$'$ bands are not shown. By forcing a stopband constraint of width Δ when designing h_{aa}, the L$'$ band is left with the same relative deadzone width that we assumed for the original source. The method is thus repeatable over successive DWT levels, leaving each successively lower L$'$ band with a similar spectral rolloff, as illustrated in the logarithmic plot of Figure 3.

It is worth noting that the modified lifting structure no longer corresponds to a wavelet transform. The presence of multiplications by $(-1)^n$ yield 4 instead of 2 distinct basis functions for a single stage of the transform. Furthermore, unlike wavelets, indefinite iteration of the multi-resolution transform proposed here does not generally converge to an underlying continuous function.

(a) DWT analysis (b) FM lifting

Fig. 4. Visual comparison of aliasing

From a visual standpoint, the benefits of our proposed transform are clearly evidenced by Figure 4(b). This figure shows a cropped portion of the LL band produced after analysing City with the modulated lifting approach, separably extended to two dimensions. Regular analysis with the 9/7 DWT kernel produces strong aliasing on the skyscraper, which is noticeable even in a still image.

3. FILTER DESIGN GUIDELINES

Our derivation in the previous section yielded an expression for the effective antialiasing filter (7). Designing h_{aa} is essentially the same as designing the end-to-end filter $h_{\text{ee}} = h_t * h_c$. Following the discussion in the preceding section, we want $\hat{h}_{\text{aa}}(\omega)$ to be a high-pass filter with stopband width Δ; equivalently, $\hat{h}_{\text{ee}}(\omega)$ should be a low-pass filter with passband width Δ. We have already noted that since h_c acts on the H band, its stopband must extend for all frequencies outside the deadzone - which is of width 2Δ. This means h_c and correspondingly h_{ee} may have transition bands of width Δ.

In this paper, we adopt an equi-ripple design strategy, based on the well-known method of Parks and McClellan. To simplify matters, we begin by considering filters h_t and h_c with identical magnitude responses. Let a and b denote the maximum passband and stopband ripples, respectively, associated with $\hat{h}_{\text{aa}}(\omega)$. Similarly, let c and d denote the maximum passband and stopband ripples associated with $\hat{h}_t(\omega)$. These are depicted in Figure 5. Stopband ripple b determines how much content remains in the L$'$ deadzone after FM lifting. This must be negligible if iteration on further DWT levels is desired. Stopband ripple d controls how much energy from H leaks into L, potentially causing further aliasing if poorly specified. Currently, we take the conservative approach of setting

$$b = d \tag{9}$$

not choosing to exploit the fact that the power spectrum throughout the H band is typically smaller than in the L band.

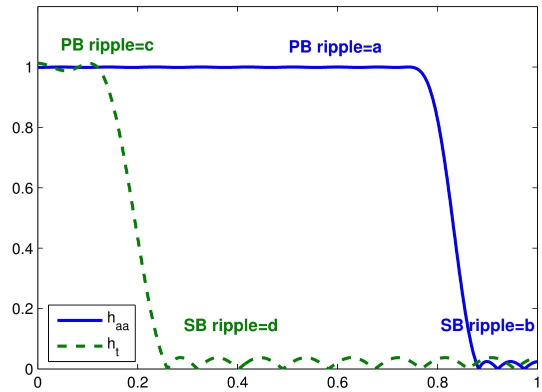

Fig. 5. Ripple specifications for h_{aa} and h_f

Since the passband of h_{aa} is equivalent to the stopband of h_{ee} and $|\hat{h}_{\text{ee}}(\omega)| = |\hat{h}_t^2(\omega)|$, we have the relationship $a = d^2$. Combining this with (9), we arrive at ripple constraints for h_{ee}: the passband is constrained to a ripple of $\pm b$, while the stopband ripple is constrained to the interval b^2. Our objective, then, is to design the end-to-end filter $\hat{h}_{\text{ee}}(\omega)$ with passband width Δ and ripple b, an unconstrained transition band of width Δ and stopband with width $\pi - 2\Delta$ and ripple b^2. h_t and h_c are obtained by a suitable factorisation of h_{ee}.

Note that h_{aa}, and correspondingly h_{ee} must be linear phase so that edges are preserved in L$'$; however, there is no such restriction on h_t and h_c separately. In fact, since the video frames are of finite size, recursive IIR filters could even be used, with a causal h_t applied to the modulated L band in one direction (e.g., left-right or top-bottom) and an anti-causal h_c applied to the H band in the other direction (e.g., right-left or bottom-top). For FIR filter factorisations, we can easily choose h_t and h_c as the minimum and maximum phase factors of the symmetric filter h_{ee}.

The above analysis, based on equal magnitude responses for h_t and h_c, yields a simple design procedure of minimum and maximum phase factorisations of the designed h_{ee}. From a coding efficiency perspective however, it is desirable to transfer as little energy from the L band to the H band as possible, suggesting that the stopband constraints on h_t may be more important than those on h_c. Preliminary results have shown improved compression efficiency for heuristic factorisations, in which the factors for h_t are selected to provide faster rolloff into the stopband than for h_c. We thus have a two-step design procedure, with antialiasing performance driving the design of h_{ee} and coding efficiency driving the factorisation of h_{ee} into h_t and h_c. Of course, a joint design may be capable of achieving a better trade-off between these two objectives.

4. COMPRESSION EFFICIENCY

Like all frequency transforms, the DWT generates coding gain by dividing the spectrum into bands whose relative energies vary. Natural images typically have spectra greatest in magnitude at dc and fall away smoothly with frequency, so the L band invariably captures much greater energy than the H band. However, the proposed modulated lifting transform transfers energy between the bands such that the resulting L$'$ has less energy than L, and H$'$ has greater energy than H. Thus, there is a decrease in the difference in energy between the two bands, and a corresponding penalty to the compression efficiency. This penalty scales with increases in Δ.

The aliasing migrated to H$'$ has a spectral density of high magnitude, within a deadzone of negligible energy. Conversely, the antialiased region in L$'$ is significantly attenuated relative to the rest of the band. Both spectral borders are highly discontinuous due to the artificial nature of their construction. If the bands can be further divided at these bordering regions, compression efficiency can be restored.

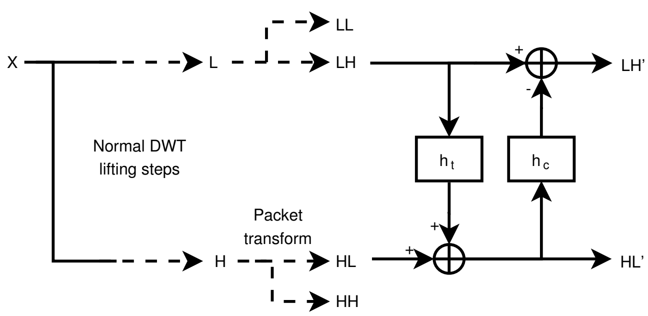

Fig. 6. Packet lifting structure

Although only able to divide bands at dyadic fractions, packet transforms are found to significantly improve the coding efficiency. Additionally, if packet transforms are used we may fold the frequency modulated lifting deeper into the structure, as shown in Figure 6. There are a number of advantages gained from this alternative structure:

- The additional analysis places both the aliasing in LH and the deadzone of HL at the dc frequencies, allowing the removal of the frequency reversal operators. Because of their shift variance, the frequency reversal operators undesirably double the number of synthesis basis vectors for a particular stage.
- Filter design constraints are relaxed, as the passbands and transition bands of h_t and h_c are doubled to 2Δ. The extra DWT analysis is leveraged to reduce the order of the frequency modulating LPFs.

PSNR results are shown in Figure 7 for compression of City. Curves for the frequency modulated lifting method with Mallat decomposition, and the packet lifting method[2] with packet transform are shown. Both have been modulated for 1 level (the LL1 level is antialiased) while the remaining levels are normal DWT analysis. PSNR curves for the standard 9/7 transform are shown for comparison. The 9/7 was used in both the standard Mallat decomposition and the same packet transform as used for the packet lifting scheme. At 1 bit/pel rate the difference between packet lifting and the corresponding DWT comparison is 1.4dB. This penalty is due to the highly discontinuous spectrum created by the aliasing transfer. The packet transform cannot completely negate this because the discontinuities do not line up with the packet transform band divisions.

[2] $h_t[n]$ = -0.0037, 0.0285, 0.0235, -0.0063, -0.0616, -0.0684, 0.0275, 0.1898, 0.3088, 0.3034, 0.1924, 0.0799

$h_c[n]$ = 0.1597, 0.3849, 0.4470, 0.2327, -0.0674, -0.1777, -0.0695, 0.0544, 0.0570, -0.0074

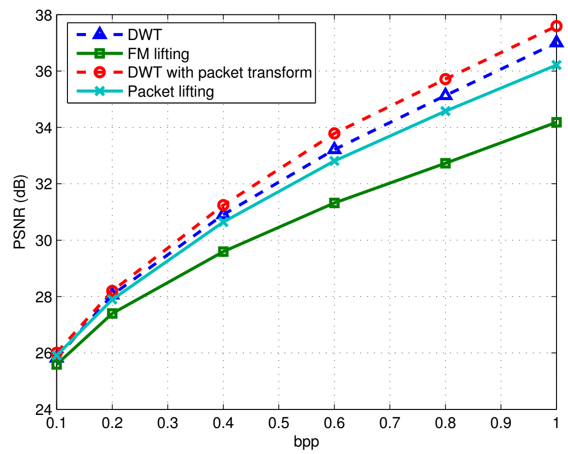

Fig. 7. PSNR results for FM lifting and packet lifting schemes

5. CONCLUSION

We have presented a novel modification to the DWT lifting structure which suppresses aliasing at reduced resolutions, by transferring frequency components between subbands. This allows us to overcome the constraint of equation (1), which applies to all true wavelet transforms. The aggressiveness of the antialiasing behaviour is controlled by a parameter Δ. The proposed scheme is reversible and iterable to multiple DWT levels. Currently, the use of our proposed transform imposes a significant compression penalty, with respect to the regular DWT. In ongoing work, we are pursuing promising directions for further reducing this penalty.

6. REFERENCES

[1] T. Nguyen and M. Li, “Optimal wavelet filter design in scalable video coding,” *Proc. IEEE Int. Conf. Image Proc.*, vol. 1, pp. 473–476, 2005.

[2] V. Bottreau, C. Guillemot, R. Ansari, and E. Francois, “Coding of moving pictures and audio,” *ISO/IEC JTC1/SC29/WG11*, 2004.

[3] G. C. K. Abhayaratne, “Reducing aliasing in wavelets based downsampling for improved resolution scalability,” *Proc. IEEE Int. Conf. Image Proc.*, vol. 2, pp. 898–901, 2005.

[4] H. Schwarz, T. Hinz, D. Marpe, and T. Wiegand, “Constrained inter-layer prediction for single-loop decoding in spatial scalability,” *Proc. IEEE Int. Conf. Image Proc.*, vol. 2, pp. 870–873, 2005.

[5] D. Santa-Cruz, J. Reichel, and F. Ziliani, “Opening the laplacian pyramid for video coding,” *Proc. IEEE Int. Conf. Image Proc.*, vol. 3, pp. 672–675, 2005.