NON-SEPARABLE WAVELET-LIKE LIFTING STRUCTURE FOR IMAGE AND VIDEO COMPRESSION WITH ALIASING SUPPRESSION

J. Gan and D. Taubman

The University of New South Wales

ABSTRACT

Wavelet-based approaches to resolution scalable video are significantly hindered due to the phenomenon of aliasing at reduced spatial resolutions. Improvements to the wavelet filters themselves have limited success, because they are fundamentally constrained by the perfect reconstruction conditions. We propose a packet lifting structure that is capable of perfect reconstruction, yet achieves superior antialiasing performance to improved wavelet designs. While the aliasing removal process incurs some loss of overall compression performance, this is offset by increased compressibility of the lower resolution information. Preliminary results suggest performance competitive with pyramid schemes for achieving similar goals.

Index Terms— antialiasing, scalable video, packet lifting

1. INTRODUCTION

In the domain of image compression, wavelet transforms have proved a valuable tool in the creation of efficient representations which also have useful spatial scalability attributes. Reduced resolution images in this case are obtained simply by discarding the higher frequency subbands and performing a partial wavelet synthesis. This property is exploited by the JPEG2000 image compression standard, and is also appealing for other applications such as video compression. Unfortunately, wavelet subbands contain aliasing which cannot be removed during partial reconstruction, without reference to the discarded high frequency subbands. These aliasing artefacts are typically tolerated in still imaging applications, but become much more noticeable in video, where the lack of shift invariance adversely interacts with motion.

As the standard 5/3 spline and 9/7 Daubechies wavelets used in images have rather poor rolloff, one direction has been in designing longer filters with faster rolloff in the low-pass filter ([1],[2]). Another approach, proposed in [3], adaptively adjusts the low-pass wavelet filter upon the detection of edges. However, these approaches are fundamentally limited, since any pair of wavelet analysis filters, h_0 and h_1, must obey the perfect reconstruction equation:

$$\hat{h}_0(\omega)\hat{h}_1(\pi-\omega)+\hat{h}_0(\pi-\omega)\hat{h}_1(\omega)=\kappa$$

where the constant κ depends on normalisation conventions. At $\omega=\frac{\pi}{2}$ this necessitates that the response of the low-pass filter cannot be zero; additionally, even modest attenuation will produce excessive gain in the complementary high-pass filter. The tight link between h_0 and h_1 near $\frac{\pi}{2}$ renders it impossible to design a low-pass wavelet filter with good aliasing suppression, without incurring a large penalty in compression performance.

Another approach is to avoid using wavelet transforms for resolution scalability altogether. In particular, the Laplacian pyramid structure imposes no constraints on the selection of the low-pass downsampling filter. [4] proposes a novel modification of the classical Laplacian pyramid which largely orthogonalises the quantisation noise contributed from the coarse and difference signals. Even with this modification, however, pyramids have a natural disadvantage in compression efficiency because they are oversampled.

Our approach in this paper is to start with a conventional wavelet transform and employ additional lifting steps to move offending aliased frequency content between subbands. More specifically, we employ a packet wavelet structure, in which high-pass subbands are further decomposed at least one level. A sequence of three lifting steps is then used to exchange frequency content between pairs of subbands with non-contiguous band edges. We recently introduced a closely related scheme in [5]. In the present work, however, we focus on a non-separable lifting structure and add a third lifting step. We also provide a more comprehensive comparison with other related approaches.

In Section 2, we present a one-dimensional inter-packet lifting structure with only two lifting steps and discuss how it transfers aliased content between the low-pass (L) subband and the high-pass (H) subband. In Section 3, we extend this structure to a non-separable 2-dimensional transform, so as to limit complexity and confine frequency swapping to only those components which are relevant to the aliasing problem. In Section 4 we show how introducing a third lifting step can improve the orthogonality of the synthesis vectors. Section 5 outlines our design procedures for producing good packet lifting filters. We estimate computational complexity in Section 6. Section 7 shows visual results at half resolution, as well as compression performance for half and full resolution frames.

2. INTER-PACKET LIFTING

A key observation behind our approach is that video sequences typically have little energy present in the spatial frequencies near Nyquist. This property certainly appears in video that has been processed digitally by interpolation and/or downsampling filters. The property can also be observed in original video content produced by good video cameras, since these contain optical anti-aliasing elements such as bi-refringent filters in their lens assembly. In the extreme case, we can think of these processes as creating a "deadzone" in the vicinity of the Nyquist frequency. Regardless of whether a deadzone exists or not, natural scenes exhibit a rolloff in energy toward higher frequencies. As an example, a clear deadzone can be observed in the horizontal frequency spectrum of various 4CIF MPEG test sequences. The vertical spectra of these sequences, however, are more typical of natural scene rolloff.

The basic idea behind packet lifting is illustrated in Fig. 1, for one dimension. The range of frequencies that will contribute to aliasing when the image is reconstructed at half resolution may be found

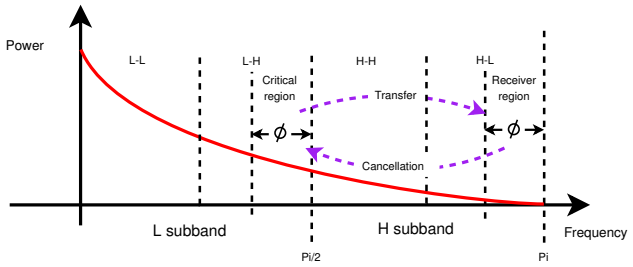


Fig. 1. Spectral representation of packet lifting. Notation for the packet subbands corresponds to the series of low and high-pass wavelet analyses used to produce that band.

in the low frequency content of the L-H subband[1]. We call this the "critical region" and denote its width by ϕ. Of course, ϕ depends upon the underlying wavelet filters; for our work with the 9/7 transform, a value of $\phi = \frac{\pi}{8}$ appears to be most suitable. The objective of inter-packet lifting is to exchange content from the critical region with a so-called "receiver region" in the highest frequency subband, H-L. As explained above, we expect these to have comparatively little energy.

A basic lifting structure for the exchange of content between two bands in a one dimensional transform is shown in Fig. 2. The first lifting step low-pass filters the L-H subband and adds the result to H-L. We call this the "transfer" step since its purpose is to transfer aliased content from the critical region to the receiver region. Note that the frequency reversal associated with high pass subbands means that both the critical region and the receiver region are located in the vicinity of DC within their respective subbands. The second step low-pass filters H-L to isolate the aliased content now found in the receiver region, subtracting this from L-H. We call this the "cancellation" step.

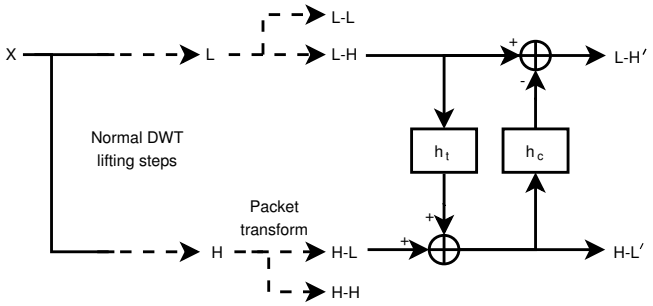


Fig. 2. 1-dimensional packet lifting analysis

Reconstruction is accomplished by simply reversing the lifting steps. On synthesis, we first reverse the cancellation step by filtering H-L′ and *adding* this to L-H′; we then filter L-H and subtract this from H-L′, recovering our original wavelet subbands.

When recovering the signal at half resolution by partial wavelet synthesis, H-L and H-H are discarded, while L-L and L-H are synthesised. In the absence of our proposed lifting steps, the resulting signal will simply be the low-pass subband L produced by a single level of wavelet analysis. The proposed transfer and cancellation lifting steps remove most of the aliased frequency content from L-H, so that it is also removed from the synthesised L subband.

The transfer and cancellation lifting steps can be expressed in the frequency domain as

$$\begin{aligned}\hat{f}_{\text{H-L}'}(\omega) &= \hat{f}_{\text{H-L}}(\omega) + \hat{h}_t(\omega)\hat{f}_{\text{L-H}}(\omega) \\ \hat{f}_{\text{L-H}'}(\omega) &= \hat{f}_{\text{L-H}}(\omega) - \hat{h}_c(\omega)\hat{f}_{\text{H-L}'}(\omega) \\ &= \left[1 - \hat{h}_t(\omega)\hat{h}_c(\omega)\right] \hat{f}_{\text{L-H}}(\omega) - \hat{h}_c(\omega)\hat{f}_{\text{H-L}}(\omega) \qquad (1)\end{aligned}$$

which shows that the effective antialiasing filter operating on L-H is $1-\hat{h}_t(\omega)\hat{h}_c(\omega)$. We should take care to ensure that $1-\hat{h}_t(\omega)\hat{h}_c(\omega)$ is close to 0 at high frequencies, since these must be left unchanged by the lifting process to avoid the creation of additional aliasing artefacts when L-L and the modified subband L-H′ are synthesised into a modified L′ subband. Accordingly, the *end-to-end* filter $h_{ee} = h_t * h_c$ should have good passband and stopband properties; this is covered in greater depth in Section 5.

The second term of equation (1) shows that spectral content is moved into L-H from the H-L band. It would appear that this adds yet more aliasing. In an ideal "deadzone" scenario this contribution has zero energy. In a non-ideal situation we are still better off. Due to the characteristic rolloff in natural video, the energy at near-Nyquist frequencies is much lower than that in the critical region near $\frac{\pi}{2}$. This is borne out in our experiments with the 4CIF MPEG test sequence, "City." This sequence does not have a forced deadzone in the vertical spectrum. After processing with packet lifting, the LL_1 subband has substantially reduced aliasing in original problem areas, while "swapped in" aliasing is comparatively unnoticeable.

It is worth noting that the aliasing cancellation of packet lifting creates an artifically attenuated region of frequencies in the L′ subband. If we repeat the packet lifting procedure on this low resolution information, we now have a deadzone which is roughly twice the width of the new critical region produced at that level. Thus, this method extends well to multi-resolution transforms, in which each successively lower resolution component is substantially antialiased.

3. EXTENSION TO TWO DIMENSIONS

We have introduced packet lifting in one dimension; we now turn our attention to 2-dimensional signals. It is trivial to implement the extension as a separable transform; however, separable application of the 1-dimensional transform in each direction results in unnecessary processing of frequency regions which do not contribute to aliasing in the reduced resolution image, indicated by the shaded regions in Fig. 3. Bi-directional arrows indicate the restricted application of our 1-dimensional packet lifting scheme, in the horizontal or vertical direction as appropriate. In this scheme, no information is

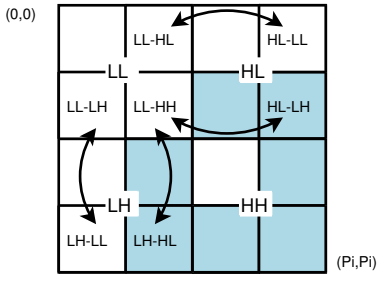


Fig. 3. Non-separable packet lifting on 2-dimensional signals

[1]Due to frequency reversal in high-pass subbands, the low frequency content in L-H actually corresponds to the higher frequencies, in the vicinity of $\frac{\pi}{2}$, in the original image.

exchanged with the HH subband. Evidently, the order of the lifting steps is important because the LL-HH subband is processed twice.

4. ADDING A THIRD LIFTING STEP

The 2 step packet lifting structure yields overall transforms whose synthesis vectors are substantially non-orthogonal. To understand the problem, consider the 1-dimensional reconstruction process illustrated in Fig. 4, ignoring the h_p filter, which we introduce shortly. In the figure, N_1 and N_2 represent quantisation noise processes. We assume that the noise processes are uncorrelated, which should at least be valid at high bit-rates. The cancellation step is inverted first, leaving the L-H subband with the noise process

$$\hat{N}_1'(\omega) = \hat{N}_1(\omega) + \hat{h}_c(\omega)\hat{N}_2(\omega)$$

The transfer step then leaves the H-L subband with the noise process

$$N_2' = -\hat{h}_t(\omega)\hat{N}_1(\omega) + \left(1 - \hat{h}_t(\omega)\hat{h}_c(\omega)\right)\hat{N}_2(\omega)$$

Over the range of frequencies which we are hoping to swap, $1-\hat{h}_t(\omega)\hat{h}_c(\omega)$ should be close to 0 while $\hat{h}_c(\omega)$ should be close to 1, so that the N_2 spectrum is not duplicated and its power is roughly preserved by the subsequent DWT packet synthesis operations[2]. By contrast, the N_1 spectrum appears in both the L-H and H-L subbands. Assuming a white spectrum, the power of this quantisation noise process is amplified by

$$1 + \frac{1}{2\pi}\int_{-\pi}^{\pi}\left|\hat{h}_t(\omega)\right|^2 d\omega.$$

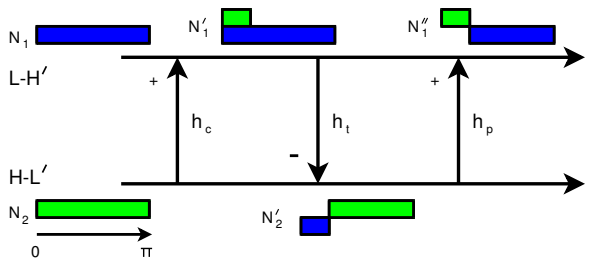


Fig. 4. Lifting of quantisation noise during synthesis

To address this problem, we add the "precompensating" lifting step, $\hat{h}_p(\omega)$, shown in Fig. 4. During analysis, this step is applied first, which means that it is the last to be inverted during synthesis. This leaves the L-H subband with the noise process

$$\begin{aligned}\hat{N}_1''(\omega) &= \left(1 - \hat{h}_p(\omega)\hat{h}_t(\omega)\right)\hat{N}_1(\omega) \\ &\quad + \left(\hat{h}_c(\omega) + \hat{h}_p(\omega)\left(1 - \hat{h}_t(\omega)\hat{h}_c(\omega)\right)\right)\hat{N}_2(\omega)\end{aligned} \quad (2)$$

If all the filters are good low-pass filters, $1 - \hat{h}_p(\omega)\hat{h}_t(\omega)$ will be close to 0 over the critical region, so that the duplicate copy of $\hat{N}_1(\omega)$ is cancelled. In fact, with ideal low-pass filters, one can verify that the original quantisation noise processes are precisely swapped in the critical region, as shown in Fig. 4.

Another interpretation of the precompensating lifting step is that it reduces the correlation between the transform's underlying synthesis basis vectors. Table 1 compares the correlation coefficients with and without this third step, for the filters used in Section 7.

[2]The 9/7 DWT transform we use here is close to orthonormal and hence nearly energy preserving.

Table 1. Correlation coefficients without, and with precompensation

	2-step lifting	3-step lifting
r(L-L, L-H)	-0.0278	0.0274
r(L-L, H-L)	-0.0523	-0.0592
r(L-L, H-H)	-0.0184	-0.0184
r(L-H, H-L)	0.3306	-0.2043
r(L-H, H-H)	0.0726	0.0727
r(H-L, H-H)	-0.0749	-0.0628

5. FILTER DESIGN

The design of the packet lifting filters is influenced by the width of the critical region, ϕ. As noted earlier, we find $\phi = \pi/8$ to be a suitable choice for use with the 9/7 DWT. Equation (1) shows that the effective antialiasing filter is equivalent to $1 - \hat{h}_{ee}(\omega)$, where the end-to-end filter $\hat{h}_{ee}(\omega) = \hat{h}_t(\omega)\hat{h}_c(\omega)$. Rather than designing h_t and h_c separately, we take the approach of designing the composite filter directly. h_{ee} must achieve aliasing cancellation for frequencies in the interval $(\pi/2 - \phi, \pi/2]$ in the original image spectrum. Because it operates on the L-H packet subband, the passband region for the end-to-end filter is actually $[0, 4\phi)$, which is $[0, \pi/2)$ for our choice of $\phi = \pi/8$. In the stopband, we must be careful to avoid significantly modifying any of the frequency components near the L-L band edge. This sensitive information is required to avoid aliasing when synthesising the low-pass signal L$'$. The width of this sensitive region is again related to the rolloff of the wavelet filters, but one level of analysis deeper than those which established our original critical bandwidth of ϕ. Accordingly, the sensitive region for h_{ee} has a width of 2ϕ in the L-H subband, corresponding to the interval $(\pi - 2\phi, \pi] = (3\pi/4, \pi]$.

We presently use the method of Parks and McClellan to design equiripple filters for h_{ee}. Note that the component filters h_t and h_c should also have reasonable passbands and stopbands, but only h_{ee} is required to be linear phase, since it operates directly on the low resolution signal content. Because of this, an easy factorisation strategy constrains h_{ee} to have strictly non-negative response, then factorises it into minimum-phase and maximum-phase components, h_t and h_c, having equal magnitude responses.

The spectral factorisation strategy would perform well if the filters were equally important; however, the spectral energy they are exchanging is far from symmetrical. In general, the largest adverse impact on the energy compaction and hence compression performance of the overall transform, arises from the transfer of high energy content from L-H to H-L via h_t. This implies that h_t should be designed to fall away faster than h_c, suggesting an unequal factorisation of h_{ee}.

Our design procedure for h_t and h_c is as follows. We initially create a component h_z that will be assigned entirely to h_t. In particular, we currently set $h_z(z) = \frac{1}{4}z + \frac{1}{2} + \frac{1}{4}z^{-1}$, having two zeros at the Nyquist frequency. We then use the Parks and McClellan method to design a complementary linear phase filter h_{qq}, with a strictly non-negative equiripple stopband and a passband objective precompensated for the fact that $\hat{h}_{ee}(\omega) = \hat{h}_{qq}(\omega)\hat{h}_z(\omega)$. After factorising $\hat{h}_{qq}(\omega)$ into minimum and maximum phase components, $\hat{h}_q(\omega)$ and $\hat{h}_q^*(\omega)$, we assign $\hat{h}_t(\omega) = \hat{h}_z(\omega)\hat{h}_q(\omega)$ and $\hat{h}_c(\omega) = \hat{h}_q^*(\omega)$.

As for the precompensation filter h_p, one reasonable candidate would be to re-use h_c. Instead, we design h_p directly to minimise the power of the noise process N_1'', described by equation (2). If the transform's synthesis vectors were truly orthonormal, a rate-distortion optimised compressor would assign quantisation pa-

rameters so as to equalise the quantisation noise powers in all subbands. Noting that this is almost exactly what happens when good pre-compensation is employed, our objective is to minimise

$$\int_{-\pi}^{\pi} \left[\left| 1 - \hat{h}_p(\omega)\hat{h}_t(\omega) \right|^2 + \left| \hat{h}_c(\omega) + \hat{h}_p(\omega) \left(1 - \hat{h}_{ee}(\omega) \right) \right|^2 \right] d\omega$$

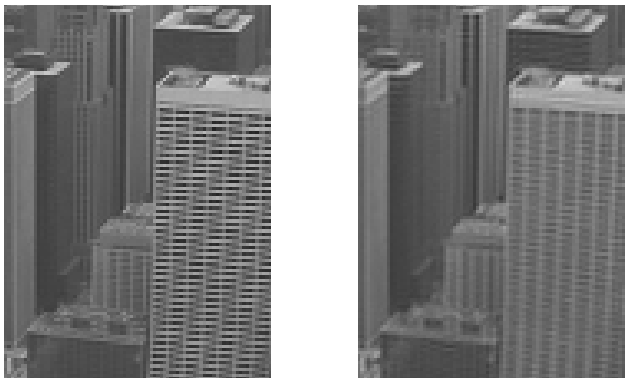
For fixed h_t and h_c, this reduces to a simple least squares optimisation problem. Assigning h_p the same region of support as h_c, we find that the resulting precompensation filters always outperform h_c in terms of compression performance, although the impact can be quite small.

6. COMPUTATIONAL COMPLEXITY

The computational complexity of packet lifting with the modifications of Sections 3 and 4 incorporated is low. When using the reduced packet lifting scheme over one dimension, we lift between 2 pairs of packet subbands, with each packet having $\frac{1}{16}^{\text{th}}$ the samples of the full image. Between each pair of packets we apply 3 filters in turn, each filtering across the samples contained in a single packet. If the number of taps in the filters are L_{h_p}, L_{h_t} and L_{h_c}, then the total number of multiplications in the 2-dimensional scheme is $\frac{\left(\mathrm{L}_{h_p}+\mathrm{L}_{h_t}+\mathrm{L}_{h_c}\right)}{4}$ per LL_0 sample. For our chosen filters with 12, 10 and 10 taps this works out to 8 multiplications per sample.

7. RESULTS

We conducted our experiments on the 4CIF MPEG test sequence City. We use filters[3] designed as described in Section 5. Fig. 5 shows a comparison of half-resolution images. We compared the scheme



(a) DWT analysis (b) Packet lifting

Fig. 5. Antialiasing performance vs regular dwt at half resolution

against ordinary DWT compression, both with baseline Mallat decomposition and the same packet transform used in packet lifting. We have also compared with the Laplacian pyramid with update (LPU) scheme of [4]. The coarse and difference components are compressed by the same wavelet coder using regular Mallat decomposition. Both our packet lifting scheme and the comparison LPU scheme are conducted for a single level of antialiasing.

We compressed both the full resolution and half resolution images. Fig. 6 shows a coding penalty is observed relative to the DWT. Our packet lifting scheme is approximately 1.5dB below the DWT with packet transform at a rate of 1 bit per pixel. The LPU is approximately 1.4dB below the corresponding regular DWT. The packet lifting scheme has larger gain in performance at half resolution because additional subband decomposition in the lower resolution successfully isolates the cancelled regions of low energy.

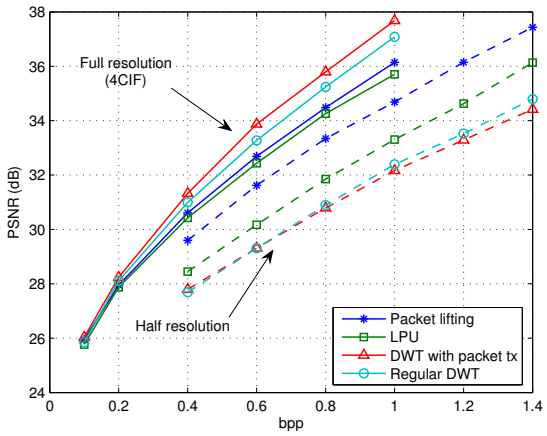


Fig. 6. PSNR comparisons computed from 10-frame MSE averages

8. CONCLUSION

We have proposed a lifting structure modification for antialiasing by transferring spectral content between non-contiguous packet subbands. We also present a design strategy for the filters within the lifting steps. The transform is computationally economical, and has reasonable compression performance compared with other resolution scalability methods.

9. 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] 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.

[5] J. Gan and D. Taubman, "Antialiasing scalable video with a modulated lifting structure," *Proc. IEEE Int. Conf. Image Proc.*, vol. 1, pp. 997–1000, 2007.

[3] $h_t = 0.0184, 0.0035, -0.0592, -0.0281, 0.0583, 0.0140, -0.1108, -0.0629, 0.2140, 0.4367, 0.3506, 0.1080$
$h_c = 0.4323, 0.5377, 0.2387, -0.1590, -0.1721, 0.0600, 0.1082, -0.0433, -0.1337, 0.0738$
$h_p = 0.3553, 0.4797, 0.2313, -0.0725, -0.1602, -0.0240, 0.0851, 0.0154, -0.0385, -0.1177$