OPTIMISED DELTA LOW-PASS FILTER FOR UMCTF

Nikola Šprljan, Marta Mrak and Ebroul Izquierdo

Multimedia and Vision Research Group, Dept. of Electronic Engineering
Queen Mary, University of London, London E1 4NS, United Kingdom
Email: {nikola.sprljan, marta.mrak, ebroul.izquierdo}@elec.qmul.ac.uk

Keywords: wavelet based scalable video coding, wavelet transform

Abstract

This paper provides a performance analysis of the delta low-pass filterbank used in Unconstrained Motion Compensated Temporal Filtering (UMCTF) of video frames. Any wavelet filterbank for which a two-step (one prediction and one update) lifting implementation exists has a delta low-pass filter variant, where the modification consists of not performing the update step. Benefits of using such filters are better temporal scalability performance and lower decoding delay. Here it is argued that in that case the normalisation factor for the high-pass subband needs to be carefully selected in order to improve the compression performance. For the case of Haar wavelet with delta low-pass filter the expression for the Riesz bounds is derived and the measure of the expected difference of energies in signal and transform domain is introduced. Then, the normalisation factor is optimised in respect to the minimal Riesz bounds ratio and minimal expected difference. The application of optimised filterbank with delta low-pass in a scalable video environment shows enhanced decoding performance on a wide range of bit-rates and lower temporal resolutions.

1 Introduction

Research on scalable video coding has resulted in efficient frameworks [1, 2] using combinations of spatio-temporal transform techniques and embedded entropy coding. All these frameworks consist of Motion Compensated Temporal Filtering (MCTF) and spatial wavelet transform (2D-DWT), producing a set of spatio-temporal subbands. The multiresolution structure resulting from the temporal and spatial wavelet transforms naturally leads to temporal and spatial resolution scalability. This is followed by a 2D or 3D embedded coding scheme enabling a fine granular quality scalability on all processed spatial and temporal resolutions. While the conventional hybrid coding systems (*e.g.* MPEG-4) use the closed-loop schemes, which imply prediction based on previously quantised samples, the coding systems employing MCTF are open-loop, where the subbands are independently encoded in a progressive fashion. When used in a quality scalability setting, the main drawback of the conventional closed-loop schemes becomes the frame-to-frame dependent quantisation structure, as the optimal bit allocation problem can become too complex.

In the early work on 3D wavelet-based scalable video coding, the wavelet transform in the temporal direction has been performed either without [3] or with [4] motion-compensated prediction. Although these 3D wavelet based methods did not yield high coding gains compared to the algorithms that use inter-frame motion-compensated prediction techniques, they provided a framework for highly scalable video coding. In the subsequent improvement of MCTF, the motion compensation was incorporated into the temporal wavelet transform, thus improving both coding gain and visual quality. By interleaving the temporal lifting steps with spatial interpolation in the MCTF, it is possible to perform compensation of any precision while maintaining the prefect reconstruction property [1]. A highly flexible variant of MCTF was introduced in [5], called Unconstrained MCTF (UMCTF). In a contrast to MCTF, in UMCTF the selection of the temporal filter is adaptive, and depends on the video contents and transmission delay requirements. For example, for a highly correlated sequences with low motion activity a filtering with longer wavelets can yield increased coding gain. However, the motion compensation process cannot always capture the motion correctly, e.g. in the cases of complex and irregular motion, or in presence of fast scene changes. In those cases a poor prediction is obtained and the overall compression is of lower efficiency. This also introduces visually unpleasant artifacts in the low-pass filtered frames, as some areas contain the result of filtering over poorly matched areas in neighbouring frames. These artifacts are relevant for temporal scalability, where only the temporal low-pass frames on a required level of temporal decomposition are decoded. When dyadic tem-

This research was supported by the European Commission under contract FP6-001765 aceMedia.

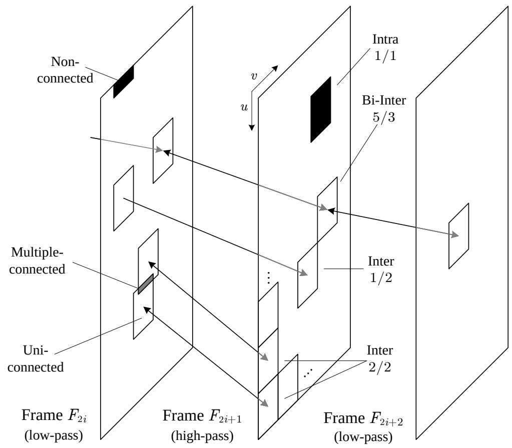

Fig. 1. Frame areas and corresponding filters

poral decomposition is used, this enables extraction and decoding of a range of frame rates that are in the powers of two relation. Then ideally the lower frame rate sequence should be a downsampled original with temporal aliasing suppressed, what is hardly achievable when the motion is incorrectly captured. Several approaches exist for reducing such artifacts, which can be classified accordingly to the applied temporal filter length. Regarding the employed filterbank, motion-compensated areas of video frames can be classified into three groups:

- using filterbank with non-delta low-pass, for instance the Haar $2/2$ filterbank or the LeGall $5/3$ filterbank;
- using filterbank with delta low-pass, which is effectively $1/g$ for filterbank with high-pass filter of $g > 1$ taps. This is equivalent to disabling the update step. The delta low-pass filterbanks are beneficial for avoiding artifacts related to the temporal averaging, *e.g.*, temporal blurring and mismatched frame areas, and also for lowering the decoding delay;
- introduction of intra areas in the high-pass frames, which is effectively lazy wavelet or $1/1$ "filterbank". Intra areas are utilised because some areas in the reference frame cannot be matched in the neighbouring frames, due to occlusions or fast motion. These areas are then declared as isolated and are not temporally filtered.

These cases are depicted on Fig. 1 where different blocks of frame F_i are compensated using various filters. These three filters can be used globally for the whole frames, or can be chosen adaptively according to the content, where the optimal one of the three is chosen for each frame area. In this paper the focus is on the second case, for which the optimised normalisation coefficients for the equivalent filterbank is derived, with respect to the desired wavelet properties and decoding distortion.

2 Delta low-pass wavelet filterbank

If a signal is split into two parts, x_k and y_k, comprising even and odd samples, respectively, one can write the general lifting steps for a filterbank with maximum filter length of 2 as:

$$\begin{aligned} y'_k &= y_k + c_x \cdot x_k \\ x'_k &= x_k + c_y \cdot y'_k \\ x''_k &= K_x \cdot x'_k \\ y''_k &= K_y \cdot y'_k, \end{aligned} \tag{1}$$

where the coefficients c_y and c_x determine the filter support and type, while K_y and K_x are the normalisation factors. The signals x'' and y'' are the wavelet transform coefficients, where the type of subband they represent depends on the choice of c_x and c_y. For the orthonormal Haar wavelet, where y'' are the high-pass coefficients and coefficients x'' are the low-pass:

$$\{K_x, K_y, c_x, c_y\} = \left\{\sqrt{2}, \frac{1}{\sqrt{2}}, -1, \frac{1}{2}\right\}. \tag{2}$$

Then the first lifting step in (1) is the prediction step, and second is the update step. The coefficient c_y for the Haar wavelet with delta low-pass filter equals zero, implying that the update step is omitted. Obviously, that filterbank is not orthogonal, since generally $E_{\mathcal{T}} \neq E$, where $E_{\mathcal{T}}$ is the energy in the wavelet (transform) domain, and E the actual signal energy. For the normalisation factor K_x the same value as in the Haar wavelet is chosen, so as to ensure the same DC gain for the case of temporal filtering with adaptive choice of filter. However, since only the ratio between normalisation factors K_x/K_y determine the performance of the quantisation, K_x can be set to any constant and optimisation performed with respect to K_y only.

For the filterbank with delta low-pass filter $c_y = 0$ is set. With this the orthogonality of the filterbank is lost, *e.g.*, generally $E_{\mathcal{T}} \neq E$, where $E_{\mathcal{T}}$ is the energy in the wavelet (transform) domain, and E the signal energy. The same normalisation factor K_x as in Haar wavelet is used, so that the low-pass temporal frame is globally normalised with the same factor in a case of adaptive temporal filtering. Also, only in that case when $x_k = y_k$ also $E_{\mathcal{T}} = E$, meaning that the energy of DC component is preserved. However, since only the ratio between normalisation factors K_x/K_y determine the performance of quantisation, K_x can be set to any constant and perform optimisation with respect to K_y only.

In terms of temporal filtering the samples of the signal as analysed here relate to pixels of consecutive frames. Here odd samples belong to the compensated frames, while even samples represent pixels from reference frames, spatially shifted relatively to the pixels in odd frames as defined by motion vectors. In a case of sub-pixel accurate motion vectors, the interpolation is applied on reference frame prior to the prediction step. As it considerably complicates the analysis, in the following it was not taken into account. Also, the reported experiments were performed with integer precision motion estimation only.

3 Analysis

Orthogonality of a filterbank can be defined as how close the observed filterbank is to being orthogonal. Orthogonality has been recognised as a crucial property of the wavelet regarding its compression performance [6]. Several measures for orthogonality have been proposed, each of them tailored to a particular application. In [6] it is argued that the orthogonality measures based on energy preservation are the most relevant for applications related to signal approximation and compression. In this work the same approach is adopted. This can be interpreted in a way that the energy of a wavelet coefficient needs to be as close as possible to the energy it represents in the signal domain. Due to the linearity of the transform the same holds for the error signals, thus the orthogonality is decisive for estimation of the signal domain error that is done in the rate-distortion algorithm of the encoder. The variation of the difference between energies in transform and signal domain can be expressed with the Riesz bounds [6]. The Riesz bounds are defined as two values A and B, for which:

$$A \cdot E_{\mathcal{T}} \leq E \leq B \cdot E_{\mathcal{T}}, \tag{3}$$

so that for the orthonormal filterbank $A = B = 1$, while the ratio of the bounds for non-orthogonal filterbanks is $B/A > 1$. If employed as a measure of orthogonality the task of optimising the wavelet coefficients becomes the task of minimising its B/A ratio. To simplify the notation in the following analysis, it is defined $x = x_k$, $y = y_k$, $x_{\mathcal{T}} = {x'}_k$, $y_{\mathcal{T}} = {y'}_k$ and $\eta = K_y$, so that (1) and the energy relations for the delta low-pass filterbank can be written as:

$$\begin{aligned} E &= x^2 + y^2 \\ y_{\mathcal{T}} &= \eta \cdot (y - x), \\ x_{\mathcal{T}} &= \sqrt{2}x, \\ E_{\mathcal{T}} &= x_{\mathcal{T}}^2 + y_{\mathcal{T}}^2 = 2x^2 + \eta^2 \cdot (y - x)^2. \end{aligned} \tag{4}$$

Basically, the goal here is to determine whether $\eta_{\mathrm{opt}} = \eta_{\mathrm{Haar}}$, where η_{Haar} denotes the high-pass normalisation coefficient as used for the Haar wavelet, $\eta_{\mathrm{Haar}} = 1/\sqrt{2}$, and with η_{opt} the optimal high-pass normalisation factor for the delta low-pass filterbank. Temporally adjacent pixels, x and y, can be regarded as observation pairs of zero-mean random variables in the range $x, y \in [-M/2, M/2]$. The optimisation problem now can be formulated as:

$$\eta_{\mathrm{opt}}^{\mathrm{R}} = \arg\min_{\eta} \rho = \arg\min_{\eta} \frac{\max_{x,y} \frac{E_{\mathcal{T}}}{E}}{\min_{x,y} \frac{E_{\mathcal{T}}}{E}}, \tag{5}$$

where $\eta_{\mathrm{opt}}^{\mathrm{R}}$ is the optimal normalisation factor for the high-pass subband, with respect to the Riesz bounds ratio. By observing that the ratio $E_{\mathcal{T}}/E$ depends on the x/y ratio only, and by converting to the polar coordinates $(x, y) \rightarrow (r, \theta)$, one can derive:

$$\theta_{\min} = \arg\min_{\theta} E_{\mathcal{T}}/E,$$
$$\theta_{\max} = \arg\max_{\theta} E_{\mathcal{T}}/E.$$

Using this, the Riesz bounds in (5) become:

$$\rho = \frac{\max_{\theta} \frac{E_{\mathcal{T}}}{E}}{\min_{\theta} \frac{E_{\mathcal{T}}}{E}} = \frac{1 + \eta^2 + \frac{1+\eta^4}{\sqrt{1+\eta^4}}}{1 + \eta^2 - \frac{1+\eta^4}{\sqrt{1+\eta^4}}}. \tag{6}$$

Differentiating (6) w.r.t η for solving (5) leads to $\eta_{\mathrm{opt}}^{\mathrm{R}} = 1$, and $B/A = 3 + 2\sqrt{2}$.

Observe that this is the optimal solution in the sense of reducing the maximum range of wavelet domain energy variation. Therefore a measure for the expected energy difference is introduced here:

$$\bar{E}_{\Delta} = E\left[\left|\gamma^2 \cdot E - E_{\mathcal{T}}\right|\right],$$

where $\gamma = K_x \cdot K_y$ is the total scaling factor of the filterbank. Now the task of minimising the averaged energy difference is $\eta_{\mathrm{opt}}^{\mathrm{E}} = \arg\min_{\eta} \bar{E}_{\Delta}$. Using (4), $\bar{E}_{\Delta}$ can be written as:

$$\bar{E}_{\Delta} = \iint_{-M/2}^{M/2} \left|(x + y)^2 \eta^2 - 2x^2\right| p(x, y) dx dy, \tag{7}$$

where $p(x, y)$ is the joint probability distribution of x and y. If it is assumed that x and y are uncorrelated, *i.e.*, $p(x, y) = 1/M^2$, then $\eta_{\mathrm{opt}}^{\mathrm{E}} = \eta_{\mathrm{Haar}} = 1/\sqrt{2}$. It can be also shown that the same result is obtained for the case where $p(x, y)$ is a bivariate normal distribution, which is more realistic if x and y are correlated since they represent motion aligned pixels from neighbouring frames in video. It can be also verified that for $\{c_x, c_y\} = \left\{-1, \frac{1}{2}\right\}$ this method gives $K_x = 1/K_y = \sqrt{2}$, *i.e.*, it yields the Haar wavelet. Interestingly, for lazy wavelet, where $\{c_x, c_y\} = \{0, 0\}$, it gives $K_x = K_y$, *i.e.*, the logical result of same normalisation factor for both subbands.

Lastly, the distribution of the quantisation energy in the reconstructed signal domain is considered. The wavelet coefficients are quantised with quantisation step q, which if is small enough ensures that quantisation error signals in low-pass and high-pass subbands are practically uncorrelated. In tha case, it can be written:

$$\begin{aligned} E_{q_x} &= q^2/24 \\ E_{q_y} &= E_{q_x} + q^2/12\eta^2, \end{aligned}$$

where E_{q_x} and E_{q_y} are quantisation error energies in the reconstructed even and odd frames, respectively. Error energies are directly related to any additive measure of distortion that measures overall distortion of the sequence, here the mean square error (MSE) is used. It can be seen that $E_{q_x} < E_{q_y}$, and that by increasing η a more uniform distribution of error energy between frames can be achieved. If the error energy is expressed as a function of η, it can be written:

$$\begin{aligned} E_{q_x}(\eta_2) &= E_{q_x}(\eta_1) + \Delta E_{q_x} \\ E_{q_y}(\eta_2) &= E_{q_y}(\eta_1) - \Delta E_{q_y}, \end{aligned}$$

Table 1. Decoding results for coding with different normalisation factors η and Haar ("Foreman", CIF, 30 fps, first 8 frames); PSNR_Y [dB]

bit rate	128 kpbs	512 kpbs	768 kpbs	1 Mbps	8 Mbps
η_{Haar}	26.93	32.91	33.69	34.00	45.68
$\eta_{\mathrm{opt}}^{\mathrm{R}}$	26.95	32.90	33.67	34.07	46.76
$\eta_{\mathrm{opt}}^{\mathrm{q}}$	26.96	32.93	33.68	34.25	46.85
Haar	27.01	33.23	33.97	34.30	46.56

where ΔE_{q_x} and ΔE_{q_y} are the positive values corresponding to the energy difference caused by using η_2 instead of η_1. It can be seen that η_2 can be increased as long as $\Delta E_{q_y} - \Delta E_{q_x} > 0$ so that improvement in the overall MSE can be achieved. Generally, the value of $\eta_{\mathrm{opt}}^{\mathrm{q}}$, defined as the optimal η with respect to the minimisation of reconstruction distortion, will depend on the rate-distortion curve of a particular sequence. For various sequences on a large range of bit rates $\eta = \sqrt{2}$ was found to perform well, so $\eta_{\mathrm{opt}}^{\mathrm{q}} = \sqrt{2}$ was used in the experiments.

4 Results

The rate-distortion performance of delta low-pass filterbank have been tested in a scalable video codec [7]. The evaluation has been carried out for the application of η_{Haar}, $\eta_{\mathrm{opt}}^{\mathrm{R}}$ and $\eta_{\mathrm{opt}}^{\mathrm{q}}$ as factors for normalisation of the high-pass subband. For a comparison, the results are obtained also for the application of Haar filter. The results, presented in Table 1, have been conducted for the MPEG test sequences "Foreman" of CIF (352×288) resolution and 30 fps frame rate. Due to the different distribution of quantisation error for different η, the PSNR can vary dramatically between the reconstructed even and odd frames, *e.g.*, it can vary up to 5 dB for high bit rates and low η. For this reason the overall PSNR has been computed from the average MSE of the sequence, and not by averaging the PSNR, which is not an additive measure. This provides a more correct estimation of the average error in the sequence when the variance of error energy across the frames is large. The application of Haar wavelet outperforms the delta low-pass filterbank on all bit rates. However, the introduction of the optimised normalisation factor in respect to the error energy difference $\eta_{\mathrm{opt}}^{\mathrm{q}}$ improves the performance of delta low-pass filter on almost all bit rates when compared to other η. The results for temporal scalability (30 fps $\rightarrow$ 15 fps) are presented in Table 2. The delta low-pass filterbanks here significantly outperform Haar wavelet on all bit rates. As the reference sequence is given by subsampling of the original one, the PSNR performance of the Haar wavelet saturates at high bit rates and is around 40 dB, while for the delta low-pass filterbanks a near-lossless performance can be measured.

Table 2. Decoding results for temporal scalability 30 fps $\rightarrow$ 15 fps; PSNR_Y [dB]

bit rate	128 kpbs	384 kpbs	640 kpbs	1 Mbps	4 Mbps
delta lp.	27.30	32.19	34.52	36.27	48.13
Haar	27.15	31.65	33.59	34.76	39.48

5 Conclusion

In this paper the optimal normalisation coefficients for delta low-pass filterbank have been derived for the application in wavelet-based video coding. Two approaches have been analysed: reduction of the range of wavelet domain energy variation and the minimisation of the variation of error energy between low-pass and high-pass frames, resulting in normalisation coefficients $\eta_{\mathrm{opt}}^{\mathrm{R}} = 1$ and $\eta_{\mathrm{opt}}^{\mathrm{q}} = \sqrt{2}$, respectively. It has been shown that both cases introduce gain in scalable video coding. In the future work, performance using rate-distortion optimisation between high-pass and low-pass temporal subbands will be analysed, what can be regarded as normalisation with varying η. Also the performance of the optimised delta low-pass filterbanks in content adaptive UMCTF will be investigated.

References

[1] A. Secker and D. Taubman, "Lifting-based invertible motion adaptive transform (LIMAT) framework for highly scalable video compression," *IEEE Trans. Image Processing*, vol. 12, no. 12, pp. 1530–1542, Dec. 2003.

[2] Peisong Chen and John W. Woods, "Bidirectional MC-EZBC with lifting implementation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 10, pp. 1183–1194, Oct. 2004.

[3] D. Taubman and A. Zakhor, "Multirate 3D subband coding of video," *IEEE Trans. Image Processing*, vol. 3, no. 5, pp. 572–589, Sept. 1994.

[4] S.-J. Choi and J. Woods, "Motion-compensated 3-D subband coding of video," *IEEE Trans. Image Processing*, vol. 8, no. 2, pp. 155–167, Feb. 1999.

[5] Mihaela van der Schaar and Deepak S. Turaga, "Unconstrained motion compensated temporal filtering (UMCTF) framework for wavelet video coding," in *Proc. IEEE Int. Conf. Acoust. Speech and Signal Proc. (ICASSP)*, 2003, vol. 3, pp. 81–84.

[6] F. Moreau de Saint-Martin, P. Siohan, and A. Cohen, "Biorthogonal filterbanks and energy preservation property in image compression," *IEEE Trans. Image Processing*, vol. 8, no. 2, pp. 168–178, Feb. 1999.

[7] N. Šprljan, M. Mrak, G. C. K. Abhayaratne, and E. Izquierdo, "A scalable coding framework for efficient video adaptation," in *Proc. Int. Work. on Image Analysis for Multimedia Interactive Services (WIAMIS)*, Montreux, Switzerland, Apr. 2005.