

FOUR-BAND LINEAR-PHASE ORTHOGONAL SPATIAL FILTER BANK FOR SUBBAND VIDEO CODING

Grégoire Pau and Béatrice Pesquet-Popescu

GET-ENST, Signal and Image Proc. Dept.
46 Rue Barrault, 75634 Paris, France
e-mail: {gpau,pesquet}@tsi.enst.fr

ABSTRACT

In wavelet-based scalable video coding schemes, temporal interframe redundancy is exploited by applying a temporal wavelet transform along the motion trajectories. A spatial decomposition of the temporal subbands is further performed to take advantage of the spatial redundancy of the filtered frames. However, most of the $t + 2D$ video codecs do not take into account the very different spatial characteristics of the temporal subband frames and use indifferently the same spatial 9/7 biorthogonal transform to decompose them. In this paper, we present a spatial transform based on a four-band filter bank, whose frequency selectivity properties are shown to be more suited to represent detail frames. We give the analytical form of a linear-phase, orthogonal and regular four-band filter bank and we show by experimental results conducted on video sequences that significant improvements in terms of PSNR can be obtained using the proposed filter bank to decompose the detail frames.

1. INTRODUCTION

Subband motion-compensated temporal filtering (MCTF) video codecs have attracted recently [1, 2] a lot of attention, due to their high compression performance comparable to state-of-the-art hybrid codecs and due to their scalability features. The spatio-temporal subband scheme ($t + 2D$) exploits the temporal interframe redundancy by applying a temporal wavelet transform along the motion trajectories on the frames of a video sequence. A spatial decomposition of the temporal subbands is then done to take advantage of the spatial redundancy of the filtered frames and the resulting wavelet coefficients can be encoded by different algorithms such as 3D-SPIHT [3], 3D-ESCOT [4] or MC-EZBC [1].

Approximation frames result from the low-pass temporal filtering of video frames and look very similar to natural images, with large piecewise smooth areas. Detail frames result from the high-pass temporal filtering and do not exhibit a natural image behavior. Sharp edges resulting from the temporal misprediction of moving areas and high-frequency textures are predominant.

Most of the MCTF-based video codecs do not take into account the very different spatial characteristics of the temporal approximation and detail frames and use the popular 9/7 biorthogonal wavelet transform to spatially decompose them, independently of their type. The spatial 9/7 transform has been shown to perform well in the case of natural images [5] and therefore should be appropriate for approximation frames. However, it may not be so well-suited to represent detail frames which have a significant amount of intermediate and high frequencies. Since detail frames usually constitutes a major portion of the temporal subband frames, an effective and parsimonious spatial representation of these frames is highly desirable.

Some previous works [6] considered the use of wavelet packets to decompose displacement frame differences (DFD) which are close to detail frames, but the computational complexity of wavelet packet best basis algorithms is high, the corresponding filter banks are not always very selective and the global results were not very satisfactory.

We present in this paper a spatial transform based on M -band filter banks, whose frequency selectivity properties are more suited to represent detail frames. We derive analytically from the general design framework proposed by Alkin and Caglar [7] a linear-phase, orthogonal and regular 4-band filter bank. We show by experimental results conducted on video sequences that significant improvements in terms of PSNR can be obtained using the proposed 4-band filter bank to decompose the detail frames.

The paper is organized as follows: in the next section, we study the characteristics of temporal subband frames. In Section 3 we review the M -band filter bank and in Section 4 we design an optimal filter bank for transforming the detail frames. Section 5 illustrates by simulation results the coding performance of the proposed spatial decomposition. We conclude in Section 6.

2. CHARACTERISTICS OF SUBBAND FRAMES

In order to study the spatial characteristics of the temporal subband frames, we compute their averaged 2D power spec-

trum repartition. Tab. 1 and 2 show this spatial power spectrum repartition on approximation and detail frames respectively, computed on a 4-level motion-compensated temporal decomposition of the video sequence “Mobile”.

f_h/f_v	$[0, \frac{1}{8}[$	$[\frac{1}{8}, \frac{1}{4}[$	$[\frac{1}{4}, \frac{3}{8}[$	$[\frac{3}{8}, \frac{1}{2}]$
$[0, \frac{1}{8}[$	97.99%	0.40%	0.19%	0.05%
$[\frac{1}{8}, \frac{1}{4}[$	0.66%	0.23%	0.12%	0.02%
$[\frac{1}{4}, \frac{3}{8}[$	0.18%	0.07%	0.04%	0.01%
$[\frac{3}{8}, \frac{1}{2}]$	0.03%	0.01%	0%	0%

Table 1. Average 2D power spectrum repartition of the *approximation* frames on a 4-level temporal decomposition of the video sequence “Mobile” with respect to horizontal (f_h) and vertical (f_v) frequency range.

f_h/f_v	$[0, \frac{1}{8}[$	$[\frac{1}{8}, \frac{1}{4}[$	$[\frac{1}{4}, \frac{3}{8}[$	$[\frac{3}{8}, \frac{1}{2}]$
$[0, \frac{1}{8}[$	18.80%	10%	7.13%	3.27%
$[\frac{1}{8}, \frac{1}{4}[$	14.91%	9.20%	6.65%	2.50%
$[\frac{1}{4}, \frac{3}{8}[$	9.71%	6.09%	4.12%	1.21%
$[\frac{3}{8}, \frac{1}{2}]$	3.30%	1.75%	1.05%	0.29%

Table 2. Average 2D power spectrum repartition of the 4^{th} level *detail* frames on a 4-level temporal decomposition of the video sequence “Mobile” with respect to horizontal (f_h) and vertical (f_v) frequency range.

These tables clearly show that for the detail frames, in contrast with the approximation frames, the spectral information is far from being mainly localized in low-frequencies. An example of a detail frame is given in Fig. 2 where the sharp edges and high-frequency textures can be observed. These observations motivate us to investigate the use of an M -band filter bank whose frequency selectivity properties are more suited to decompose the detail frames.

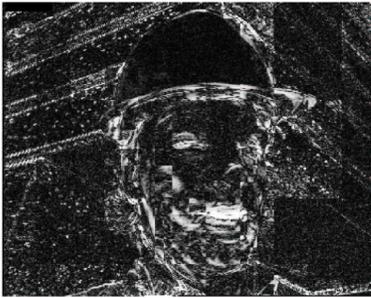


Fig. 1. Contrast-enhanced 2^{nd} level detail frame issued from the temporal decomposition of the video sequence “Foreman”.

3. M -BAND FILTERS

An M -band filter bank is an extension of the dyadic filter bank: an input signal x is decomposed into M decimated

bands with a set of analysis filters with impulse responses $\{h_k\}_{0 \leq k < M}$. The reconstructed signal \tilde{x} is obtained by summing the interpolated subband signals, filtered by the synthesis filters with impulse responses $\{\tilde{h}_k\}_{0 \leq k < M}$. Fig. 2 illustrates an M -band filter bank.

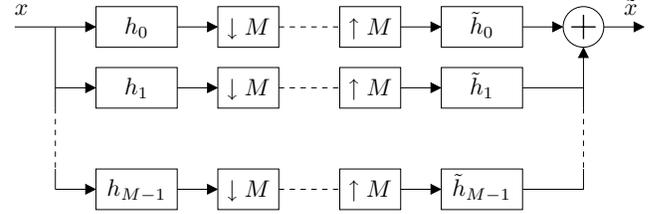


Fig. 2. M -band filter bank structure.

By choosing $\tilde{h}_k(n) = h_k^*(-n)$, it can be shown that the perfect reconstruction property $x = \tilde{x}$ is equivalent to the following time-domain orthogonality condition:

$$\forall k \forall i, j \sum_n h_i(n) h_j^*(n - Mk) = \delta_k \delta_{i-j} \quad (1)$$

A general design technique to produce a set of linear-phase and orthogonal filters $\{h_k\}$ which fulfills Eq. (1) has been given in [7]. The filters are of length L , where L must be a multiple of $2M$. Given a suitable low-pass h_0 filter, the remaining $M - 1$ filters are obtained through the use of shuffling operators on the low-pass filter h_0 , which must be symmetric and must satisfy:

$$\forall k \sum_n h_0(n) h_0^*(n - Mk) = \delta_k \quad (2)$$

The M -band filter bank is thus entirely characterized by its low-pass prototype filter h_0 . Moreover, due to the symmetry requested for h_0 and due to Eq. (2), there are $L/2 - L/M$ degrees of freedom to choose h_0 . For the case $M = 4$ and given a low-pass filter h_0 , the design technique determines the remaining $M - 1$ filters:

$$\begin{aligned} h_1(n) &= (-1)^{\lfloor n/2 \rfloor} h_0(n + (-1)^n) \\ h_2(n) &= (-1)^n h_1(n) \\ h_3(n) &= (-1)^n h_0(n). \end{aligned} \quad (3)$$

4. PROTOTYPE H_0 FILTER DESIGN

We consider here designing $M = 4$ band filter banks because this is the most simple non-dyadic case as M must be a power of two in this framework.

We first start with a $L = 8$ tap orthogonal linear-phase filter bank. According to the last section, there is 2 degrees of freedom to choose h_0 . We decide to design explicitly the filter bank by using these degrees of freedom to maximize the number of vanishing moments of the underlying

wavelets ψ_1, ψ_2 and ψ_3 , respectively associated to the filters h_1, h_2 and h_3 , in order to improve the regularity of the filter bank.

On the contrary, in the case $M = 4$, Alkin et al. used a numerical constrained optimization technique to maximize the coding gain of the filter bank to choose h_0 . The so-obtained coefficients can be found in Tab. 3.

Firstly, h_0 must be symmetric w.r.t. $-1/2$ and fulfills Eq. (2); this can be expressed as:

$$\begin{cases} h_0(0)^2 + h_0(1)^2 + h_0(2)^2 + h_0(3)^2 = 1/2 \\ h_0(0)h_0(3) + h_0(1)h_0(2) = 0. \end{cases}$$

The wavelet ψ_i associated to the filter h_i has p vanishing moments if and only if the z -transform of the filter $H_i(z)$ has a zero at 1 with a multiplicity of p . Thus, due to the anti-symmetry of h_1 and h_3 , ψ_1 and ψ_3 have at least one vanishing moment. We impose ψ_2 to have at least one vanishing moment. In the design method proposed in [7], the condition is not strictly satisfied, and this causes the filter bank to produce non-zero coefficients in case of a constant signal, with annoying effects in the context of an embedded zero-tree coder.

We impose an equivalent but explicit condition, leading to more accurate coefficients, which is to find h_2 such that $H_2(1) = 0$. This yields a new equality constraint on h_0 :

$$h_0(0) + h_0(3) = h_0(1) + h_0(2).$$

We also impose ψ_3 to have at least two vanishing moments. This is equivalent of finding h_3 such the derivative $H'_3(1) = 0$, leading to $\sum_k k h_3(k) = 0$. Due to the anti-symmetry of h_3 , this leads to $\sum_{k \geq 0} (2k + 1)h_3(k) = 0$, which gives a new equality constraint:

$$-7h_0(3) + 5h_0(2) - 3h_0(1) + h_0(0) = 0$$

This leads to the following system:

$$\begin{cases} h_0(0)^2 + h_0(1)^2 + h_0(2)^2 + h_0(3)^2 = 1/2 \\ h_0(0)h_0(3) + h_0(1)h_0(2) = 0 \\ h_0(0) + h_0(3) - h_0(1) - h_0(2) = 0 \\ -7h_0(3) + 5h_0(2) - 3h_0(1) + h_0(0) = 0 \end{cases}$$

which yields:

$$\begin{cases} h_0(0) = (5 + \sqrt{15})/16 \\ h_0(1) = (3 + \sqrt{15})/16 \\ h_0(2) = (5 - \sqrt{15})/16 \\ h_0(3) = (3 - \sqrt{15})/16. \end{cases}$$

The obtained low-pass filter h_0 coefficients, referred as FB1 can be compared to Alkin et al's ones for the case $M = 4$ and $L = 8$ in Tab. 3. Complete filters coefficients of the 4-band 8-tap FB1 filter bank can be found using the shuffling operations stated in Eqs. (3). Fig. 3 shows the frequency

responses of the filters $\{h_k\}_{0 \leq k < 4}$ where we can observe the frequency selectivity of the filter bank.

It is worth noting that due to the symmetry of h_2 w.r.t. $-1/2$, it can be shown that $H_2(1) = 0$ implies $H'_2(1) = 0$ and therefore ψ_2 has two vanishing moments. Moreover, due to the anti-symmetry of h_3 w.r.t. $-1/2$, $H'_3(1) = 0$ and this causes ψ_3 to have three vanishing moments.

As a conclusion, the proposed 4-band 8-tap FB1 filter bank strictly satisfies by design the following properties: perfect reconstruction, linearity of the phase, orthogonality, ψ_1 has one vanishing moment, ψ_2 has two vanishing moments and ψ_3 has three vanishing moments.

n	$h_0(n)$ Alkin	$h_0(n)$ FB1	$h_0(n)$ FB2
0	0.567030813	0.554561459	0.591506350
1	0.406151488	0.429561459	0.158493649
2	0.094517754	0.070438540	0.341506350
3	-0.067700953	-0.054561459	-0.091506350

Table 3. Low-pass h_0 filter coefficient comparison.

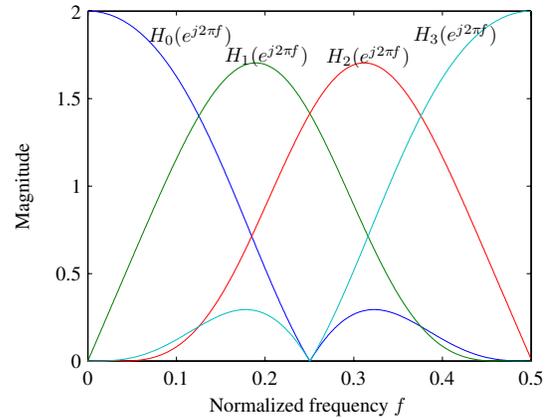


Fig. 3. Frequency responses of the filters $\{h_k\}_{0 \leq k < 4}$ of the 4-band 8-tap FB1 filter bank.

Note that it is also possible to set different constraints on the number of vanishing moments of the related wavelet functions. For example, having three vanishing moments on ψ_1 , two vanishing moments on ψ_2 and one vanishing moment on ψ_3 leads to:

$$\begin{cases} h_0(0) = (3 + \sqrt{3})/8 \\ h_0(1) = (3 - \sqrt{3})/8 \\ h_0(2) = (1 + \sqrt{3})/8 \\ h_0(3) = (1 - \sqrt{3})/8. \end{cases}$$

The resulting coefficients are reported in Tab. 3 and the corresponding filter bank is referred in the sequel by FB2. The frequency responses of the filter bank FB2 plotted in Fig. 4 exhibit strong secondary lobes, leading to poor frequency selectivity.

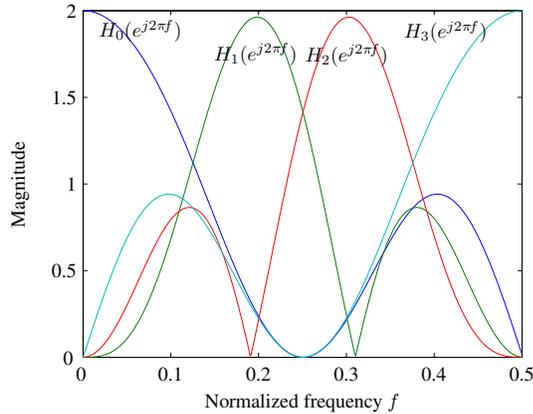


Fig. 4. Frequency responses of the filters $\{h_k\}_{0 \leq k < 4}$ of the 4-band 8-tap FB2 filter bank.

5. EXPERIMENTAL RESULTS

In our simulations, we consider two representative test color video sequences in CIF format at 30 fps: “Mobile” and “City”, which have been selected for their very different motion and texture characteristics. These video sequences have been decomposed over five temporal levels with the uniform motion-compensated 5/3 temporal filter [8]. The simulations were conducted with the MC-EZBC codec.

Temporal *approximation* subbands are decomposed over 5 levels with biorthogonal 9/7 wavelets. Temporal approximation and detail subbands coefficients are encoded using the MC-EZBC algorithm, which has been adapted to work with 4-band 2D signals.

We compare in Tabs. 4 and 5 the coding performance achieved at several bitrates with the proposed FB1 and FB2 filter banks, with the 9/7 and the 5/3 biorthogonal filter banks, used for decomposing the detail frames over 2 spatial levels.

YSNR (in dB)	384 kbs	512 kbs	768 kbs	1024 kbs
4-band FB1	27.93	29.95	32.16	33.65
4-band FB2	27.80	29.70	31.70	33.05
9/7	27.71	29.76	32.00	33.47
5/3	27.40	29.52	31.60	33.05

Table 4. Rate-distortion comparison of three spatial filters on “Mobile” CIF video sequence (30fps).

We observe that the use of the 4-band 8-tap FB1 filter bank for decomposing detail frames leads to better performance of almost 0.2 dB at all bitrates compared to the dyadic biorthogonal 9/7 filter bank and more than 0.5 dB compared to the 5/3 filter bank. We also observe that the FB2 filter bank does not perform as well as the FB1 does. This could be expected due to the frequency responses less selective than the ones of FB1.

YSNR (in dB)	384 kbs	512 kbs	768 kbs	1024 kbs
4-band FB1	33.61	35.45	37.65	39.26
4-band FB2	33.44	35.10	37.08	38.59
9/7	33.40	35.23	37.47	39.09
5/3	33.16	34.92	37.06	38.73

Table 5. Rate-distortion comparison of three spatial filters on “City” CIF video sequence (30fps).

6. CONCLUSION

We have presented in this paper a spatial transform based on a 4-band filter bank, whose frequency selectivity properties are more suited to represent detail frames. By imposing regularity conditions on the filter bank based on the number of vanishing moments of the underlying wavelets, we have derived analytically a linear-phase, orthogonal and regular 4-band filter bank. Experimental results conducted on video sequences have shown significant improvements by using the FB1 filter bank for decomposing detail frames. Future investigations with a larger number of bands and longer filters will be conducted in order to assess the global efficiency of the proposed approach.

7. REFERENCES

- [1] S.-T. Hsiang, J.W. Woods, and J.-R. Ohm, “Invertible temporal subband/wavelet filter banks with half-pixel-accurate motion compensation,” *IEEE Transactions on Image Processing*, vol. 13, pp. 1018–1028, August 2004.
- [2] A. Secker and D. Taubman, “Highly scalable video compression with scalable motion coding,” *IEEE Transactions on Image Processing*, vol. 13, pp. 1029–1041, August 2004.
- [3] B.-J. Kim, Z. Xiong, and W.A. Pearlman, “Very low bit-rate embedded video coding with 3-D set partitioning in hierarchical trees (3D-SPIHT),” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 8, pp. 1365–1374, 2000.
- [4] S. Li, J. Xu, Z. Xiong and Y.-Q. Zhang, “3D embedded subband coding with optimal truncation (3D-ESCOT),” *Applied and Computational Harmonic Analysis*, vol. 10, pp. 589, May 2001.
- [5] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, “Image coding using wavelet transform,” *IEEE Transactions on Image Processing*, vol. 1, pp. 205–220, 1992.
- [6] P.-Y. Cheng, J. Li, and C.-C. Jay Kuo, “Video coding using embedded wavelet packet transform and motion compensation,” in *Proc. of SPIE Visual Information Processing*, Orlando, FL, April 1996.
- [7] O. Alkin and H. Caglar, “Design of efficient M -band coders with linear-phase and perfect-reconstruction properties,” *IEEE Transactions on Signal Processing*, vol. 43, pp. 1579–1590, 1995.
- [8] G. Pau and B. Pesquet-Popescu, “Uniform motion-compensated 5/3 filterbank for subband video coding,” in *Proc. of IEEE Int. Conf. on Image Processing*, Singapore, October 2004.