14th European Signal Processing Conference (EUSIPCO 2006), Florence, Italy, September 4-8, 2006, copyright by EURASIP

IMAGE CODING WITH RATIONAL SPATIAL SCALABILITY

Grégoire Pau and Béatrice Pesquet-Popescu

Dept. of Image and Signal Processing, GET-Télécom Paris
37-39 rue Dareau, 75014 Paris, France
Phone: +33 145817137, Fax: +33 145817144, email:{gpau,pesquet}@tsi.enst.fr
Web: http://www.tsi.enst.fr/∼{gpau,pesquet}

ABSTRACT

In the digital TV and mobile operators scene, there is an increasing demand for a wider scalability range when delivering content to heterogeneous client devices which have different screen sizes. Subband image coding with classical wavelet transform only provide dyadic scalability factors but we have shown that M-band transforms can offer rational scalability factors with an appropriate modification of the synthesis filters. In this paper, we evaluate the image coding efficiency offered by some M-band transforms combined with the rational scalability feature. We show that they offer an increased compression performance compared to the popular dyadic 9/7 wavelet transform.

1. INTRODUCTION

With the expansion of multimedia applications and the need for delivering compressed bitstreams for heterogeneous devices, scalability has become an important feature for still image and video coders. It is then highly desirable to have an encoding scheme able to provide an *unique* bitstream which can be decoded by a large variety of devices which share different screen sizes.

Subband image coding with classical wavelet transforms inherently provides a dyadic scalability since the low-pass approximation band can be viewed as a half-sized version of the original image. Subsequent decompositions of the approximation band further provide other scalability factors that are powers of two, but no other factors can be achieved.

However, in the digital TV and mobile operators scene, there is an increasing demand for a wider range of scalability when delivering content. As illustrated by recent proposals [1] concerning the rising high definition TV (HDTV) video formats, we can also observe the need for rational scalability factors. As an example, the HDTV 720p and 1080p video formats are linked by a scaling ratio of 2/3; with classical wavelet transforms, it is therefore not possible to create a scalable video stream able to represent the video content in both resolutions. It is also not possible to represent two video formats which have a different aspect ratio, for instance the wide-screen 16/9 and 4/3 video formats.

Classical solutions to circumvent these scalability limitations exist and can be classified into three types: simulcast strategies, layered predictive coding schemes and post-resampling solutions. In simulcast strategies, several versions of the same multimedia content at different resolutions are simultaneously broadcasted. This allows each device to decode the bitstream adapted to its screen size but this is a bandwith-expensive solution. Layered predictive coding schemes reduce the overhead created in simulcast strategies by using predictive mechanisms between each version of the different compressed contents. The main drawback of such schemes [1] is that they can cope only with a fixed set of video resolutions already known at the encoding time, offering then only a coarse scalability. In post-resampling solutions such as [2], the devices decode the image content at full resolution and then apply a resizing operator in order that it fits with their own screen sizes. However, this is a costly solution in terms of complexity as the image has to be fully reconstructed and resampled. It is worth noting that other techniques exist as in [3] where the authors propose an image resizing strategy in the compressed DCT domain using a transcoding module. This is however a computationally expensive solution which needs an external transcoder device. Other approaches based on the image decomposition into basis functions [4] are also able to provide a fine spatial scalability but they still need performance improvement at medium and high coding bitrates.

How can we design an efficient transform which supports extended scalability features ? In the context of M-band filter banks [5], we have presented in [6] how an M-band *synthesis* filter bank can be modified in order to directly reconstruct an output image of a resolution reduced by an arbitrary rational factor M/P, where P is any integer lower than M. This property then brings a scalability of a variable factor M/P to any M-band subband coding scheme, without any change in the analysis filter bank. We have also shown that this approach has a lower computational complexity than the aforementioned post-resampling solution.

We propose in this paper to evaluate the image coding efficiency offered by some M-band transforms combined with the rational scalability feature. More precisely, we provide a quantitative evaluation of the scalabiliy feature in M-band and dyadic structures, from a coding point of view and for different scalability factors.

The paper is organized as follows: in the next section, we review the basic properties of M-band filter banks. In Section 3, we present how an M-band synthesis filter bank can be modified in order to produce an output signal resized by a factor M/P. Section 4 illustrates the coding efficiency of the method and we conclude in Section 5.

2. M-BAND FILTER BANK

An M-band filter bank is an extension of a dyadic filter bank where M is higher than 2. For an M-band filter bank, an input signal x is decomposed into M subbands $\{y_k\}_{0\leq k<M}$ with a set of analysis filters with impulse responses $\{h_k\}_{0\leq k<M}$ followed by decimators of a factor M. The reconstructed signal $\tilde{x}$ is then obtained by summing the M-fold upsampled subband signals, filtered by the synthesis filters with impulse responses $\{g_k\}_{0\leq k<M}$. Fig. 1 illustrates an M-band analysis-

synthesis filter bank structure.

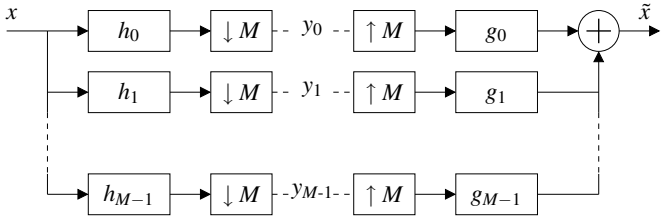

Figure 1: M-band analysis-synthesis filter bank structure.

The filter bank is said to ensure perfect reconstruction (PR) if the original signal can be reconstructed at the decoding side. It can be shown [7] that this property $x = \tilde{x}$ is equivalent to the following time-domain conditions:

$$\forall k, 0 \leq i, j < M, \quad \sum_n h_i(n) g_j^*(n - Mk) = \delta_k \delta_{i-j}$$

Many transforms commonly used in subband image coding belong to the class of M-band PR filter banks such as DCT, lapped orthogonal transform (LOT) [8], lapped biorthogonal transform (LBT) [9] or GenLOT [10]. Alkin and Caglar have also proposed in [11] a framework to design arbitrary M-band PR filter banks where M is a power of two with several degrees of freedom. Dyadic wavelet packet transforms with uniform subband bases also belong to the class of M-band filter banks with $M = 2^L$.

The M-band filter bank typically offers a scalability of factor M, since the low-pass band y_0 can be used at the synthesis side as an M-fold downsampled version of the original signal. We recall in Section 3 how a simple modification of the synthesis filters allows to *directly* reconstruct the signal at any M/P resizing factor.

3. MODIFIED SYNTHESIS FILTER BANK

3.1 Rational Resampling Operator

First, we need to define precisely what we are meaning by rational M/P resizing. Image resizing is classically done by using a resampling operator. A resampling operator by a rational factor M/P is made by cascading a P-fold upsampler $[\uparrow P]$, a resampling filter w and an M-fold downsampler $[\downarrow M]$. It is denoted by $[\downarrow \frac{M}{P}]$ and is illustrated in Fig.2.

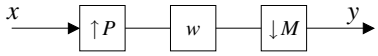

Figure 2: Resampling $[\downarrow \frac{M}{P}]$ operator by a factor M/P.

The choice of the filter w is generally a trade–off between its cut–off frequency and its length, in order to prevent blurring and aliasing artifacts. Linear filters like nearest neighbor, linear, bicubic or Lanczos are commonly used, depending on the application requirements. Other efficient resampling filters designed in the polyphase domain [12] with reduced aliasing effets can also be used. However, the optimal choice of the resampling filter w is beyond the scope of this paper.

A classical approach to obtain reduced images of factor M/P when reconstructing an image with an M-band synthesis filter bank is to simply use a resampling operator after having reconstructed the full image. However, this is a costly solution in terms of complexity as the full image has to be reconstructed and resampled. We show in the next subsection how rational scalability can be obtained with an M-band filter bank.

3.2 Modified Synthesis Filter Bank

We have shown in [6] that, given a perfect reconstruction M-band filter bank, it is sufficient to replace the M-fold upsamplers by P-fold ones and to use M/P resampled synthesis filters $\{\widetilde{g}_k = [\downarrow \frac{M}{P}] g_k\}$ instead of $\{g_k\}$ in order to have a perfect M/P resampled signal at the output, i.e. $\tilde{y} = [\downarrow \frac{M}{P}] x$. This is illustrated in Fig.3.

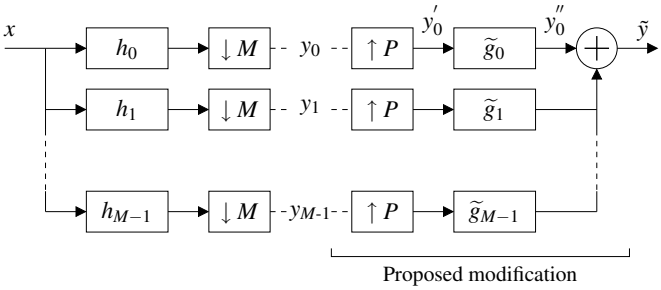

Figure 3: M/P resizing by modification of the synthesis filter bank.

It is clear that this property is able to bring a spatial scalability of a variable factor M/P, for all the integers $P < M$, to any M-band subband coding scheme, without any change in the analysis filter bank.

It is also worth noting that the reduced-resolution signal can be reconstructed using only the first Q subbands $\{y_k\}_{0 \leq k < Q}$ in the original synthesis filter bank, with $Q \neq P$. This flexibility offered by M-band transforms provides a finer granularity over dyadic ones when some subbands can be lost, as in broadcasting applications for instance.

4. EXPERIMENTAL RESULTS

In this section, we want to evaluate the image coding efficiency offered by some M-band transforms combined with the rational scalability feature. Simulations are also conducted for comparison purposes with the popular 9/7 dyadic transform used in the JPEG-2000 standard.

The following 8-band transforms have been considered in our experiments : the Alkin 16-tap [11] filter bank, DCT, LOT, LBT and the 9/7 wavelet packet transform with an uniform subband basis. Test images are decomposed over 3 levels with the dyadic 9/7 transform and over one level with the 8-band transforms. The DCT, LOT and LBT block transforms coefficients are reordered as a regular 8-band wavelet 2D decomposition, as illustrated in Fig. 4 where the basis decomposition structure used for the dyadic 9/7 transform is also presented. The transforms have been integrated in the JPEG-2000 VM 8.0 codec. Malvar's software implementation [13] is used for DCT, LOT and LBT transforms. All the simulations have been conducted with the JPEG-2000 codec default parameters.

In the sequel, M denotes the number of subbands in the transforms and is set to 8. The parameter P controls the M/P reduction factor. Q and Q' denote respectively the number of

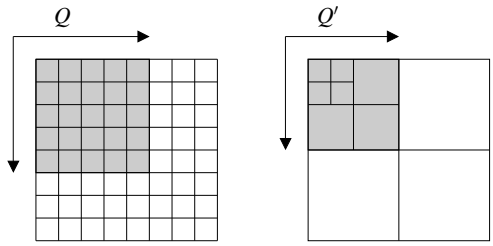

Figure 4: Basis decomposition structures used for 8-band transforms (left) and the dyadic 9/7 transform (right). Q and Q' denote the number of subbands used for the reconstruction.

subbands used for reconstructing the signal in 8-band transforms and in the dyadic transform.

In order to ensure the rational scalability feature, synthesis filters have to be resampled. We have used a linear interpolator which is a good compromise bewteen cut-off frequency, length and complexity. Linear interpolation can be done by choosing a triangle w filter of length $2P+1$, which impulse response is defined by :

$$w[k] = \begin{cases} 1-|k/P|, & \text{if } -P < k < P \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

4.1 Full Reconstruction

We first compare in Tab. 1 and 2 the coding efficiency obtained on Barbara and Lena for the proposed dyadic 9/7 and 8-band transforms with full reconstruction. All the subbands are used ($Q=8$) and images are fully reconstructed ($P=8$).

Bitrate (bpp)	0.1	0.2	0.4	1.0
9/7-dyadic	24.47	27.02	30.61	37.02
Alkin	24.29	26.63	29.79	35.25
LOT	25.00	27.84	31.34	37.23
LBT	**25.27**	**28.17**	**31.76**	**37.70**
DCT	23.92	26.44	29.78	36.00
9/7-packet	**25.27**	28.00	31.38	37.43

Table 1: Rate-distortion PSNR (in dB) comparison of the dyadic 9/7 transform and 8-band transforms on Barbara with full reconstruction ($P=Q=8$).

Bitrate (bpp)	0.1	0.2	0.4	1.0
9/7-dyadic	29.62	32.74	**36.05**	**40.36**
Alkin	28.31	31.09	34.06	38.79
LOT	29.06	32.05	35.27	39.47
LBT	29.48	32.63	35.71	39.82
DCT	28.20	31.29	34.73	39.27
9/7-packet	**29.72**	**32.78**	35.80	39.85

Table 2: Rate-distortion PSNR (in dB) comparison of the dyadic 9/7 transform and 8-band transforms on Lena with full reconstruction ($P=Q=8$).

We observe that the 8-band LBT transform has the best coding efficiency on Barbara at almost all bitrates whereas the dyadic 9/7 transform performs the best on Lena. On both images, it appears that the 8-band LBT and the dyadic 9/7 transform have the best overall coding efficiency. This is confirmed in Fig. 5 where we see that the LBT and 9/7 transforms perform far better than the Alkin or DCT ones, which exhibit severe blocking artefacts. In accordance with [14], this therefore shows that the LBT is an efficient transform, which is competitive against the dyadic 9/7 transform.

Figure 5: Zooms on reconstructed Lena images with the dyadic 9/7 transform and several 8-band transforms at 0.1 bpp with full reconstruction ($P=Q=8$).

4.2 Rational Scalability

Let us now suppose a network where a content server broadcasts a compressed image to heterogeneous client devices which have different screen sizes. Some PDA clients are able to display images of 512 pixel width, most smartphones can only handle 320 or 192 pixels resolutions whereas mobile phones often have screens of 128 pixels wide. This leads to a set of $\{8/8, 8/5, 8/3, 8/2\}$ scalability factors. We want to investigate which solution among the dyadic 9/7 transform and 8-band transforms gives the best coding efficiency at the client side in this scenario. Note that the interest of the method from the computational point of view was extensively studied in [6].

As noticed before, a subband coding scheme with the classical dyadic wavelet transform is not able to support these rational factors. Simulcast is a strategy too bandwith-expensive and the only solution for the low-resolution devices is to reconstruct the full image and then resize it with a post-resampling operator. Client devices are then able to get Q' bands from the bitstream depending of their bandwidth capabilities. Because of the pyramidal structure of dyadic decompositions, Q' must be a power of 2. They can further reconstruct the images at the resolution $8/P$, according to their screen size.

On the contrary, with an 8-band filter bank subband coding scheme and the proposed modification of the synthesis filter bank, it is possible to make a scalable bitstream able to handle these rational scalability factors. Client devices are then able to get Q bands from the bitstream and directly reconstruct the images at the resolution $8/P$, according to their screen size. M-band transforms offer a finer granularity than dyadic ones and Q can be freely choosen between 1 and M.

We first compare in Tab. 3 and 4 the coding efficiency obtained on Barbara and Lena for the 9/7 dyadic transform and 8-band transforms with $P = Q = 5$. This is equivalent to the situation where a client device is only able to get $Q = 5$ bands from the bitstream and has a resolution reduced by a factor $8/5$ compared to the original image size. Due to the dyadic decomposition restrictions, the number of subbands in the 9/7-dyadic case has been lowered to $Q' = 4$.

Bitrate (bpp)	0.1	0.2	0.4	1.0
9/7-dyadic	25.44	26.71	28.10	28.88
Alkin	25.50	27.42	29.08	30.47
LOT	26.18	28.51	30.72	32.43
LBT	**26.51**	**28.79**	**30.90**	**32.48**
DCT	25.10	27.24	29.77	32.18
9/7-packet	25.62	27.19	28.28	28.93

Table 3: Rate-distortion PSNR (in dB) comparison of the dyadic 9/7 transform and 8-band transforms on Barbara reduced by a factor 8/5 ($P = Q = 5$).

Bitrate (bpp)	0.1	0.2	0.4	1.0
9/7-dyadic	30.89	34.23	36.76	38.08
Alkin	29.42	32.35	34.65	36.40
LOT	30.34	33.97	37.45	40.51
LBT	30.72	**34.41**	**38.04**	**41.14**
DCT	29.15	32.82	36.83	40.54
9/7-packet	**31.01**	34.24	36.56	38.18

Table 4: Rate-distortion PSNR (in dB) comparison of the dyadic 9/7 transform and 8-band transforms on Lena reduced by a factor 8/5 ($P = Q = 5$).

We observe that the 8-band LBT transform performs the best on both images and outperforms the 9/7 dyadic transform by up to 3 dB. We also observe that the 9/7 dyadic transform is far below all the other 8-band transforms. This can be explained by the fact that the dyadic pyramidal structure restrictions force the 9/7 dyadic transform to use a number of subbands $Q' = 4$ lower than 8-bands tranforms, where $Q = 5$ bands are used. The observed gain is therefore explained by the lack of flexibility of dyadic decompositions.

Despite this lack of flexibility and in order to compare the dyadic and 8-band transforms with the same number of bands, the simulations have been conducted on Lena with $P = 5$ and $Q = Q' = 4$. The results are presented in Tab. 5.

We observe again the better coding efficiency of the 8-band 9/7-packet and LBT transforms over the dyadic 9/7 decomposition. The gain is however lower than the one observed in Tab. 4. This can be explained by the weak frequency selectivity of the 9/7 filters, already noticed by [15]

Bitrate (bpp)	0.1	0.2	0.4	1.0
9/7-dyadic	30.89	34.23	36.76	38.08
Alkin	29.34	31.88	33.54	34.46
LOT	30.34	33.83	36.52	38.28
LBT	30.72	34.19	**37.03**	**38.60**
DCT	29.14	32.64	35.75	37.54
9/7-packet	**31.01**	**34.24**	36.55	38.09

Table 5: Rate-distortion PSNR (in dB) comparison of the dyadic 9/7 transform and 8-band transforms on Lena reduced by a factor 8/5 ($P = 5,\ Q = 4$).

who have observed that the low-pass Daubechies 9-tap filter has too much energy leakage in stopband. This may give rise to large wavelet coefficients in the high-pass subbands, therefore reducing the overall coding efficiency.

We now compare in Tab. 6 the coding performance achieved for the reconstructed images reduced by a factor $8/3$ with $P = 3$ and $Q = 4$. We observe again that the 8-band transforms outperform the 9/7 dyadic transform. However, the gain is lower than the one observed in the $8/5$ case.

Bitrate (bpp)	0.1	0.2	0.4	1.0
9/7-dyadic	30.88	34.14	36.74	38.09
Alkin	29.01	31.52	33.11	33.98
LOT	30.15	33.55	36.22	37.96
LBT	30.65	34.04	**36.83**	**38.36**
DCT	28.70	32.14	35.29	37.24
9/7-packet	**30.98**	**34.18**	36.53	38.12

Table 6: Rate-distortion PSNR (in dB) comparison of the dyadic 9/7 transform and 8-band transforms on Lena reduced by a factor 8/3 ($P = 3,\ Q = 4$).

Fig. 6 shows the reconstructed image Lena for the case $P = Q = 3$ at a bitrate of 0.4 bpp, with the 9/7 dyadic and 8-band transforms. The results are compared with the reference uncoded Lena image reduced by a factor $8/3$ with a linear interpolator resampler. We clearly observe the better image quality obtained with the Alkin and LBT cases, compared to the dyadic 9/7 one.

Another example is illustrated in Fig. 7 where the image Barbara has been reconstructed at the reduced resolution 8/6 with $Q = 6$ bands. One can notice the strong aliasing artefacts on Barbara's scarf obtained with the 9/7 dyadic transform which are not visible with the 8-band Alkin and LBT transforms.

5. CONCLUSION

In the context of subband image coding and content broadcasting to heterogeneous devices which share different screen sizes, we have shown that M-band transforms are well-suited to perform the image decomposition. First, they offer a rational scalability allowing the devices to directly reconstruct an image at the desired resolution, without any post-resampling. We have shown in this paper with experimental results that the 8-band 9/7-packet and LBT transforms have a better coding efficiency for different rational scalability factors than the popular 9/7 dyadic transform with post-

Figure 6: Zooms on reconstructed Lena images reduced by a factor 8/3 with several 8-band transforms at 0.4 bpp with $P = Q = 3$.

Figure 7: Zooms on reconstructed Barbara images reduced by a factor 8/6 with several 8-band transforms at 1.0 bpp with $P = Q = 6$.

resampling. Future work will focus on demonstrating coding effiency in video compression, as well as designing optimal resampling filters.

REFERENCES

[1] G. Marquant, E. François, N. Burdin, P. Lopez, and J. Viéron, "Extended spatial scalability for non dyadic video formats: from SDTV to HDTV," in *Proc. of SPIE Visual Communications and Image Processing*, Beijing, China, July 2005.

[2] T. Nakachi, T. Sawabe, J. Suzuki, and T. Fujii, "A study on non-octave scalable coding with filter bank and its performance evaluation using EBCOT," in *Int. Symposium on Communications and Information Technologies (ISCIT'04)*, Sapporo, Japan, Oct. 2004.

[3] Y. S. Park and H. W. Park, "Arbitrary-ratio image resizing using fast DCT of composite length for DCT-based transcoder," *IEEE Transactions on Image Processing*, vol. 15, no. 2, pp. 494–500, Feb. 2006.

[4] R. M. Figueras, P. Vandergheynst, and P. Frossard, "Low-rate and flexible image coding with redundant representations," *IEEE Transactions on Image Processing*, vol. 15, no. 3, pp. 726–739, March 2006.

[5] P. P. Vaidyanathan, *Multirate systems and filter banks*, Prentice Hall, Englewood Cliffs, 1993.

[6] G. Pau, B. Pesquet-Popescu, and G. Piella, "Modified M-band synthesis filter bank for fractional scalability of images," *IEEE Signal Processing Letters*, June 2006.

[7] P. Steffen, P. N. Heller, R. A. Gopinah, and C. S. Burrus, "Theory of regular *M*-band wavelet bases," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3497–3511, December 1993.

[8] H. S. Malvar and D. H. Staelin, "The LOT: transform coding without blocking effects," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-37, pp. 553–559, April 1989.

[9] H. S. Malvar, "Biorthogonal and nonuniform lapped transforms for transform coding with reduced blocking and ringing artifacts," *IEEE Transactions on Signal Processing*, vol. 46, no. 4, pp. 1043–1053, April 1998.

[10] R. L. de Queiroz, T. Q. Nguyen, and K. R. Rao, "GenLOT: generalized linear-phase lapped orthogonal transforms," *IEEE Transactions on Signal Processing*, vol. 44, pp. 497–507, April 1996.

[11] 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.

[12] D. Seidner, "Polyphase antialiasing in resampling of images," *IEEE Transactions on Image Processing*, vol. 14, no. 11, pp. 1879–1889, November 2005.

[13] H. Malvar, "Lapped transforms software for image processing," *http://research.microsoft.com/∼malvar/software/programs.aspx*, January 2006.

[14] H. S. Malvar, "Fast progressive image coding without wavelets," in *Proceedings of IEEE Data Compression Conference*, Snowbird, UT, March 2000, pp. 243–252.

[15] M. Li and T. Nguyen, "Optimal wavelet filter design in scalable video coding," in *Proc. of the IEEE Int. Conf. on Image Processing*, Genova, Italy, September 2005.