A FAST LAPPED TRANSFORM FOR IMAGE CODING

Ricardo L. de Queiroz[1] *and Trac D. Tran*[2]
[1]Xerox Corporation, Webster, NY, 14580
queiroz@wrc.xerox.com
[2]The Johns Hopkins University, ECE Department, Baltimore, MD, 21218
ttran@ece.jhu.edu

ABSTRACT

This paper introduces a class of linear phase lapped biorthogonal transforms with basis functions of variable length. A lattice is used to enforce both linear phase and perfect reconstruction properties as well as to provide a fast and efficient transform implementation for image coding applications. In the proposed formulation which we call fast lapped transform (FLT), the higher frequency filters (basis functions) are those of the DCT, which are compact to limit ringing. The lower frequency filters (basis functions) are overlapped for representing smooth signals while avoiding blocking artifacts. A great part of the FLT computation is spent at the DCT stage, which can be implemented through fast algorithms, while just a few more operations are needed to implement the extra stages. For example, compared to the DCT, an FLT with good performance can be implemented with only 8 extra additions and 6 extra multiplications for an 8-sample block. Yet, image coding examples show that the FLT is far superior to the DCT and is close to the 9/7-tap biorthogonal wavelet in subjective coding performance.

1. INTRODUCTION

Block transforms are widely used in image coding. For instance, the discrete cosine transform (DCT) is used in the JPEG image compression standard.[1] At high bit rates, JPEG offers excellent image quality and almost visually lossless reconstruction. However, as compression increases, the reconstructed image is often subjected to blocking and ringing artifacts. The development of the lapped orthogonal transform (LOT),[2] its generalized version GenLOT,[3] and the extensions to biorthogonality[4–7] help solve the blocking problem by borrowing pixels from the adjacent blocks to produce the transform coefficients of the current block. Lapped transforms outperform the DCT on two counts: (i) from the analysis viewpoint, it takes into account inter-block correlation, hence, provides better energy compaction; (ii) from the synthesis viewpoint, its basis functions decay asymptotically to zero at the ends, reducing blocking discontinuities drastically. Nevertheless, lapped transforms have not yet been able to replace the DCT in international standards. One reason is the increase in computational complexity.

In this paper, we introduce the fast lapped transform (FLT) which has filters with variable length and can be implemented by adding minimal complexity to the DCT. Despite its simplicity the FLT greatly reduces said artifacts at high compression ratios and is competitive with much more complex filter banks.

The motivation for the FLT is the fact that blocking is most noticeable in smooth image regions. Thus, in order to reduce blocking artifacts, filters covering high-frequency bands do not have to be long and overlapped, i.e. long basis functions are better suited for lower frequencies only. Moreover, shorter basis functions for high-frequency signal components can effectively limit the ringing artifacts.

2. LATTICE STRUCTURE FOR THE FLT

2.1. Review of the general case

In this paper's context, lapped transforms are M-channel uniform linear phase perfect reconstruction filter banks (LPPRFBs). The most general lattice for this FB subclass named the generalized lapped biorthogonal transform (GLBT) is presented in a recent paper.[6,7] The GLBT's polyphase matrix $\mathbf{E}(z)$ can be factorized as

$$\mathbf{E}(z) \;=\; \mathbf{G}_{K-1}(z)\,\mathbf{G}_{K-2}(z)\,\cdots\,\mathbf{G}_1(z)\;\mathbf{E}_0, \tag{1}$$

$$\mathbf{G}_i(z) = \tfrac{1}{2}\begin{bmatrix} \mathbf{U}_i & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_i \end{bmatrix}\begin{bmatrix} \mathbf{I} & \mathbf{I} \\ \mathbf{I} & -\mathbf{I} \end{bmatrix}\begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & z^{-1}\mathbf{I} \end{bmatrix}\begin{bmatrix} \mathbf{I} & \mathbf{I} \\ \mathbf{I} & -\mathbf{I} \end{bmatrix}$$

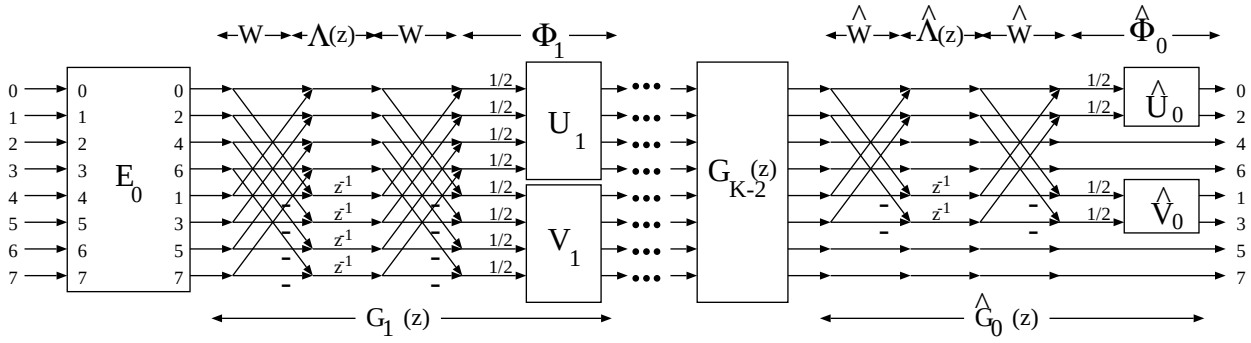

Figure 1. Detailed lattice structure for a VLGLBT (drawn for $M = 8$ and $N = 4$).

$$\stackrel{\triangle}{=} \tfrac{1}{2}\, \mathbf{\Phi}_i\, \mathbf{W}\, \mathbf{\Lambda}(z)\, \mathbf{W}, \quad \text{and} \tag{2}$$

$$\mathbf{E}_0 \;=\; \tfrac{1}{\sqrt{2}} \begin{bmatrix} \mathbf{U}_0 & \mathbf{U}_0\, \mathbf{J}_{\frac{M}{2}} \\ \mathbf{V}_0\, \mathbf{J}_{\frac{M}{2}} & -\mathbf{V}_0 \end{bmatrix}. \tag{3}$$

This lattice results in all filters having length $L = KM$. Each cascading structure $\mathbf{G}_i(z)$ increases the filter length by M. All $\mathbf{U}_i$ and $\mathbf{V}_i$, $i = 0, 1, \ldots, K-1$, are arbitrary $\frac{M}{2} \times \frac{M}{2}$ invertible matrices, and they can be completely parameterized by their SVD decomposition, i.e.,

$$\mathbf{U}_i = \mathbf{U}_{i0} \mathbf{\Gamma}_i \mathbf{U}_{i1} \text{ and } \mathbf{V}_i = \mathbf{V}_{i0} \mathbf{\Delta}_i \mathbf{V}_{i1}, \tag{4}$$

where $\mathbf{U}_{i0}$, $\mathbf{U}_{i1}$, $\mathbf{V}_{i0}$, $\mathbf{V}_{i1}$ are orthogonal matrices, and $\mathbf{\Gamma}_i$, $\mathbf{\Delta}_i$ are diagonal matrices with positive elements. For fast implementations, the initial stage $\mathbf{E}_0$ can be replaced by the DCT. Further trade-off between the FB's speed and performance can be elegantly carried out by setting some of the diagonal multipliers to 1 or some of the rotation angles to 0. See the appropriate references[6,7] for details on the GLBT.

2.2. Variable-length lattices

Let us first consider an M-channel LPPRFB with variable-length filters: M is even, N filters of length MK, $(M-N)$ filters of length $M(K-1)$, each analysis filter $h_i[n]$ and the corresponding synthesis filter $f_i[n]$ have the same length $L_i, 0 \le i \le M-1$. We can show[8] that

- The number of long filters N and the number of short filters $(M-N)$ must both be even.
- Half of the long (short) filters are symmetric.

Let $\mathbf{E}_L(z)$ be the $N \times M$ polyphase matrix of order $(K-1)$, representing the long analysis filters, and $\mathbf{E}_S(z)$ be the $(M-N) \times M$ polyphase matrix of order $(K-2)$, representing the shorter analysis filters. Similarly, let $\mathbf{R}_L(z)$ and $\mathbf{R}_S(z)$ represent the long and the short synthesis filters respectively. Without any loss of generality, the long filters are permuted to be on top. The following factorization establishes the completeness of our solution.

Since all filters have linear phase, $\mathbf{E}(z)$ also has to satisfy the LP property:

$$\begin{cases} \mathbf{E}_L(z) &=& z^{-(K-1)}\, \mathbf{D}_L\, \mathbf{E}_L(z^{-1})\, \mathbf{J}_M \\ \mathbf{E}_S(z) &=& z^{-(K-2)}\, \mathbf{D}_S\, \mathbf{E}_S(z^{-1})\, \mathbf{J}_M \end{cases} \tag{5}$$

where $N \times N$ $\mathbf{D}_L$ and $(M-N) \times (M-N)$ $\mathbf{D}_S$ are diagonal matrices whose entries are $+1$ when the corresponding filter is symmetric and -1 when the corresponding filter is antisymmetric. $\mathbf{E}_L(z)$ now forms a remarkably similar system to an N-channel order-$(K-1)$ GLBT.[6]

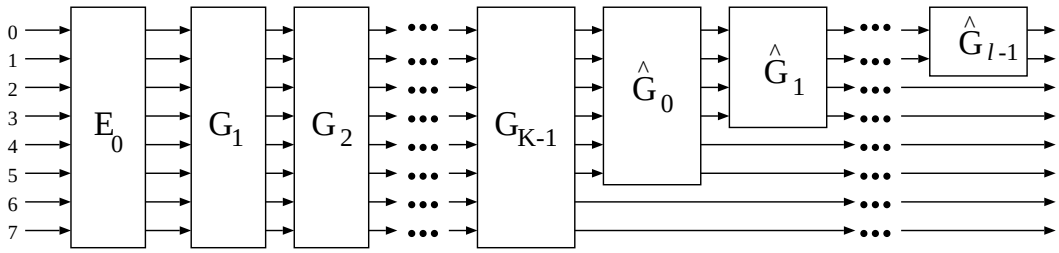

Figure 2. General lattice structure for the VLGLBT.

From the work on the GLBT,[6,7] there always exists a factorization similar to the one shown in Eq.(1) that reduces the order of the polyphase matrix $\mathbf{E}_L(z)$ by one. Hence, the polyphase matrix $\mathbf{E}(z)$ can always be factorized as follows (the $\frac{N}{2}$ long symmetric filters are arranged on top)

$$\mathbf{E}(z) \;=\; \hat{\mathbf{G}}_0(z)\,\mathbf{E}_{K-2}(z), \tag{6}$$

where

$$\hat{\mathbf{G}}_0(z) \;\stackrel{\triangle}{=}\; \tfrac{1}{2}\,\hat{\mathbf{\Phi}}_0\,\hat{\mathbf{W}}\,\hat{\mathbf{\Lambda}}(z)\,\hat{\mathbf{W}}, \quad \text{and} \tag{7}$$

$$\hat{\mathbf{W}} = \begin{bmatrix} \mathbf{I}_{\frac{N}{2}} & \mathbf{I}_{\frac{N}{2}} & \mathbf{0} \\ \mathbf{I}_{\frac{N}{2}} & -\mathbf{I}_{\frac{N}{2}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \end{bmatrix},\; \hat{\mathbf{\Lambda}}(z) = \begin{bmatrix} \mathbf{I}_{\frac{N}{2}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & z^{-1}\mathbf{I}_{\frac{N}{2}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \end{bmatrix}, \tag{8}$$

$$\hat{\mathbf{\Phi}}_0 \;=\; \begin{bmatrix} \hat{\mathbf{U}}_0 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \hat{\mathbf{V}}_0 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \end{bmatrix}. \tag{9}$$

The above factorization leaves $\mathbf{E}_S(z)$ untouched, it reduces the length of the longer filters by M, so all filters now have the same length of $M(K-1)$. $\mathbf{E}_{K-2}(z)$ is the familiar polyphase matrix of an order-$(K-2)$ GLBT, and it can be factorized into the familiar cascade structure in Eq.(1). The complete factorization is

$$\mathbf{E}(z) \;=\; \hat{\mathbf{G}}_0(z)\,\mathbf{G}_{K-2}(z)\,\cdots\,\mathbf{G}_1(z)\;\mathbf{E}_0, \tag{10}$$

and is depicted in Figure 1. The inverse transform is obtained by inverting each building block and causality can be achieved by a z^{-1} shift, i.e.,

$$\mathbf{R}(z) \;=\; \mathbf{E}_0^{-1}\,z^{-1}\,\mathbf{G}_1^{-1}(z)\,\cdots\,z^{-1}\,\mathbf{G}_{K-2}^{-1}(z)\,z^{-1}\,\hat{\mathbf{G}}_0(z).$$

We should also mention that the factorization in Eq.(10) can be proven to be minimal, i.e., the resulting lattice employs the least number of delays in the implementation.[9] We use the term variable length GLBT or VLGLBT to refer to such a structure. Of course, more VL structures $\hat{\mathbf{G}}_i(z)$ can be added to increase the frequency resolution of the long filters. Each $\hat{\mathbf{G}}_i(z)$ block increases the length of N_i filters by M and leaves the rest intact. The most general lattice for the VLGLBT is shown in Figure 2.

2.3. The FLT

In order to make a lapped transform competitive with the DCT, we are concerned with high-performance, yet low-complexity, lapped transforms where the newly-found flexibility of the VLGLBT is exploited to our advantage. To minimize complexity, we use very few long filters and set (i) $K = 1$; (ii) $\mathbf{G}_0(z) = \mathbf{E}_0 = \mathbf{C}_8$ where $\mathbf{C}_n$ is the $n \times n$ DCT matrix; (iii) N_i constant. In other words the FLT is defined as having the analysis polyphase transfer matrix as

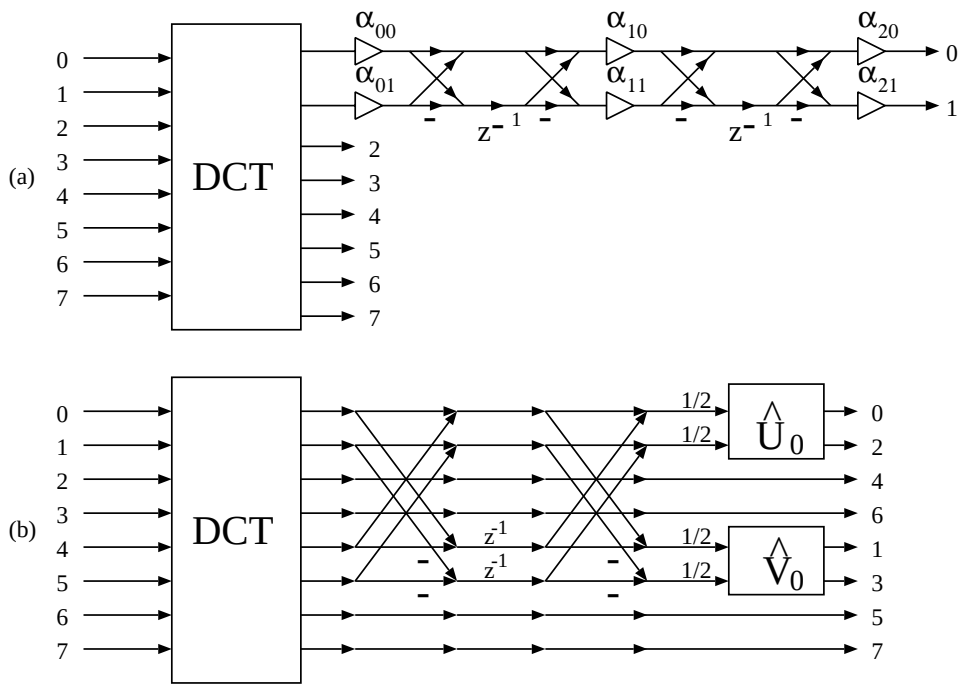

Figure 3. The FLT lattice drawn for $M = 8$. (a) FLT-I. (b) FLT-III.

$$\mathbf{E}(z) = \begin{bmatrix} \mathbf{F}(z) & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{M-N} \end{bmatrix} \mathbf{C}_M \tag{11}$$

where $\mathbf{F}(z)$ is an $N \times N$ transfer matrix, post-processing the N lower frequency outputs of the DCT. We have two main approaches for the design of the FLT.

The first approach is to use one or two variable-length structures $\hat{\mathbf{G}}_0(z)$ and $\hat{\mathbf{G}}_1(z)$ in which $\hat{\mathbf{U}}_i$ and $\hat{\mathbf{V}}_i$ are biorthogonal. A design example can be obtained using unconstrained nonlinear optimization where the cost function is a weighted combination of coding gain, DC leakage, stopband attenuation, and mirror frequency attenuation.[8,9]

The second approach is to decompose $\mathbf{F}(z)$ as

$$\mathbf{F}(z) = \mathbf{F}'(z)\mathbf{C}_N^T \tag{12}$$

i.e. apply an inverse DCT to obtain a pseudo non-uniform band[10] and then apply any N-channel filter bank $\mathbf{F}'(z)$ that one sees fit.

From those two methods we derived 3 instances of the FLT which we found useful for image coding and we will describe next.

FLT-I - We select ($N = N_0 = N_1 = 2$) where the $\hat{\mathbf{U}}_i$ and $\hat{\mathbf{V}}_i$ matrices become scalar parameters α_{ij}. The resulting lattice is shown in Fig. 3(a) for $M = 8$. The extra complexity is only 8 additions and 6 multiplications, i.e. 14 operations per block of 8 samples. Our particular design example used

$$\{\alpha_{00}, \alpha_{01}, \alpha_{10}, \alpha_{10}, \alpha_{20}, \alpha_{21}\} = \{1.9965, 1.3193, 0.4388, 0.7136, 0.9385, 1.2878\}.$$

The impulse responses of the two lowest frequency (longer) filters for this example FLT are depicted in Figure 4. Note that the last filters (bases) of the FLT are those of the DCT.

FLT-II - The second type also uses $N = 2$, but $\mathbf{F}'(z)$ is any filter bank, preferably suited for image coding. We carried tests using the popular 9/7 biorthogonal filter bank[11] as well for Johnston's 16-tap QMF bank[11] and Malvar's

2-channel ELT.[2] Those filter banks can be implemented with 26, 32, and 14 operations per block, respectively. Note that the inverse 2×2 DCT can be at least partially incorporated in the filter banks. For example, in the ELT case, the DCT is cascaded with its first stage with the effect of simply generating a new plane rotation as the ELT's first stage.

FLT-III - The third type can be devised using either approach. We select $N = 4$ and use a 4-channel GLBT[6] as $\mathbf{F}'(z)$. As the first stage of the GLBT is commonly selected as the $N \times N$ DCT, said first stage cancels the factor $\mathbf{C}_N^T$ and only the remaining parts need to be implemented. Those remaining parts are exactly the factors described in Sec. 2.2. Hence, one can view the FLT-III as a system with $N = 4$ where $\hat{\mathbf{G}}_0(z)$ has the same matrix parameters as the 4×8 GLBT.[6] The extra complexity to implement $\hat{\mathbf{G}}_0(z)$ is about 20 operations per block. The resulting lattice is shown in Figure 3(b) in an example for $M = 8$. The 4 impulse responses of the modified (longer) filters for this example are depicted in Figure 5. The remaining 4 impulse responses are the 8-tap high frequency bases of the DCT.

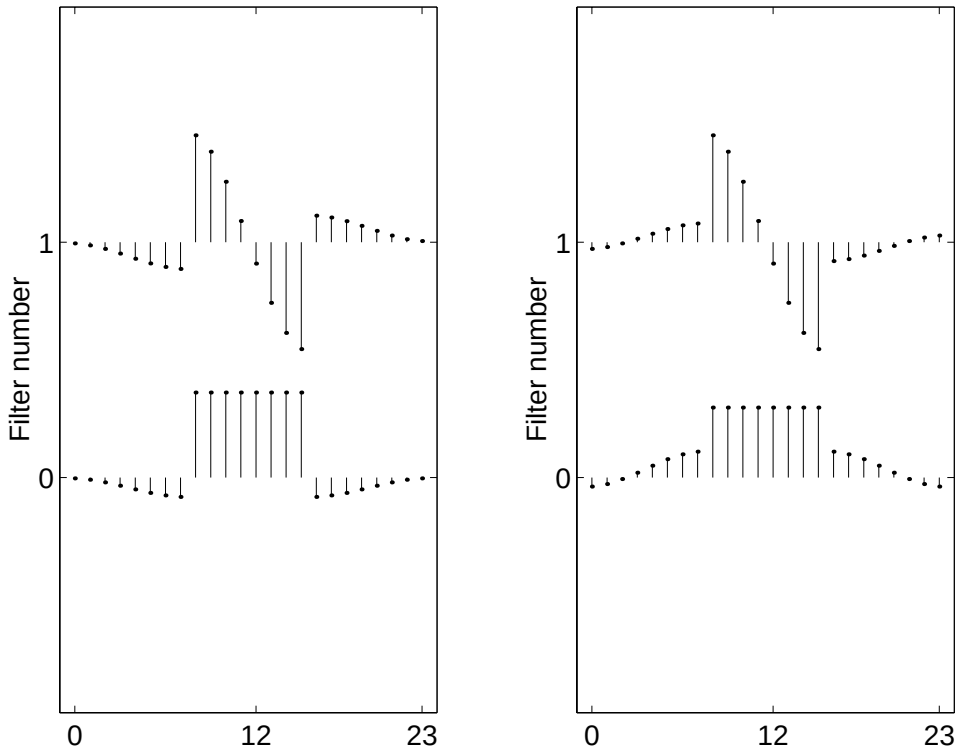

Figure 4. The two longer bases for the example FLT-I with $M = 8$. Left: analysis; right: synthesis. The remaining six bases are those of the DCT.

3. APPLICATION TO IMAGE CODING

The transforms were coupled with a high-performance image compression system,[12] which is an adaptation of SPIHT[13] for lapped transforms. The DC subband is further decorrelated using 3 additional dyadic iterations using the 9/7 filters in order to have a fair comparison with coders such as SPIHT. The complexity of the extra stages is negligible since the DC subbands is 64 times smaller than the original image.

We carried objective comparisons for popular images using PSNR (dB) as a distortion measure. Among the 3 variations of the FLT-II, the difference in PSNR was always below 0.1 dB for compression ratios below 16:1, and below 0.2 for higher compression. The 9/7-tap biorthogonal filter bank always slightly led the pack in PSNR. We feel it does not matter much which one to test, however we will use the 9/7 filter bank when referring to FLT-II.

We compare 6 transforms, the 8×8 DCT, Malvar's 8×16 fast LOT,[2] 3 stages of 9/7-taps filter bank (i.e a 6-stage SPIHT coder), and the 3 FLTs. Objective comparisons are shown in Table 1 while an example of subjective comparison is shown in Fig. 6.

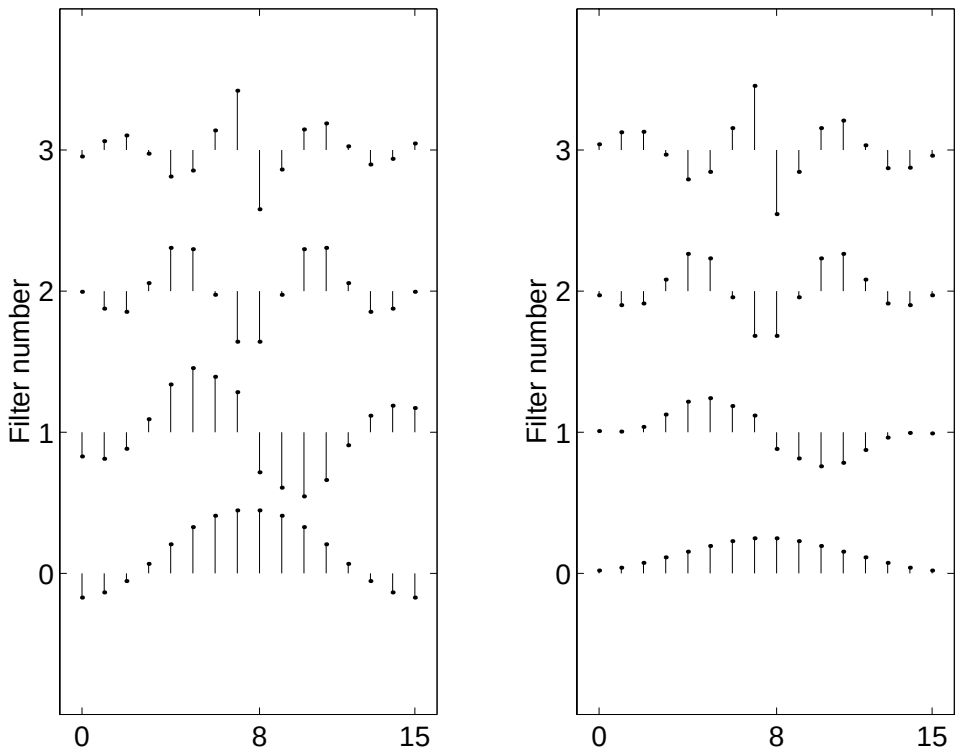

Figure 5. The four longer bases for the example FLT-III with $M = 8$. Left: analysis; right: synthesis. The remaining four bases are those of the DCT.

As expected, the FLT offers a $0.3 - 0.5$ dB improvement over the DCT at medium and low bit rates. Also, the FLT provides a significant improvement in image quality over the traditional DCT: blocking is avoided while ringing is suppressed. In fact, the FLT is much better than the fast LOT^2 in blocking elimination. The FLT's visual performance comes quite close to those of state-of-the-art wavelets at a much lower computational cost.

4. CONCLUSIONS

We have presented in this paper a class of variable-length lapped biorthogonal transform. The FLT is based on a fast, efficient, robust, and modular lattice structure. With low overhead, comparing to the DCT, our transform offers a fast, low-cost, VLSI-friendly implementation while providing high-quality reconstructed images at medium and low bit rates as demonstrated in the image coding example. Its block-based nature also supports parallel processing mode, facilitates region-of-interest coding/decoding, and is capable of processing large images under limited resource constraint.

(a) (b)

(c) (d)

Figure 6. Coding results of Lena at 1:32 compression ratio. Enlarged portions. (a) DCT (b) FLT-I (c) FLT-II (d) FLT-III (e) LOT (f) SPIHT, 9/7-tap biorthogonal wavelet.

Table 1. Objective comparison among transforms. Rate is given as compression ratio (CR), distortion is given in PSNR (dB) and complexity (OPS) is given in operations necessary to compute a 1D block of 8 samples. Compression using the DWT 9/7 is equivalent to the SPIHT coder.

CR	DCT	FLT-I	FLT-II	FLT-III	LOT	WT 9/7
	Lena					
8:1	39.91	39.89	40.05	40.11	40.01	40.41
16:1	36.38	36.51	36.67	36.83	36.68	37.21
32:1	32.90	33.25	33.44	33.63	33.48	34.11
64:1	29.67	30.15	30.35	30.66	30.42	31.10
	Goldhill					
8:1	36.25	36.22	36.36	36.40	36.56	36.55
16:1	32.76	32.76	32.92	32.98	33.11	33.13
32:1	30.07	30.25	30.34	30.45	30.51	30.56
64:1	27.93	28.17	28.29	28.41	28.36	28.48
OPS	40	54	68	60	82	182

(e) (f)

Figure 6. (Continued) Coding results of Lena at 1:32 compression ratio. Enlarged portions. (a) DCT (b) FLT-I (c) FLT-II (d) FLT-III (e) LOT (f) SPIHT, 9/7-tap biorthogonal wavelet.

REFERENCES

1. W. B. Pennebaker and J. L. Mitchell, *JPEG: Still Image Compression Standard*, Van Nostrand Reinhold, 1993.
2. H. S. Malvar, *Signal Processing with Lapped Transforms*, Artech House, 1992.
3. R. de Queiroz, T. Q. Nguyen, and K. Rao, "The GenLOT: generalized linear-phase lapped orthogonal transform," *IEEE Trans. on Signal Processing,* vol. 40, pp. 497-507, Mar. 1996.
4. S. C. Chan, "The generalized lapped transform (GLT) for subband coding applications," *Proc. Intl. Conf. Acoust. Speech Signal Processing,* Detroit, pp. 1508–1511, May 1995.
5. H. S. Malvar, "Lapped biorthogonal transforms for transform coding with reduced blocking and ringing artifacts," *Proc. Intl. Conf. Acoust. Speech Signal Processing,* Munich, Germany, April 1997.
6. T. D. Tran, R. de Queiroz, and T. Q. Nguyen, "The generalized lapped biorthogonal transform," *Proc. Intl. Conf. Acoust. Speech Signal Processing,* Seattle, VOl. III, pp. 1441–1444, May 1998.
7. T. D. Tran, R. L. de Queiroz, and T. Q. Nguyen, "Linear phase perfect reconstruction filter bank: lattice structure, design, and application in image coding," *IEEE Trans. on Signal Processing*, Vol. 44, Jan. 2000.
8. T. D. Tran, R. L. de Queiroz, and T. Q. Nguyen, "The variable-length generalized lapped biorthogonal transform," *Proc. Intl. Conf. on Image Processing*, Chicago IL, 1998.
9. T. D. Tran, *Linear phase perfect reconstruction filter banks: theory, structure, design, and application in image compression*, Ph.D. thesis, University of Wisconsin, Madison, WI, May 1998.
10. R. L. de Queiroz, "Uniform filter banks with non-uniform bands: post-processing design," *Proc. Intl Conf. Acoust. Speech Signal Processing,* Seattle, WA, Vol. III, pp. 1341–1344, May 1997.
11. G. Strang and T. Nguyen, *Wavelets and Filter Banks,* Wellesley, MA: Wellesley-Cambridge, 1996.
12. T. D. Tran and T. Q. Nguyen, "A lapped transform embedded image coder," *Proc. Intl. Symp. on Circuits and Systems,* Monterey, CA, May 1998.
13. A. Said and W. A. Pearlman, "A new fast and efficient image codec based on set partitioning in hierarchical trees," *IEEE Trans on Circuits Syst. Video Tech.,* vol. 6, pp. 243-250, June 1996.