ERROR RESILIENT PRE-/POST-FILTERING FOR DCT-BASED BLOCK CODING SYSTEMS

Chengjie Tu, Trac D. Tran, and Jie Liang

The Johns Hopkins University
Department of Electrical and Computer Engineering
Baltimore, MD 21218
Email: {cjtu, trac, jliang}@jhu.edu

ABSTRACT

Pre-/post-filtering can be attached to a DCT-based block coding system to improve coding efficiency as well as to mitigate blocking artifacts. Previously designed pre-/post-filters are optimized to maximize coding efficiency solely. For image and video communication over unreliable channels, those pre-/post-filters are sensitive to transmission errors. This paper addresses the problem of designing pre-/post-filters which are more error resilient. A family of pre-/post-filters are designed to provide desired trade-offs between coding efficiency and robustness to transmission errors. These filters achieve superior reconstruction performance without sacrificing much coding performance.

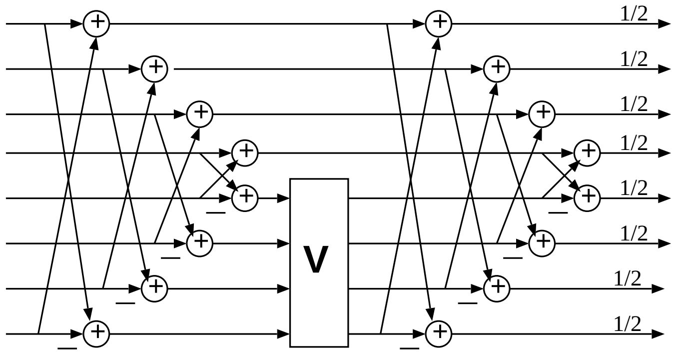

Fig. 2. Structure of pre-filter P.

1. INTRODUCTION

Block coding based on the Discrete Cosine Transform (DCT) is very popular in image and video compression. It enjoys the DCT's excellent energy compaction capability, low complexity, and high flexibility on a block-by-block basis. Unfortunately, its coding efficiency heavily suffers from ignoring correlation between blocks. More annoyingly, blocking artifacts manifest at low bitrates.

An elegant remedy to DCT-based block coding is recently proposed in [1]: employing pre-filtering in the encoder and post-filtering in the decoder along the block boundaries of the DCT coding framework. Pre-/post-filtering operates in the time domain and is completely outside the existing DCT-based infrastructure. It has been demonstrated that by turning on pre-/post-filtering, DCT-based block coding can achieve much better performance at the cost of slightly increased complexity [2].

The problem of designing pre-/post-filters optimized for coding efficiency has been extensively investigated in [1]. For image and video delivery over unreliable channels, those filters are sensitive to transmission errors. A comprehensive review on techniques combating transmission errors can be found in [3]. This paper concentrates on the design of error resilient pre-/post-filters. Since pre-/post-filtering combined with the DCT is a particular implementation of Lapped Transforms (LTs), the paper can be viewed as an extension to the pioneering work of designing error resilient lapped orthogonal transforms (LOTs) in [4, 5]. However, our approach provides the following advantages: designing pre-/post-filters is far easier than designing LOTs directly; pre-/post-filter based LTs are computationally efficient; orthogonality is not imposed since biorthogonal LTs have higher coding efficiency and are better at eliminating coding artifacts; unlike LOTs which can only improve error distribution, pre-/post-filters can be designed to not only improve error distribution but also notably decrease average error, resulting in improved reconstructions.

Notation-wise, $\mathbf{I}_N$, $\mathbf{J}_N$, $\mathbf{0}_N$ denote the $N \times N$ identity, reversal identity, and null matrices, respectively. The $AR(1)$ image model with unit variance σ_x^2 and intersample correlation $\rho = 0.95$ is assumed throughout.

2. PRE-/POST-FILTERING

The pre-/post-filtering framework [1] is illustrated in Fig. 1: the pre-filter $\mathbf{P}$ and the post-filter $\mathbf{P^{-1}}$ operate on the boundaries of the DCT without affecting the existing DCT-

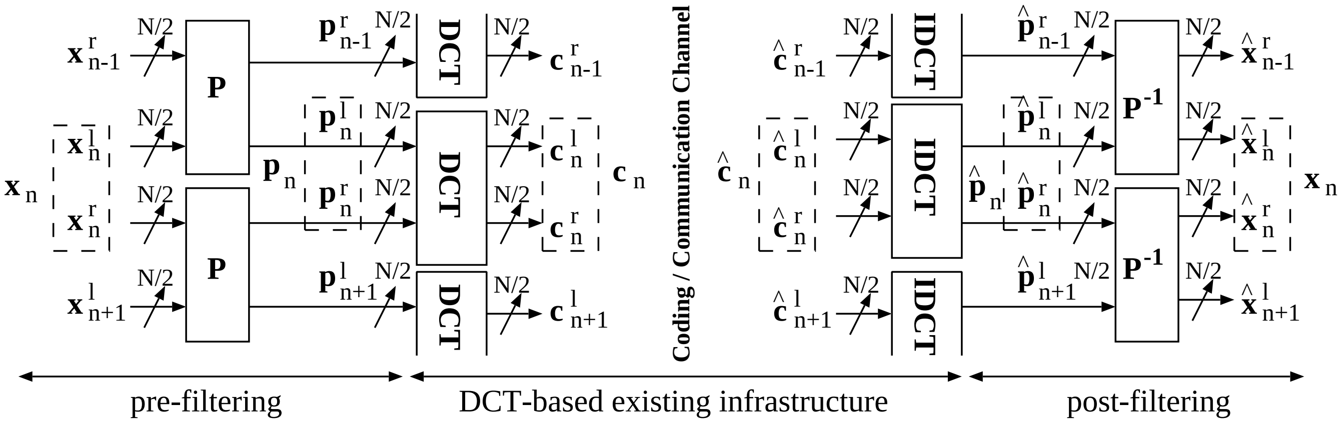

Fig. 1. The pre-/post-filtering framework

based infrastructure. $\mathbf{P}$ ($\mathbf{P}^{-1}$) has a specific structure as shown in Fig. 2: two stages of butterflies with an $\frac{N}{2} \times \frac{N}{2}$ matrix $\mathbf{V}$ ($\mathbf{V}^{-1}$) between them. So pre-/post-filters are uniquely specified by the $\frac{N}{2} \times \frac{N}{2}$ matrix $\mathbf{V}$. Consequently, designing pre-/post-filters ($\frac{N^2}{4}$ free parameters involved) is far more tractable than designing LTs directly ($2N^2$ free parameters involved). Furthermore, pre-/post-filtering is computationally inexpensive ($\frac{N^2}{4}$ multiplications per block).

Partitioning $\mathbf{P}$ and $\mathbf{P^{-1}}$ into square $\frac{N}{2} \times \frac{N}{2}$ submatrices:

$$\mathbf{P} = \begin{bmatrix} \mathbf{P}_{00} & \mathbf{P}_{01} \\ \mathbf{P}_{10} & \mathbf{P}_{11} \end{bmatrix}, \quad \mathbf{P^{-1}} = \begin{bmatrix} \mathbf{T}_{00} & \mathbf{T}_{01} \\ \mathbf{T}_{10} & \mathbf{T}_{11} \end{bmatrix} \tag{1}$$

and with the notation shown in Fig. 1, we have

$$\begin{bmatrix} \mathbf{p}_{n-1}^{r} \\ \mathbf{p}_{n}^{l} \end{bmatrix} = \mathbf{P} \begin{bmatrix} \mathbf{x}_{n-1}^{r} \\ \mathbf{x}_{n}^{l} \end{bmatrix} = \begin{bmatrix} \mathbf{P}_{00}\mathbf{x}_{n-1}^{r} + \mathbf{P}_{01}\mathbf{x}_{n}^{l} \\ \mathbf{P}_{10}\mathbf{x}_{n-1}^{r} + \mathbf{P}_{11}\mathbf{x}_{n}^{l} \end{bmatrix}, \tag{2}$$

$$\begin{bmatrix} \hat{\mathbf{x}}_{n-1}^{r} \\ \hat{\mathbf{x}}_{n}^{l} \end{bmatrix} = \mathbf{P^{-1}} \begin{bmatrix} \hat{\mathbf{p}}_{n-1}^{r} \\ \hat{\mathbf{p}}_{n}^{l} \end{bmatrix} = \begin{bmatrix} \mathbf{T}_{00}\hat{\mathbf{p}}_{n-1}^{r} + \mathbf{T}_{01}\hat{\mathbf{p}}_{n}^{l} \\ \mathbf{T}_{10}\hat{\mathbf{p}}_{n-1}^{r} + \mathbf{T}_{11}\hat{\mathbf{p}}_{n}^{l} \end{bmatrix}. \tag{3}$$

It can be shown that the pre-/post-filtering framework generates an N-band $2N$-tap ($N \times 2N$) LT with a specific forward transform $\mathbf{H}$ and inverse transform $\mathbf{F}$ [1]. The coding performance of a pre-/post-filter pair is measured via the coding gain of the corresponding LT:

$$G_{TC} = 10log_{10}\left(\frac{\frac{1}{N}\sum_{i=0}^{N-1}\sigma_i^2}{\left(\prod_{i=0}^{N-1}\sigma_i^2 f_i^2\right)^{\frac{1}{N}}}\right), \tag{4}$$

where σ_i^2 and f_i^2 are the i^{th} diagonal entry of $\mathbf{H}^T\mathcal{R}_{xx}\mathbf{H}$ and $\mathbf{F}\mathbf{F}^T$ respectively whereas $\mathcal{R}_{xx} = [\rho^{|i-j|}]_{2N}$ is the autocorrelation matrix of the input signal.

3. RECONSTRUCTION CRITERIA

The imperfect transmission of an N-point coefficient block $\mathbf{c}_n$ affects the reconstruction of a $2N$-point signal block $[(\mathbf{x}_{n-1}^r)^T, (\mathbf{x}_n)^T, (\mathbf{x}_{n+1}^l)^T]^T$. Denote the reconstruction error as $\mathbf{e}_n$ and its auto correlation matrix as $\mathcal{R}_{ee}$. The $i^{th}(i = 0, \cdots, 2N-1)$ diagonal entry of $\mathcal{R}_{ee}$, e_i^2, is the mean squared error (MSE) of the i^{th} reconstructed sample.

The objective quality of the reconstructed block in terms of the peak signal-noise ratio (PSNR) depends on the MSE for the entire block,

$$MSE = \frac{1}{2N}\sum_{i=0}^{2N-1} e_i^2 = \frac{1}{2N}Tr(\mathcal{R}_{ee}). \tag{5}$$

Besides the MSE, the distribution of e_i^2 also has a significant impact on the visual quality of the reconstructed image. A more uniform error distribution has less artifacts and is visually more pleasing. The reconstruction gain can be defined as

$$G_R = \frac{\left(\prod_{i=0}^{2N-1} e_i^2\right)^{\frac{1}{2N}}}{\frac{1}{2N}\sum_{i=0}^{2N-1} e_i^2} \tag{6}$$

which measures how uniform the distribution is.

4. OPTIMAL PRE- AND POST-FILTER DESIGN

From the previous section, designing an error-resilient pre-/post-filter pair is equivalent to finding a $\mathbf{V}$ to optimize the corresponding G_{TC}, MSE, and G_R. Unfortunately, G_{TC}, MSE, and G_R can not reach their optimal values simultaneously. We typically setup an optimization procedure to maximize a weighted sum, $G_{TC} - \alpha \times 10log_{10}(MSE) + \beta \times 10log10(G_R)$, taking $\mathbf{V}$ as argument, where α and β are non-negative weights. The choice of the weights highly depends on the loss pattern.

The key of the design procedure is to find the error autocorrelation matrix $\mathcal{R}_{ee}$ according to the specific loss pattern and the recovery method. Without loss of generality, we assume that a lost block is recovered by the mean of its perfectly received neighboring blocks:

$$\hat{\mathbf{c}}_n = \frac{1}{2}(\mathbf{c}_{n-1} + \mathbf{c}_{n+1}). \quad (7)$$

Other reconstruction methods do not affect the design procedure; only $\mathcal{R}_{ee}$ might take a different form. Under the pre-/post-filtering framework, (7) is equivalent to

$$\hat{\mathbf{p}}(n) = \frac{1}{2}(\mathbf{p}_{n-1} + \mathbf{p}_{n+1}), \quad (8)$$

and so the recovery can be performed in the time domain after the IDCT.

Define

$$\begin{aligned} \mathbf{A} &= \begin{bmatrix} \mathbf{T}_{01} \\ \mathbf{T}_{11} \end{bmatrix} \begin{bmatrix} \mathbf{P}_{10} & \mathbf{P}_{11} \end{bmatrix}, \\ \mathbf{B} &= \begin{bmatrix} \mathbf{T}_{00} \\ \mathbf{T}_{10} \end{bmatrix} \begin{bmatrix} \mathbf{P}_{00} & \mathbf{P}_{01} \end{bmatrix}. \end{aligned} \quad (9)$$

Noting that $\mathbf{A} + \mathbf{B} = \mathbf{I}_N$ and with the help of (2) and (3),

$$\begin{aligned} \mathbf{e}_n &= \begin{bmatrix} \mathbf{x}^r_{n-1} \\ \mathbf{x}_n \\ \mathbf{x}^l_{n+1} \end{bmatrix} - \begin{bmatrix} \mathbf{T} & \mathbf{0}_N \\ \mathbf{0}_N & \mathbf{T} \end{bmatrix} \begin{bmatrix} \mathbf{p}^r_{n-1} \\ \frac{\mathbf{p}_{n-1}+\mathbf{p}_{n+1}}{2} \\ \mathbf{p}^l_{n+1} \end{bmatrix} \\ &= -\begin{bmatrix} \frac{1}{2}\mathbf{A} & -\mathbf{A} & \frac{1}{2}\mathbf{A} & \mathbf{0}_N \\ \mathbf{0}_N & \frac{1}{2}\mathbf{B} & -\mathbf{B} & \frac{1}{2}\mathbf{B} \end{bmatrix} \begin{bmatrix} \mathbf{x}^r_{n-2} \\ \mathbf{x}_{n-1} \\ \mathbf{x}_n \\ \mathbf{x}_{n+1} \\ \mathbf{x}^l_{n+2} \end{bmatrix}, \end{aligned} \quad (10)$$

$$\mathcal{R}_{ee} = \begin{bmatrix} \mathbf{A} & \mathbf{0}_N \\ \mathbf{0}_N & \mathbf{B} \end{bmatrix} \mathcal{R}_{err} \begin{bmatrix} \mathbf{A}^T & \mathbf{0}_N \\ \mathbf{0}_N & \mathbf{B}^T \end{bmatrix}, \quad (11)$$

where the entry of the $2N \times 2N$ matrix $\mathcal{R}_{err}$ at the i^{th} row and the j^{th} column is $\frac{3}{2}\rho^{|i-j|} - \rho^{|N+i-j|} - \rho^{|N+j-i|} + \frac{1}{4}(\rho^{2N+i-j} + \rho^{2N+j-i})$. Noting that the top-left quadrant and the right-bottom quadrant of $\mathcal{R}_{err}$ are the same matrix denoted as $\mathcal{E}$, and coupling with the fact that Tr($\mathbf{XY}$)=Tr($\mathbf{YX}$),

$$\begin{aligned} MSE &= \frac{1}{2N} Tr(\mathcal{R}_{ee}) \\ &= \frac{1}{2N} Tr\left(\begin{bmatrix} \mathbf{A}^T & \mathbf{0}_N \\ \mathbf{0}_N & \mathbf{B}^T \end{bmatrix} \begin{bmatrix} \mathbf{A} & \mathbf{0}_N \\ \mathbf{0}_N & \mathbf{B} \end{bmatrix} \mathcal{R}_{err} \right) \\ &= \frac{1}{8N} Tr\left(\begin{bmatrix} 2\mathbf{I}_{\frac{N}{2}} + \mathbf{J}_{\frac{N}{2}}\mathbf{K}\mathbf{J}_{\frac{N}{2}} & \mathbf{J}_{\frac{N}{2}}\mathbf{L} \\ \mathbf{L}\mathbf{J}_{\frac{N}{2}} & 2\mathbf{I}_{\frac{N}{2}} + \mathbf{K} \end{bmatrix} \mathcal{E} \right), \end{aligned} \quad (12)$$

where $\mathbf{K} = \mathbf{V}^{-T}\mathbf{V}^{-1} + \mathbf{V}^T\mathbf{V}$, and $\mathbf{L} = \mathbf{V}^{-T}\mathbf{V}^{-1} - \mathbf{V}^T\mathbf{V}$. If orthogonality ($\mathbf{V}^{-1} = \mathbf{V}^T$) is assumed, then $\mathbf{K} = 2\mathbf{I}_{\frac{N}{2}}$ and $\mathbf{L} = \mathbf{0}_{\frac{N}{2}}$, and since Tr($\mathcal{E}$)=$\frac{1}{2}$Tr($\mathcal{R}_{err}$), (12) reduces to

$$MSE = \frac{1}{4N} Tr(\mathcal{R}_{err}) \quad (13)$$

which is a constant. So if orthogonality is imposed, a particular pre-/post-filter pair does not affect the MSE of the entire *signal* block. However, this is not true anymore if orthogonality is absent, and some $\mathbf{V}$s may give significant lower MSEs, resulting in much better reconstructions.

In the case of $N = 8$, some $\mathbf{V}$ designs are listed below:

$$\mathbf{V}_1 = \begin{bmatrix} -1.6769 & 0.6005 & -0.3369 & 0.1006 \\ -0.7091 & 1.2843 & -0.4077 & 0.1601 \\ -0.1774 & 0.7553 & -1.1195 & 0.1202 \\ -0.1131 & 0.1046 & -0.8291 & 0.9090 \end{bmatrix},$$

$$\mathbf{V}_2 = \begin{bmatrix} 1.1547 & 0.8640 & -0.1114 & -0.2059 \\ -0.2708 & 1.1865 & 0.2792 & -0.2730 \\ 0.3406 & 0.0560 & 1.0157 & -0.1406 \\ 0.1023 & 0.2514 & -0.0021 & 0.8509 \end{bmatrix},$$

$$\mathbf{V}_3 = \begin{bmatrix} 0.5183 & -0.3612 & -1.2530 & 0.8415 \\ 0.1582 & 0.8663 & -1.2547 & 0.5062 \\ 1.1711 & 0.2693 & -0.4468 & 0.4451 \\ -0.0511 & 0.2264 & -0.2225 & 0.9502 \end{bmatrix}.$$

Table 1 compares the resulting LTs, {P1, P2, P3}, with several other existing transforms: the DCT, the LOT with the known maximum coding gain 9.22 dB (LOT$_{opt}$) [1], T6 and T9 in [4]. As expected, all orthogonal transforms (the DCT, the LOT$_{opt}$, T6, T9) have the same MSE; yet coding gains and reconstruction gains vary widely. T6 and T9 have better reconstruction gains than the DCT and the LOT$_{opt}$, but with much worse coding gains. None of our designs P1-P3 is orthogonal, but each can achieve a lower MSE and/or a higher G_R without heavily sacrificing coding gains.

Table 1. Comparisons between the resulting LTs and some standard transforms.

	DCT	LOT$_{opt}$	T6	T9	P1	P2	P3
G_{TC}	8.31	9.22	7.83	6.50	6.95	9.20	8.41
MSE	0.20	0.20	0.20	0.20	0.14	0.18	0.15
G_R	0	0.44	0.55	0.85	0.67	0.30	0.64

5. TRANSFORM PERFORMANCE

We benchmark transform performance for the following 5 block loss patterns: S0) no loss; S1) 25% regular checkboard loss; S2) 50% regular checkboard loss; S3) 25% random loss; S4) 50% random loss. To reconstruct a lost block, we search the nearest layer of neighbors with at least one received block in a diamond order and the block is reconstructed as the mean of the received blocks of that layer. The 8-bit 512 × 512 Lena image is used as the test image. We use a block-based coder, L-CEB [2], to simulate an error resilient block image coding system in the following way: an image is coded by L-CEB at a given bitrate and then decoded; blocks belonging to a specific loss pattern are reconstructed prior to performing the inverse transform. The coding and communication framework is fixed. Only different transforms are tested, ensuring the comparison is as

Fig. 3. Portions of reconstructed *Lena* image (coded at 0.25 bpp), suffering 25% random loss: LOT$_{opt}$(left, 25.82 dB), T9(middle, 24.82 dB); P3(right, 27.13 dB).

fair as possible. Table 2 and Table 3 list the reconstruction results in terms of PSNR for different transforms and different loss patterns at 1 bpp and 0.25 bpp respectively. The S0 columns only illustrate coding performance since there is no coefficient loss. Penalized by their bad coding performance, the overall performance of T6 and T9 in terms of PSNR is not impressive. On the other hand, our new designs P1-P3 generally offer much more reasonable performances. Several reconstructed portions are shown in Fig. 3. The P3 reconstruct is more pleasing than the others.

Table 2. Reconstruction PSNRs in dB at 1 bpp

Transform	S0	S1	S2	S3	S4
DCT	39.78	27.24	24.22	26.18	22.63
LOT$_{opt}$	**39.95**	27.25	24.19	26.21	22.65
T6	37.21	27.10	24.15	26.16	22.66
T9	35.64	26.89	24.20	26.18	22.63
P1	34.74	**29.41**	**26.23**	**28.27**	**24.64**
P2	39.88	28.07	25.03	26.88	23.31
P3	38.38	29.33	26.02	28.05	24.28

Table 3. Reconstruction PSNRs in dB at 0.25 bpp

Transform	S0	S1	S2	S3	S4
DCT	33.05	26.63	24.03	25.71	22.54
LOT$_{opt}$	33.61	26.72	24.02	25.82	22.57
T6	28.29	26.23	23.96	25.57	22.64
T9	28.22	25.28	23.56	24.82	22.26
P1	28.05	26.45	24.85	25.96	23.84
P2	**33.69**	27.43	24.81	26.41	23.22
P3	31.97	**28.06**	**25.52**	**27.13**	**24.05**

6. CONCLUSION

This paper presents the design of error resilient pre-/post-filters for the block DCT coding framework. It can also be viewed as a fast and efficient method to design error resilient lapped transforms with low computational complexity. We demonstrate that each designed pre-/post-filter pair provides a different tradeoff between compression and reconstruction performance. Our designs provide a significant performance gain over other approaches, both objectively and subjectively.

7. REFERENCES

[1] T. D. Tran, J. Liang, and C. Tu, "Lapped transform via time-domain pre- and post-processing," *IEEE Trans. on Signal Processing, to appear*.

[2] C. Tu and T. D. Tran, "Context based entropy coding of block transform coefficients for image compression," *IEEE Trans. on Image Processing*, Nov. 2002, to appear.

[3] Y. Wang and Q.-F. Zhu, "Error control and concealment for video communication: a review," *Proc. of The IEEE*, vol. 86, No. 5, pp. 974–997, May 1998.

[4] S. S. Hemami, "Reconstruction-optimized lapped orthogonal transforms for robust image transmission," *IEEE Trans. on Circ.& Sys. Video Tech.*, vol. 6, no. 2, pp. 168–181, April 1996.

[5] D. Chung and Y. Wang, "Lapped orthogonal transform designed for error resilient image coding," *IEEE Trans. on Circ. & Sys. Video Tech.*, pp. 752–764, Sep. 2002.