DEBLOCKING OF VIDEO SEQUENCES WITH LAPPED EMBEDDED IDCT

Lorenzo Cappellari, Truong Q. Nguyen

Dept. of Electrical and Computer Engineering
University of California, San Diego - La Jolla, CA 92093-0407
kappa@dei.unipd.it, nguyent@ece.ucsd.edu

ABSTRACT

At mid to low bit-rate the standard DCT transform based video codecs produce in the reconstructed sequences annoying artifacts known as ringing and blocking effects. In this work we are concerned in blocking artifact removal with lapped embedded IDCT. In intra frames the deblocking always produces an increase in the peak signal-to-noise ratio; as a consequence of the processing on the last reference intra frame, inter frames, whose prediction error is decoded with IDCT, are free of blocking artifacts. Using the proposed algorithm on the inter frame decreases the resulting PSNR while improving the visual quality of the decoded images. In this paper, we present a theoretical bound on the PSNR reduction.

1. INTRODUCTION

Most image/video compression standards produce blocking artifacts when the bit-rate is reduced since they use DCT-IDCT pair to achieve compression. Lapped transforms that use the DCT as front-end are introduced to reduce this effect and take advantage of low complexity DCT algorithms [1], but modifications in both encoder and decoder are required.

The lapped embedded IDCT [2] has been used effectively in deblocking, avoiding modifications in the standard encoder. But, in video sequences, it prevents correct frame prediction and produces an increased error on inter frames. The loss in PSNR is theoretically analysed to show the degradations are negligible, in particular at low bit-rates.

Sections 2 and 3 review the lapped embedded IDCT and the weighting process used for deblocking. Section 4 presents theoretical analysis for the deblocking algorithm, which is tested on MPEG-2 and MPEG-4 standards in section 5. Section 6 provides comparisons of experimental and theoretical results. Conclusions are drawn in section 7.

2. THE LAPPED EMBEDDED INVERSE DISCRETE COSINE TRANSFORM

All transforms in image/video standards have linear phase and orthogonality which simplify border processing as well as preserve the signal energy in the transformed domain. The polyphase matrix of any linear-phase paraunitary filter bank (LPPUFB) with even number of filters M sharing the same center of symmetry and with the same length equal to $(N+1)M$ can be always decomposed as [3][4]

$$\mathbf{E}(z) = \prod_{i=N}^{1} \left(\mathbf{\Phi}_i \mathbf{W} \mathbf{\Lambda}(z) \mathbf{W}\right) \mathbf{E}_0 \quad . \tag{1}$$

$\mathbf{E}_0$ is an orthogonal matrix whose half rows have even symmetry and the others have odd symmetry, while

$$\mathbf{\Lambda}(z) = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & z^{-1}\mathbf{I} \end{bmatrix} \quad , \quad \mathbf{W} = \frac{1}{\sqrt{2}} \begin{bmatrix} \mathbf{I} & \mathbf{I} \\ \mathbf{I} & -\mathbf{I} \end{bmatrix} \tag{2}$$

and $\mathbf{\Phi}_i = diag\{\mathbf{U}_i, \mathbf{V}_i\}$, with $\mathbf{U}_i$ and $\mathbf{V}_i$ $M/2 \times M/2$ orthogonal matrices. A particular case of LPPUFB is obtained using DCT as the analysis front-end $\mathbf{E}_0$. This transform is a more general case of the LOT introduced in [1].

The polyphase matrix of the synthesis FB for causal reconstruction after analysis with LPPUFB has the form

$$\mathbf{R}(z) = \mathbf{E}_0^{-1} \prod_{i=1}^{N} \left(\mathbf{W} z^{-1} \mathbf{\Lambda}^{-1}(z) \mathbf{W} \mathbf{\Phi}_i^{-1}\right) \quad . \tag{3}$$

Hence, if $\mathbf{E}_0^{-1}$ is the inverse DCT, the DCT and the following synthesis FB form a LPPUFB

$$\mathbf{R}'(z) = \mathbf{R}(z) \prod_{i=N}^{1} \left(\mathbf{\Phi}_i \mathbf{W} \mathbf{\Lambda}(z) \mathbf{W}\right) \quad . \tag{4}$$

This synthesis FB, shown in Fig. 1, is called *lapped embedded inverse discrete cosine transform* (le-IDCT) [2].

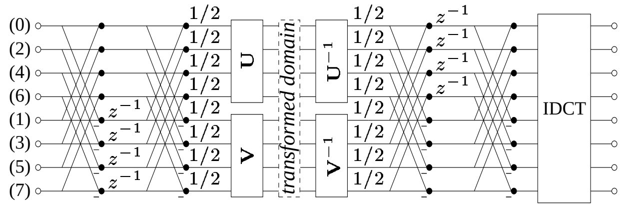

Fig. 1. Lattice structure of single stage le-IDCT.

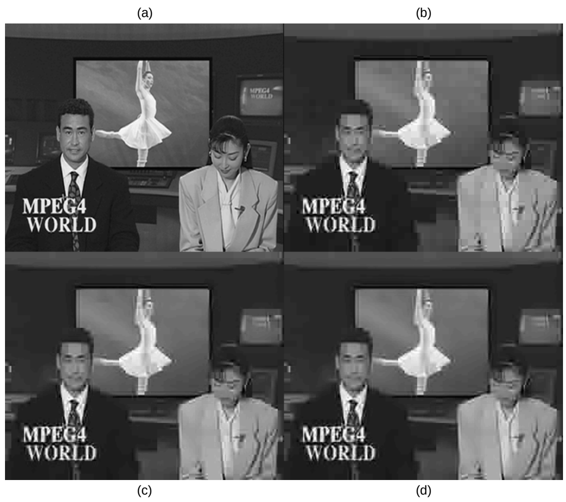

Fig. 2. Frame #120 (intra) of the sequence *news* (CIF resolution): (a) original frame, (b) standard decoder output; (c) deblocked frame (low threshold), (d) deblocked frame (high threshold).

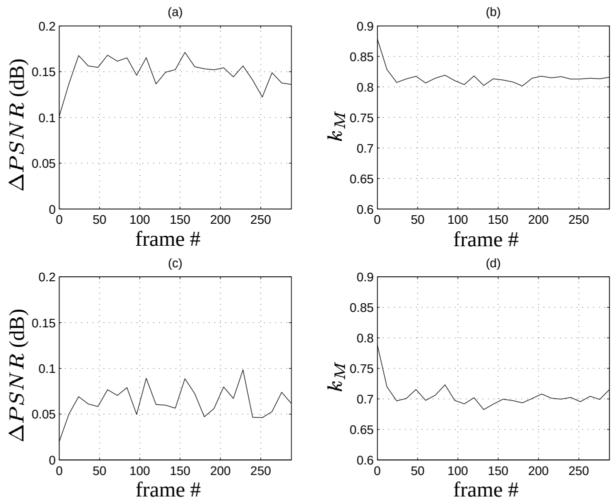

Fig. 3. Objective measurements for intra frames of the sequence *news*: (a) and (b) for low threshold; (c) and (d) for high threshold.

3. BLOCKING ARTIFACT REMOVAL

Without processing in the transformed domain le-IDCT is equivalent to IDCT. Also if better design strategies can be used, the single stage le-IDCT ($N = 1$) designed to maximise coding gain [1] produces synthesis functions that decay gracefully toward zero at their ends. Hence, the $M/2$ coefficients corresponding to odd functions measure blockishness introduced by quantization in the DCT domain. To remove the artifacts, the coefficient vector $\mathbf{y}$ is weighted in the lapped transform domain through multiplication by

$$\mathbf{A}(\mathbf{y}, \varepsilon) = diag\left\{1, \alpha\left(y_1, \varepsilon\right), 1, \alpha\left(y_3, \frac{\varepsilon}{2}\right), 1, 1, 1, 1\right\} \quad , \tag{5}$$

which represents a non-linear operation since

$$\alpha(y, \gamma) = \begin{cases} \frac{1}{2} & |y| \leq \gamma \\ 1 & |y| > \gamma \end{cases} \quad . \tag{6}$$

Details on this non-linear weighting can be found in [2].

4. THEORETICAL ANALYSIS

This section presents a one-dimensional analysis assuming that the blocked signal is a stationary Gauss-Markov process with zero mean, variance σ^2, and correlation coefficient ρ. Consequently each subband, which is then eventually weighted, is a zero-mean gaussian process. Although the analysis below only considers the one-dimensional case, the results can be generalized to two-dimensional case by using two-dimensional Gauss-Markov processes and separable filter banks.

To represent the non-linear weighting used in the lapped transform domain, we use an additive model, because the weighted signal $\hat{y}_n$ can be expressed as $\hat{y}_n = y_n + e_n$. The *weighting noise* e_n is related to the incoming samples of the signal y_n through the function

$$e(y) = \begin{cases} -\frac{y}{2} & |y| \leq \varepsilon \\ 0 & |y| > \varepsilon \end{cases} \quad . \tag{7}$$

It therefore has zero mean and its variance σ_e^2 is related to the variance of the subband σ_y^2 through

$$\frac{\sigma_e^2}{\sigma_y^2} = \frac{1}{4}\left[1 - 2Q\left(\frac{\varepsilon}{\sigma_y}\right) - 2\frac{\varepsilon}{\sigma_y}\varphi\left(\frac{\varepsilon}{\sigma_y}\right)\right] \triangleq \beta\left(\frac{\varepsilon}{\sigma_y}\right) \quad , \tag{8}$$

where $\varphi(a)$ and $Q(a)$ are respectively the normalized gaussian *pdf* and the complementary gaussian distribution function.

From previous computations, it is easy to obtain

$$\rho_{ey} = \frac{E[e_n y_n]}{\sigma_e \sigma_y} = -2\frac{\sigma_e}{\sigma_y} \quad , \tag{9}$$

which shows that, like quantization noise, the error added to the signal and the signal are negatively correlated each other. If $\varepsilon \to 0$ the error tend to be uncorrelated with the signal; numerical analysis shows that we can assume the error and the signal quite uncorrelated only if $\varepsilon \lesssim 0.1\sigma_y$.

Denote the autocorrelation matrix of the blocked signal by $\mathbf{R}_{xx}(\rho)$ and the transform matrix of subband decomposition by $\mathbf{P}$, the variances of the subbands are thus the diagonal elements of $\mathbf{R_y}(\rho) = \mathbf{P}\mathbf{R}_{xx}(\rho)\mathbf{P}^T$, each of which

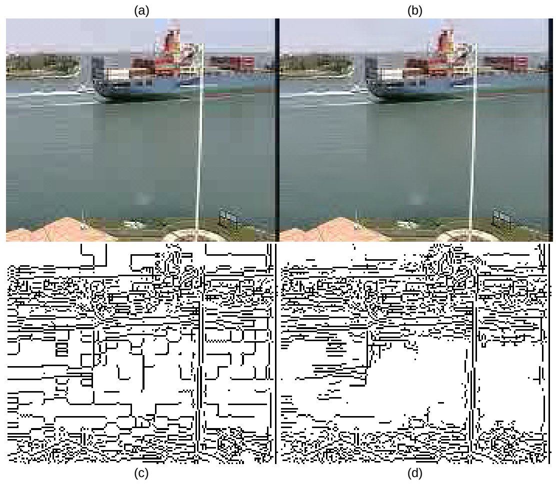

Fig. 4. Frame #225 (intra) of the sequence *container* at 10 fps and 40 kbps (QCIF resolution): (a) standard frame, (b) deblocked frame, (c) and (d) respective edge maps.

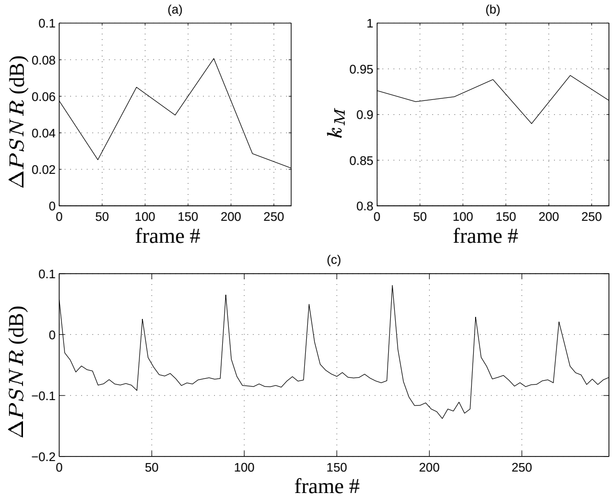

Fig. 5. Measurements for the sequence *container* at 10 fps and 40 kbps: (a) and (b) only for intra frames; (c) for all frames.

can be related to the blocked signal variance according to $\sigma_i^2 = \alpha_i(\rho)\sigma^2$. Applying our additive model, we obtain the error on each weighted subband has variance

$$\sigma_{e,i}^2 = \beta\left(\frac{\varepsilon_i}{\sigma_i}\right)\sigma_i^2 = \beta\left(\frac{\varepsilon_i}{\sigma_i}\right)\alpha_i(\rho)\sigma^2 \quad , \qquad (10)$$

which leads to

$$\bar{\sigma}_e^2 = \sigma^2\frac{1}{8}\left[\alpha_1(\rho)\beta\left(\frac{\varepsilon}{\sigma_1}\right) + \alpha_3(\rho)\beta\left(\frac{\varepsilon}{2\sigma_3}\right)\right] \qquad (11)$$

as the averaged variance of weighting error in the reconstructed signal. This error is correlated with the blocked signal, and its variance can not be added to the quantization noise variance to obtain the total error variance. In any case, the sum is an upper bound for the exact variance.

The question now is: what happens to weighting error in inter frames? The predicted signal $\hat{x}'_n$ is obtained from the previous corrupted signal $\tilde{x}'_n = \tilde{x}_n + e_n$, where $\tilde{x}_n$ is the standard reconstructed frame and e_n is the noise added by the deblocking algorithm. If d_n is the displacement vector used for each n to do the prediction, we obtain

$$\hat{x}'_n = \tilde{x}'_{n-d_n} = \tilde{x}_{n-d_n} + e_{n-d_n} = \hat{x}_n + e_{n-d_n} \quad , \qquad (12)$$

where $\hat{x}_n$ is the prediction made without deblocking algorithm. Hence, in inter frames the variance of the error is not modified. Moreover, in interframe, we can assume that the errors are more uncorrelated with the signal and quantization noise and therefore, the upper bound obtained by summing up all error variances is reasonably accurate.

Similar considerations hold in the case of two-dimensional displacements, but we must pay attention at sub-pixel accurancy. In fact, bilinear interpolation used to achieve sub-pel accurancy reduces the variance of the weighting error. Suppose that ρ_e is the autocorrelation on both rows and columns of the weighting error $e_{n,m}$, for half-pel accurancy of the displacement on one or both directions the error variances are respectively $\sigma_e^2\left(\frac{1+\rho_e}{2}\right)$ and $\sigma_e^2\left(\frac{1+\rho_e}{2}\right)^2$, which are both less than σ_e^2 for any $0 \le \rho_e < 1$.

5. EXPERIMENTAL RESULTS

The deblocking algorithm is applied in both MPEG-2 and MPEG-4 video compression standards, modifying the respective implementations by *MSSG* and ACTS *MoMuSys* project. To verify the derived theoretical bound, the variation of PSNR ($\Delta PSNR$) and the ratio of MSDS [5] (k_M) are measured for several video sequences.

The results on the MPEG-2 compressed sequence *news*, decompressed with deblocking at two different thresholds ε are reproduced in Fig. 2, which clearly demonstrate the algorithm effectiveness. Fig. 3 shows that increasing the threshold ε in the processing step reduces both $\Delta PSNR$ and MSDS ratio.

In another set of experiments, we compressed several test sequences with MPEG-4 at various bit-rates and resolutions, using always GOVs of 15 VOPs without bidirectional frames. Fig. 4 shows the results on an intra frame of the sequence *container* with the respective luminance edge maps. As shown in Fig. 5(a,b), there are improvements in both PSNR and MSDS when only intra frames are processed. However, the PSNR degrades as one also processes inter frames (Fig. 5c), in agreement with the results in section 4.

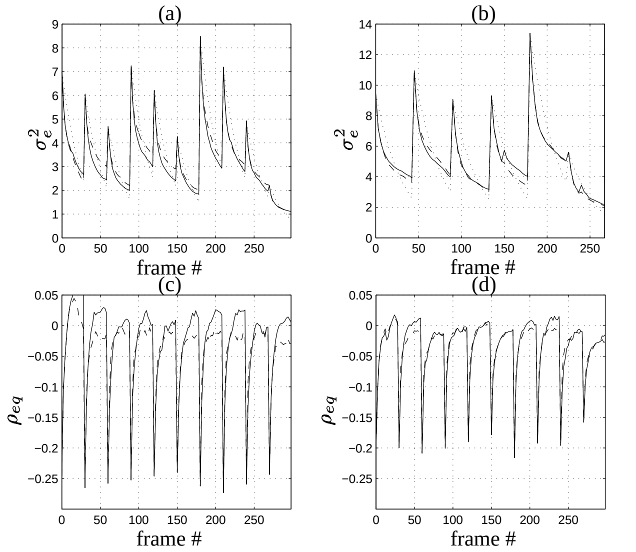

Fig. 6. (a,b): weighting error variance for the sequences *silent* (15 fps, 256 and 384 kbps) and *news* (10 fps, 40 and 56 kbps); the dotted line represents the theoretical approximation. (c,d): cross correlation of weighting and quantization error for the sequences *mother* and *silent* (CIF, 15 fps, solid and dashed line for 256 kbps and 384 kbps).

6. COMPARISON WITH THEORETICAL RESULTS

The Gauss-Markov model used in section 4 for blocked intra frames is the best approximation of our data as long as the correct correlation coefficients are used in the analysis. From experimental data we assume $\rho = 0.97$ and $\rho = 0.95$ respectively for CIF and QCIF resolution, for both rows and columns (as example, for the sequence *silent* we have respectively $\rho_{row} = 0.976$, $\rho_{col} = 0.972$ and $\rho_{row} = 0.947$, $\rho_{col} = 0.945$), and $\sigma = 255/6$. Since the threshold ε is known, the error variances are then computed. In Fig. 6(a,b) we reported the error variance for two sequences at CIF and QCIF resolution. The theoretical approximation is obtained assuming equiprobability of integer and half-pel displacements on both directions, with $\rho_e = 0.6$ and $\rho_e = 0.5$ for CIF and QCIF resolution, as measured on the average in our experiments. Moreover, Fig. 6(c,d) shows that, as expected, weighting error is correlated with quantization error only for intra frames. Therefore, in inter frames the extimated error variance is almost equal to the measured value.

The average PSNR of the standard decoded sequence is relative to the average quantization error variance. As the variance of the weighting error is theoretically known, we immediately derive the loss of PSNR produced. The results we obtained, compared with the experimental data, are showed in Fig. 7.

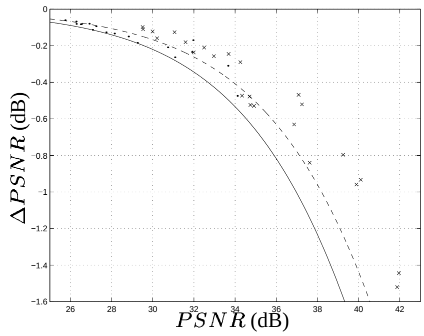

Fig. 7. Relationship between mean PSNR and mean $\Delta PSNR$ for different sequences at QCIF resolution (•) and CIF resolution ($\times$). The solid and dashed lines represent the respective theoretical bound for the two resolutions.

7. CONCLUSIONS

The presented algorithm reconstructs intra frames with alleviated blockishness and increased PSNR, while it seems to reduce quality for inter frames. The derived bound for the loss of PSNR, which is not only a lower bound, but also a good approximation, demonstrates the mean loss in PSNR is low for sequences already highly corrupted, i.e. with low PSNR. Moreover, an eventual loss in PSNR must not necessarily be interpreted as loss of perceived quality, since the PSNR, in particular for low qualities, can be misleading.

8. REFERENCES

[1] H.S. Malvar and D.H. Staelin, "The lot: Transform coding without blocking effects," *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol. 37, no. 4, pp. 553–559, Apr. 1989.

[2] S. Yang S. Kittitornkun Y. Hu T.Q. Nguyen and D.L. Tull, "Blocking artifact free inverse discrete cosine transform," *IEEE Proc. of International Conference on Image Processing*, vol. 3, pp. 872–875, 2000.

[3] A.K. Soman P.P. Vaidyanathan and T.Q. Nguyen, "Linear phase paraunitary filter banks: Theory, factorizations and designs," *IEEE Trans. on Signal Processing*, vol. 41, no. 12, pp. 3480–3496, Dec. 1993.

[4] R.L. De Queiroz T.Q. Nguyen and K.R. Rao, "The genlot: Generalized linear-phase lapped orthogonal transform," *IEEE Trans. on Signal Processing*, vol. 44, no. 3, pp. 497–507, Mar. 1996.

[5] S. Minami and A. Zakhor, "An optimization approach for removing blocking effects in transform coding," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 5, no. 2, pp. 74–82, Apr. 1995.