OPTIMAL PSNR ESTIMATED SPECTRUM ADAPTIVE POSTFILTER FOR DCT CODED IMAGES

Irving Linares

School of Electrical Engineering
Georgia Institute of Technology
Atlanta, GA 30332

ABSTRACT

Our previous work aimed at reducing the blocking distortion of DCT coded images is further improved by introducing the Optimal PSNR *Estimated Spectrum Adaptive Postfilter* (ESAP) Algorithm. ESAP searches for a one dimensional log-sigmoid weighting function which, when applied to the separable, interpolated local block estimated spectrum of the coded image, minimizes the Mean Square Error (MMSE) with respect to the original image using a 2-D steepest descent search. Convergence is obtained in a few iterations for integer parameters. A unique maximum PSNR is guaranteed given the asymptotic exponential overshoot behavior of the surface generated by ESAP. We obtained PSNR improvement of 1.5 dB over nonpostfiltered JPEG images as well as subjective improvement. ESAP is based on a DFT analysis of the DCT basis functions and uses spatially adaptive separable FIR postfilters.

1. INTRODUCTION

There have been a number of methods to reduce DCT blocking [1, 2, 3, 4]. In [5] we proposed the *Residual Spectra Adaptive Postfilter*. RSAP is an *optional* method for image enhancement external to a JPEG Coder. It improves the subjective quality and increases the objective PSNR measure over non-postfiltered images by 1.0 dB *without* increasing the bit rate.

We now further develop RSAP into the ESAP[1] algorithm (Figure 1) and apply it to 512x512 images. ESAP iteratively searches for the *optimal* adaptively postfiltered image with the MMSE with respect to the original image at the encoder using a *steepest descent* search. A one-dimensional log-sigmoid weighting function is separably applied to both the interpolated vertical and interpolated horizontal frequency images. Once the integer *steepness* and *bias* values associated with the optimal log-sigmoid frequency weighting curve are found, they are transmitted as a negligible single overhead byte to the decoder. The high four bits represent a [-3..12] steepness range while the low four bits contain a [-3..12] bias range.

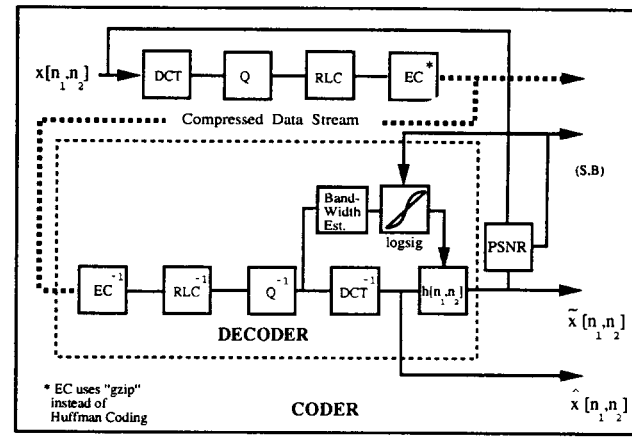


Figure 1: Block diagram of the *Estimated Spectrum Adaptive Postfilter* Algorithm extension to the JPEG DCT coder where $\tilde{x}$ is the postfiltered image and $\hat{x}$ is the non-postfiltered decoded JPEG image.

2. ESTIMATED SPECTRUM ADAPTIVE POSTFILTER

The waveform of the DFTs of each DCT basis function can be derived by using the Modulation Theorem

$$x[n]w[n] \overset{\mathcal{DFT}}{\longleftrightarrow} \frac{1}{N} X(k) \circledS W(k) \qquad (1)$$

Multiplication of a *cosine* function by a rectangular

This work was supported by the NASA Goddard Space Flight Center Part-Time Graduate Study Program.

[1] In order to avoid confusion with the standard usage of the term *Residual* which implies a *difference*, the modified algorithm is now renamed *Estimated* Spectrum Adaptive Postfilter to indicate a localized spectral estimation directly obtained from the DCT coefficients.

0-7803-2431-5/95 $4.00

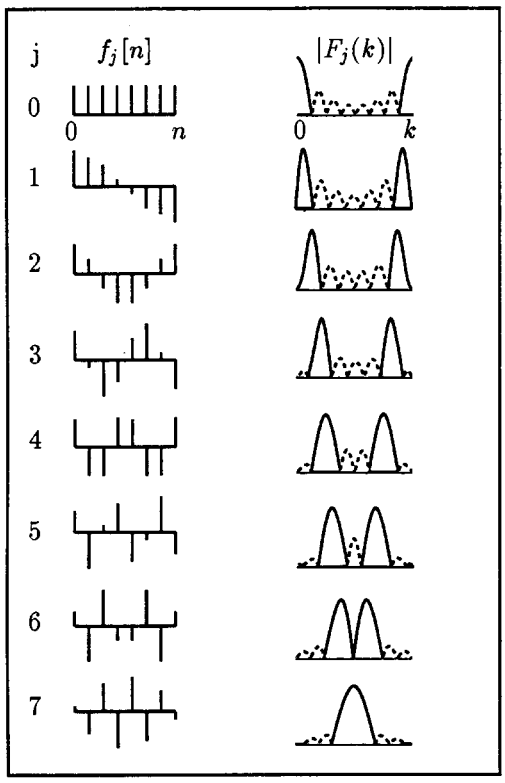
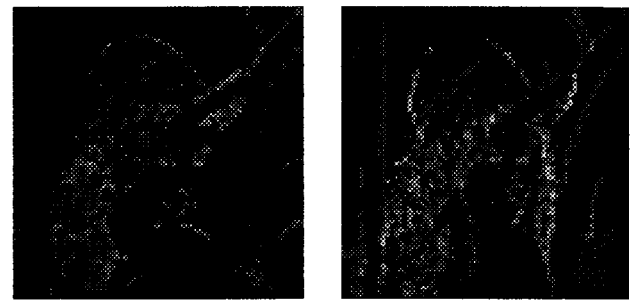


Figure 2: DCT Basis Functions $f_j[n]$ and the magnitudes of their 256-point DFTs $|F_j(k)|$, $j = 0..7, N = 8$.

window in the time domain is equivalent to the circular convolution of a sinc-shaped rectangular window transform $W(k)$ with an ideal pair of impulses $N\delta(k \pm k_0)$ producing the $|F_j(k)|$ waveforms shown in Figure 2. A DFT analysis (Sect 7.4 [6],[7]) in conjunction with Figure 2 reveals that the *ripples* or sidelobes of the DFT of each DCT basis function, shown as *dashed* lines, represent the DCT block discontinuities for any particular spatial frequency. The window's width affects the main lobe frequency localization resolution and simultaneously introduces ripples. The ripples contain most of the *out-of-band* blocking distortion while the main lobes represent the *in-band* main frequency content. Therefore, if we extend the analysis to 2-D and neglect aliasing, removal of the DCT spatial blocking is achieved by filtering the ripple in the frequency domain.

To obtain a local spectral estimate $\hat{X}_i(k_1, k_2)$ we look at the dequantized coefficient values $\hat{C}_i(k_1, k_2)$ for each 8x8 block i of the decoded image $\hat{x}[n_1, n_2]$. The current block spectral bandwidth is found by inspection of the *highest* 2-D frequency present in a block. Some of the intermediate coefficients could be 0. Based on the fact that decoded blocks mostly contain low frequency coefficients we get the coefficient-bandwidth relationship shown in Table 1. For example, if the decoded block's highest coefficient is $C(2,5)$, we say in a deterministic way, that the *vertical* bandwidth is 0.375π and the *horizontal* bandwidth is 0.750π. This 2-D local bandwidth, denoted $L_B(i)$ is spatially centered in the middle of the 8x8 block. We obtain two 64x64 frequency images for 512x512 pixel images: the Non-Interpolated DCT *Vertical* Frequency Image and the Non-Interpolated DCT *Horizontal* Frequency Image shown in Figures 3(a) and 3(b) respectively.



(a) (b)

Figure 3: (a) 64x64 Non-Interpolated Vertical Frequency (b) 64x64 Non-Interpolated Horizontal Frequency for Lena image @ 0.25 BPP.

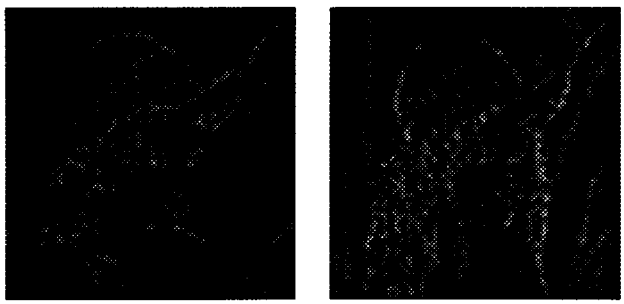
Table 1. Coefficient-Block Bandwidth relationship.

DCT Coeffs. Present	Normalized Bandwidths
c_0	0.125π
c_0, c_1	0.250π
c_0, c_1, c_2	0.375π
c_0, c_1, c_2, c_3	0.500π
c_0, c_1, c_2, c_3, c_4	0.625π
$c_0, c_1, c_2, c_3, c_4, c_5$	0.750π
$c_0, c_1, c_2, c_3, c_4, c_5, c_6$	0.875π
$c_0, c_1, c_2, c_3, c_4, c_5, c_6, c_7$	1.0π

Each non-interpolated frequency image is then interpolated to the original 512x512 size by using

$$\hat{B}_{\omega_1,\omega_2}[n_1, n_2] = 64 \sum_{i=0}^{(\frac{N}{8})^2-1} L_B(i) \times \sum_{m_1=-127/2}^{127/2} \frac{\sin\frac{\pi}{8}(n_1 + m_1 - d_{1_i})}{\pi(n_1 + m_1 - d_{1_i})} \sum_{m_2=-127/2}^{127/2} \frac{\sin\frac{\pi}{8}(n_2 + m_2 - d_{2_i})}{\pi(n_2 + m_2 - d_{2_i})} \quad (2)$$

where (m_1, m_2) are restricted to the mid-block values (...-5/2, -3/2, -1/2, 1/2, 3/2, 5/2, ...) and (d_{1_i}, d_{2_i}) are displacement values for each mid-block point given by $d_{1_i} = (i/64) * 8 + 3.5$ and $d_{2_i} = (i * 8) \bmod 512 + 3.5$. The division $i/64$ is an *integer* truncated division.



(a) (b)

Figure 4: Log-Sigmoid (a) 512x512 Interpolated Vertical Frequency (b) 512x512 Interpolated Horizontal Frequency for Lena image @ 0.25 BPP.

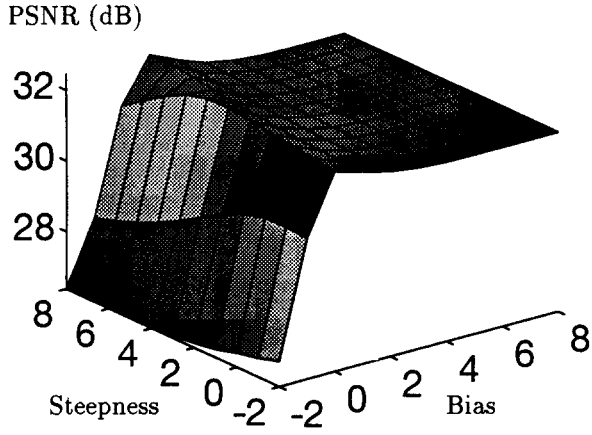


Figure 5: ESAP PSNR as function of Steepness and Bias for the 512x512 JPEG Lena's image @ 0.25 BPP.

Next both interpolated frequency images are weighted with the log-sigmoid [8] function

$$logsig(s,\omega,b) = \frac{1}{1+e^{-[s(\omega-1/2)+b]}} \qquad (3)$$

where s and b are the steepness and bias to be determined during encoding and w is the normalized frequency [0..1]. This weighting is iteratively repeated over the PSNR surface until the maximum PSNR is obtained. The optimal (S,B) values are transmitted to the decoder as a byte of side information. The result of this operation is shown in Figures 4(a) and 4(b).

Figure 5 shows the Lena ESAP surface as a function of steepness and bias at a 32:1 Compression Ratio. Note the optimal (S,B) pair at (3,0). For comparison, Figure 6 shows the optimal (3,0) logsig function and the previous RSAP frequency weighting.

The interpolated local 2-D bandwidth $\hat{B}_{\omega_1,\omega_2}[n_1,n_2]$ obtained from (2) and weighted by (3) allows a pixel-by-pixel adaptive convolution using a set of 128 precomputed Non-causal Hamming Window 1-D Low Pass Filters. ESAP convolves the current $\hat{x}[n_1,n_2]$ pixel's region with the 2-D non-causal separable adaptive LPF $h[n_1,n_2] = h_{l_1}[n_1]\ h_{l_2}[n_2]$ where $1 \le l_{1,2} \le 128$ and whose cutoffs are $\hat{\omega_1} = (l_1/128)\pi$, $\hat{\omega_2} = (l_2/128)\pi$ to obtain the optimal PSNR postfiltered image. Figure 7 shows a non-postfiltered JPEG decoded image. Figure 8 shows the optimal ESAP image.

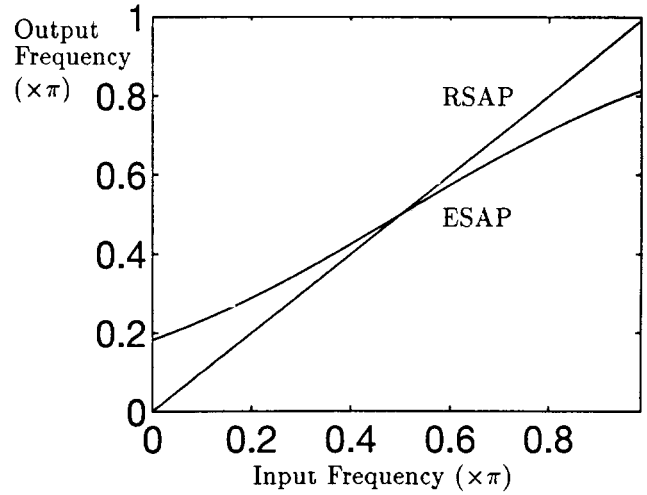


Figure 6: Log-Sigmoid frequency weighting function.

3. RESULTS

The following PSNR[2] results were obtained for the 8-BPP grey-scale test images below. All images are 512x512 except Peppers which is 256x256.

Table 2. ESAP PSNR Values

Image	BPP	PSNR	(S,B)	Δ
Peppers NPF[3]	0.25	30.89		
Peppers ESAP[4]	0.25	32.39	(4,0)	+1.50
Lena NPF	0.25	31.46		
Lena ESAP	0.25	32.45	(3,0)	+0.99
Fruits NPF	0.25	30.01		
Fruits ESAP	0.25	30.08	(9,2)	+0.07
New Orleans NPF	0.33	32.69		
New Orleans ESAP	0.33	33.46	(2,0)	+0.77
Wash D.C. NPF	0.33	34.19		
Wash D.C. ESAP	0.33	34.68	(2,0)	+0.49

[2] $\text{PSNR} = 10\log\frac{(N)^2(255)^2}{\sum_{n_1=0}^{N-1}\sum_{n_2=0}^{N-1}(x[n_1,n_2]-\hat{x}[n_1,n_2])^2}$ dB, where x is the original image, $\hat{x}$ is the decoded image and N is the image size.

[3] NPF = No Post-Filtering, JPEG decoded only.

[4] ESAP = Optimal PSNR Estimated Spectrum Adaptively Postfiltered using 17x17 FIR filter.

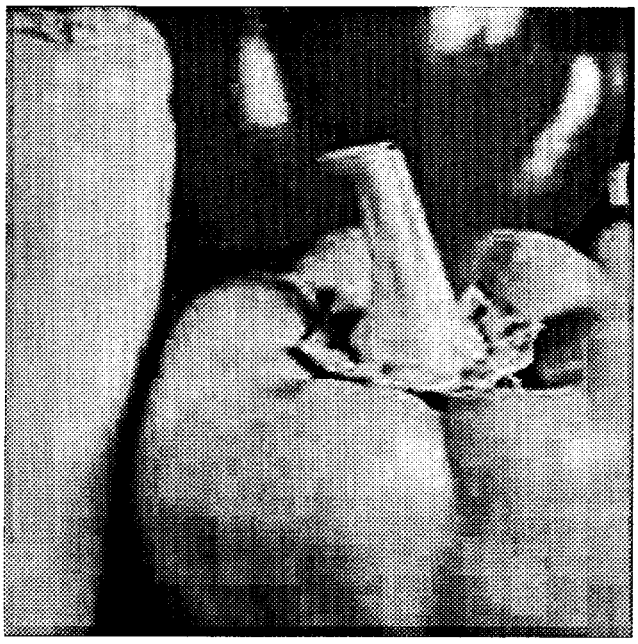


Figure 7: JPEG coded Peppers image @ 0.25 BPP, 30.89 dB PSNR.

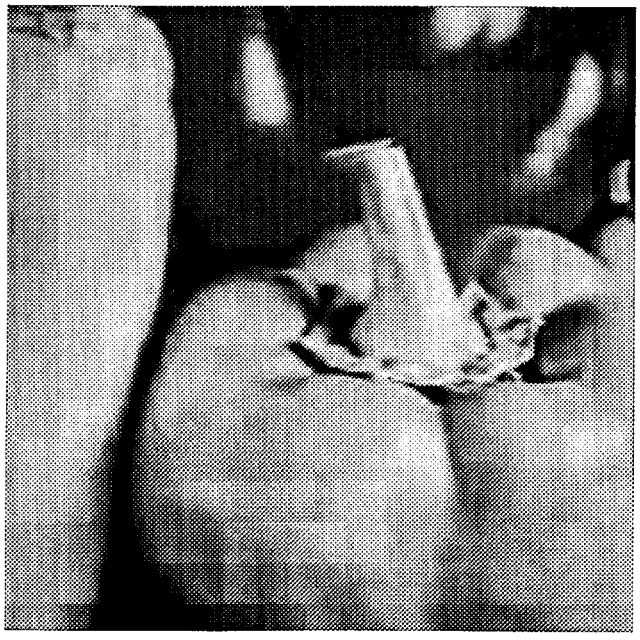


Figure 8: Optimal PSNR ESAP Peppers image @ 0.25 BPP, 32.39 dB PSNR.

4. CONCLUSION

ESAP reduces DCT blocking. Blocking appears as side lobes in the frequency domain. ESAP filters the side lobes to minimize blocking and to improve the PSNR and perceived quality. ESAP reuses each block's DCT coefficients to estimate its frequency content. This information is interpolated, then a steepest descent search for the MMSE (maximum PSNR) log-sigmoid frequency weighting function is performed at the encoder. The optimal postfilter is completely characterized by the dequantized DCT coefficients and the optimal (S,B) parameters transmitted to the decoder as a negligible single byte. ESAP exploits the Human Visual System frequency masking characteristics. Based on its tolerance to quantization errors in the high frequency regions ESAP performs almost no filtering there while it adaptively filters low frequency regions to reduce blocking and restore the smoothness without significantly reducing the crispness of the image.

PSNR improvements of 1.5 dB can be obtained with ESAP. These compare favorably to other blocking reduction algorithms [1-5]. ESAP is an asymmetric postfiltering technique. It can converge to the optimal PSNR in as few as nine iterations with the proper choice of initial integer (S,B) parameters. However, it requires only *one* iteration for *any* image at the decoder. ESAP's current computational complexity is $O(N^4/4)$ for the initial local pixel frequency interpolation and $O((MN)^2)$ adds, $O(N^2M^2/4)$ multiplications for each iteration of the adaptive convolution (N is the image size and M is the filter size). ESAP showed its robustness, data independence and adaptability by improving the PSNR of the decoded images and by effectively removing blocking while preserving important edge information.

5. REFERENCES

[1] Reeve and Lim, "Reduction of Blocking Effect in Image Coding", *ICASSP'83*, Boston MA

[2] Ramamurthi and Gersho, "Nonlinear Space-Variant Postprocessing of Block Coded Images" *Transactions* ASSP, Vol.34, October 1986

[3] Malvar and Staelin, "The LOT: Transform Coding Without Blocking Effects", *Transactions* ASSP, Vol. 37, 1989

[4] Zeng and Venetsanopoulos, "A JPEG-Based Interpolative Image Coding Scheme" *ICASSP'93*, Sect. V, pp. 393-396

[5] Linares, Mersereau and Smith, "Enhancement of Block Transform Coded Images Using Residual Spectra Adaptive Postfiltering", *Proceedings* DCC'94, pp. 321-330

[6] Oppenheim and Schafer, "Discrete-Time Signal Processing", Prentice Hall, 1989

[7] Clarke, "Spectral response of the discrete cosine and Walsh-Hadamard transforms", IEE *Proceedings*, Vol. 130, Pt. F, No. 4, June 1983

[8] Demuth and Beale, "Neural Network Toolbox Users's Guide", The MathWorks, Inc. 1992, p. 9-51