HVS-Weighted Progressive Image Transmission Using the Lapped Orthogonal Transform

Ricardo L. de Queiroz and *K. R. Rao*

Electrical Engineering Department, University of Texas at Arlington
Box 19016, Arlington, TX, 76019
email: queiroz@eepost.uta.edu or eekrr521@utacnvx.uta.edu

Abstract – Progressive transmission of images based on lapped orthogonal transform (LOT), adaptive classification and human visual sensitivity (HVS) weighting is proposed. HVS weighting for LOT is developed using a general technique that can be applied to any orthogonal transform. The method is compared with discrete cosine transform (DCT) based progressive image transmission (PIT). It is shown that LOT based PIT yields subjectively improved images compared to those based on DCT.

1 Introduction

Progressive image transmission (PIT) [1] based on transform domain prioritization has gained wide acceptance [2–7] due to various adaptive features such as classification [8–12], spectral selection [4,6], HVS weighting [2,6,13–15], as well as due to the VLSI development of operations such as transform, quantization, and variable length coding. In addition, PIT based on the DCT has been extensively investigated. For example, JPEG (Joint Photographic Experts Group) [6] algorithms for the baseline and extended systems are DCT based. At low bit rates , however, DCT introduces block structure in the reconstructed images [2]. One technique to reduce or eliminate this artifact is replacing DCT by the LOT [16,17] whose basis vectors overlap across traditional block boundaries. It is intuitively felt that LOT based PIT should yield subjectively more pleasing pictures even during the initial stages compared to the DCT. This is the objective of this paper. i.e., develop a LOT-PIT incorporating various adaptive features and compare with the DCT dependent PIT.

2 PIT with the Chen-Smith coder

The Chen-Smith coder [9] is based on the zonal sampling strategy. First, the image undergoes an orthogonal transform. The transform coefficients are stored in a buffer and some statistics are computed prior to the decision of: (i) which coefficients are transmitted; (ii) how to quantize these coefficients; (iii) the order of transmission. We will assume the image as having $N \times N$ picture elements (pixels or pels). The encoding steps can be briefly described as:

•Transform the image using blocks of $M \times M$ pels. Let $N_B = (N/M)^2$ be the total number of blocks in the image. The blocks are, then, labelled from 1 to N_B in a lexicographical order, with coefficients given as $x_i(u,v)$ for $i = 1, \ldots, N_B$ and $(u,v) \in \{(0,0), \Psi\}$, where Ψ is defined as a set of $M^2 - 1$ block-index pairs, as $\Psi \equiv \{(0,1),(0,2)\ldots,(0,M-1),(1,0),(1,1)\ldots,(M-1,M-1)\}$.

•Quantize and code separately the coefficients $x_i(0,0)$ (the DC coefficents) using uniform quantizers.

•Compute the AC energy of each block E_i as

$$E_i = \sum_{(m,n)\in\Psi} x_i^2(m,n) \qquad (1)$$

Sort the energies, and classify the blocks (in sorted order) into N_C equally populated classes [12], each with N_B/N_C blocks. Construct the *classmap* $C(i)$ with the classification of each block. $C(i) = k$ indicates the i-th block belongs to class k.

•For all blocks belonging to the same class, construct N_C *deviation maps* with the standard deviation of the coefficients, found from (δ is the Kronecker delta function)

$$\sigma_k^2(m,n) = \sum_{i=1}^{N_B} \delta(C(i) - k)\; x_i^2(m,n) \qquad (m,n) \in \Psi \qquad (2)$$

•Merge all N_C *deviation maps* and decide the bit-allocation. Based on the rate-distortion theory, we shall iteratively find a distortion value D and a set of integers $B_k(m,n)$ (for $(m,n) \in \Psi$ and $1 \leq k \leq N_C$), so that

$$B_k(m,n) = \frac{1}{2}\log_2\left[\sigma_k^2(m,n)\right] - \log_2(D) \qquad (3)$$

$$\sum_{k=1}^{N_C-1} \sum_{(m,n)\in\Psi} B_k(m,n) = (RN^2 - B_{ov})\frac{N_C}{N_B} \qquad (4)$$

for $0 \leq B_k(m,n) \leq B_{max}$, where B_{max} is the maximum number of bits allowed, B_{ov} is the number of bits required

0-7803-0950-2/93/$3.00©1993IEEE

for overhead transmission, and R is the average bit-rate in bits/pel (bpp). Create N_C bit-allocation maps with a one-to-one correspondence with the elements of the *deviation maps*.

•Re-estimate the standard deviations using the bit allocation maps.

$$\hat{\sigma}_k(m,n) = c\, 2^{B_k(m,n)-1} \quad 1 \leq k \leq N_C \quad (m,n) \in \Psi \quad (5)$$

where c is a normalization factor. It was suggested in [9] to choose c as the maximum $\sigma_k(m,n)$ for which $B_k(m,n) = 1$, for avoiding excessive clipping.

•Send $C(i)$, c, and $B_k(m,n)$ as side information.

•Quantize, encode and send all the coefficients, using the re-estimated variances. A coefficient $x_i(m,n)$ (block i), which belongs to class k ($C(i) = k$), is scaled (divided by $\hat{\sigma}_k(m,n)$), applied to a quantizer with $2^{B_k(m,n)}$ levels and encoded with $B_k(m,n)$ bits. If $B_k(m,n) = 0$, the particular coefficient is not transmitted.

The receiver may first decode the side information and the DC coefficients. Given the class map, the bit-allocation maps and the normalization factor c, the decoder can reconstruct the standard deviations used to scale the quantizers as in (??). With the maps reconstructed, and with the knowledge of the transmission order, the decoder can exactly determine the position of the incoming coefficient and its class, bit-assignment and variance. Therefore, the receiver can decode all coefficients, apply an inverse transform and obtain the image. The overhead is made by the class map, the bit allocation maps and c. Quantizing c with 16 bits, the total amount of overhead is given by:

$$B_{ov} = N_B \log_2(N_C) + N_C(M^2-1)\log_2(B_{max}+1) + 16 \quad (6)$$

In order to use PIT, we shall transmit data in the following order (i) DC coefficients in any predefined order; (ii) classmap, c, and bit-allocation maps; (iii) AC coefficients. The transmission of the AC coefficents [2] is made by spanning the blocks and sending first the elements $x_i(m,n)$ which would yield higher contribution to the reconstructed image. For minimizing the reconstruction error, we shall send the coefficients with higher variances. Alternatively, we can incorporate some information about the spatial response of the visual system, by using weighted standard deviations. If one assumes that the estimated standard deviation is a good measure of the real standard deviation of a particular coefficient (at least, is the best information we have at hand), the priority can be decided upon the weighting of the *deviation maps* by a matrix $H(m,n)$ containing spatial information about the HVS. Let

$$\eta_k(m,n) = \hat{\sigma}_k(m,n)\, H(m,n); \quad 1 \leq k \leq N_C; \quad (m,n) \in \Psi \quad (7)$$

The order for transmission of the coefficients is, then, defined by sending first the coefficients $(x_i(m,n); C(i) = k)$ which correspond to greater value of $\eta_k(m,n)$. If there is any ambiguity, send first $\eta_k(m,n)$ with smaller value of $m+n$ and larger value of k (this means lower frequency coefficients on higher AC energy classes). Using (??) and (??), we can redefine $\eta_k(m,n)$ as

$$\eta_k(m,n) = B_k(m,n) + \log_2(H(m,n)) \quad (8)$$

Although having a different value, this representation still maintains the transmission order, addressing directly the bit-allocation maps. As long as both encoder and decoder have the same maps, and use the same weighting matrix, there will be no overhead for indicating the transmission order.

The coder has some limitations and can very well be improved in many ways [18]. However, it is quite insensitive to the transform used, being this the main reason for the choice of the Chen-Smith coder.

3 The HVS weighting matrix

A complete study of the psychophysical properties of the visual system is well beyond the scope of the present paper. Our intention is restricted to the determination of a spatial response weighting matrix for use with the LOT coefficients.

In [2], the application of a linear function relating the HVS to spatial variations is discussed.

Let $H(f_c)$ be a linear transfer function representing the unidimensional spatial model of HVS (f_c given in cycles per degree of the visual angle subtended), Further usual assumptions follow: (i) the screen has an 1:1 aspect ratio and is uniformly bright for an uniform image; (ii) the viewer is situated at a distance v from the screen, right in front of its geometric center; (iii) the screen has width w and each row (column) has N pels; (iv) the viewer can observe aproximately the same density of pels-per-degree (spatial) in any region of the screen.

Let α be the ratio of viewer distance (v) by screen width (w), i.e., $\alpha \equiv v/w$. This factor is the relative distance of the observer. The maximum visible frequency in cycles per degree is obtained when the discrete signal displayed has its maximum frequency component, which is half of the sampling frequency, i.e., in N samples it is possible to observe $N/2$ cycles. The maximum visible frequency can be found as:

$$f_{max} = \frac{N/2}{2\theta} \text{ cycles/degree} \quad (9)$$

where θ in degrees is the viewing angle, from the center to the extreme of the screen, and $\tan(\theta) = w/2v = 1/2\alpha$. We, therefore, can represent a discrete sensitivity function as $H_D(e^{jw}) = H_D(e^{j2\pi f}) = H(f/f_{max})$ for $|f| < f_{max}$. This representation will be accurate if $H(f) = 0$ for $|f| > f_{max}$.

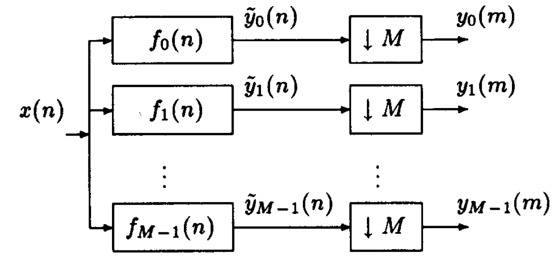

Figure 1. Analysis section of a critically decimated M-band filter bank.

An orthogonal block transform is a special case of lapped transform in which there are as many basis functions as elements in each basis function [17]. Furthermore, lapped transforms are equivalent to paraunitary filter banks [17]. Therefore, we can always regard any discrete, real, and orthogonal (lapped or block) transform as a filter bank [17,19,20]. The analysis filters' coefficients are the time reversed basis functions elements [17,19]. Suppose the M basis functions have elements $p_k(n)$ ($k = 0,1\ldots,M-1$ and $n = 0,1,\ldots,L-1$). The equivalent analysis filter bank is shown in Fig. 1, where each filter (with coefficient $f_k(n)$) is equal to a basis function of the LOT, i.e., $f_k(n) = p_k(L-1-n)$ for $n = 0,1,\ldots,L$. For the particular case of the LOT of M bands, $L = 2M$. For the DCT, we have $L = M$. In Fig. 1, $x(n)$ is the input signal to the filter bank, $\tilde{y}_k(n)$ corresponds to each subband (filtered signals), and $y_k(n)$ is the subband signal after decimation. Let $F_k(e^{j\omega})$ be the frequency response of $f_k(n)$. If, in Fig. 1, the input $x(n)$ has power spectral density (PSD) given by $S_x(\omega)$, and denoting the PSD of $\tilde{y}_k(n)$ and $y_k(n)$ as $S_{\tilde{y}_k}(\omega)$ and $S_{y_k}(\omega)$, we have:

$$S_{\tilde{y}_k}(\omega) = S_x(\omega)|F_k(e^{j\omega})|^2 \qquad (10)$$

After the decimator, $y_k(n) = \tilde{y}_k(nM)$, and

$$S_{y_k}(\omega) = \sum_{r=0}^{M-1} S_{\tilde{y}_k}\left(\frac{\omega - 2\pi r}{M}\right) \qquad (11)$$

Using variable substitution and recognizing that, for a and b real numbers,

$$\int_a^b S_{\tilde{y}_k}(\omega)d\omega = \int_{2\pi-b}^{2\pi-a} S_{\tilde{y}_k}(\omega)d\omega$$

it is easy to see that the variance of y_k is given by

$$\zeta_k^2 = \frac{1}{\pi}\int_0^{\pi} S_{y_k}(\omega)d\omega = \frac{1}{\pi}\int_0^{\pi} S_{\tilde{y}_k}(\omega)d\omega \qquad (12)$$

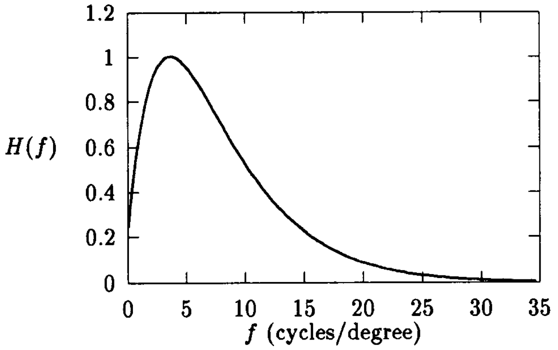

Figure 2. HVS model function used in this paper.

Alternatively, this result could be shown using the fact that if $u(n)$ is a stationary process, then var[$u(n)$]=var[$u(Mn)$]. Therefore, var[$\tilde{y}_k(n)$]=var[$y_k(n)$] and the above equation is also true.

If a white noise with unit variance is input to the linear system $H_D(e^{j\omega})$, and its output is transformed using the LOT, then (??) is given by:

$$\zeta_k^2 = \frac{1}{\pi}\int_0^{\pi} |H_D(e^{j\omega})|^2 \; |F_k(e^{j\omega})|^2 \, d\omega \qquad (13)$$

The continuous HVS model function as used in [2] is shown in Fig. 2, and is given by:

$$H(f) = 2.46(0.1 + 0.25f)\, e^{-0.25f} \qquad (14)$$

The corresponding weights ζ_k are found, for $M = 8$, using (??). The two dimensional case, is just an extension of these results, since the transform is separable. We are interested in weights $\zeta_{ij}, (i,j) \in \Psi$, which can be derived from

$$\zeta_{ij}^2 = \frac{1}{\pi^2}\int_0^{\pi}\int_0^{\pi} |H_D(e^{j\omega_1}, e^{j\omega_2})|^2 |F_{ij}(e^{j\omega_1}, e^{j\omega_2})|^2 d\omega_1 d\omega_2 \qquad (15)$$

$$H_D(e^{j\omega_1}, e^{j\omega_2}) = H_D(e^{j2\pi f_1}, e^{j2\pi f_2}) = H(f_p/f_{max}) \qquad (16)$$

$$f_p = \sqrt{f_1^2 + f_2^2}\ ;\ |f_1| < f_{max}, |f_2| < f_{max}$$

$$F_{ij}(e^{j\omega_1}, e^{j\omega_2}) = F_i(e^{j\omega_1})\, F_j(e^{j\omega_2}) \qquad (17)$$

In our application, we are weighting standard deviation values and we shall use ζ_{ij} instead of the squared value. Figs. 3 and 4 show weighting matrices containing normalized ζ_{ij} for some values of α, f_{max} and N. Values of α as 6 or 7 are more representative for broadcast TV viewing. Values of 4 or 5 fit very well modern PIT needs, and approximate the situation where a 256 × 256 pels image is displayed on the 640 × 480 resolution mode on a regular home PC monitor, and the observer is in front of it, working on the computer.

0.685	0.870	0.988	1.000	0.955	0.870	0.771	0.679
0.870	0.937	0.993	0.982	0.929	0.846	0.747	0.660
0.988	0.993	0.996	0.961	0.899	0.815	0.719	0.636
1.000	0.982	0.961	0.911	0.846	0.766	0.675	0.598
0.955	0.929	0.899	0.846	0.782	0.707	0.624	0.554
0.870	0.846	0.815	0.766	0.707	0.641	0.567	0.505
0.771	0.747	0.719	0.675	0.624	0.567	0.503	0.449
0.680	0.660	0.636	0.598	0.554	0.505	0.449	0.402

(a) $\alpha = 4$; $f_{max} = 9$ cycles/degree

0.863	1.000	0.975	0.823	0.649	0.493	0.351	0.277
1.000	0.993	0.918	0.767	0.605	0.462	0.330	0.262
0.975	0.918	0.821	0.682	0.540	0.416	0.300	0.238
0.823	0.768	0.682	0.570	0.455	0.354	0.258	0.206
0.649	0.605	0.540	0.455	0.368	0.290	0.214	0.172
0.493	0.462	0.416	0.354	0.290	0.231	0.173	0.139
0.351	0.330	0.300	0.258	0.214	0.173	0.133	0.107
0.277	0.262	0.238	0.206	0.172	0.139	0.107	0.087

(b) $\alpha = 7$; $f_{max} = 15.7$ cycles/degree

Figure 3. 2D HVS weighting matrices for the LOT, $N = 256$.

0.894	1.000	0.921	0.730	0.537	0.386	0.254	0.200
1.000	0.964	0.847	0.668	0.494	0.358	0.236	0.187
0.921	0.847	0.727	0.575	0.429	0.314	0.210	0.166
0.729	0.668	0.575	0.459	0.348	0.258	0.175	0.139
0.537	0.494	0.429	0.348	0.268	0.202	0.140	0.111
0.386	0.358	0.314	0.258	0.202	0.154	0.109	0.086
0.254	0.236	0.210	0.175	0.140	0.109	0.079	0.063
0.200	0.187	0.166	0.139	0.111	0.086	0.063	0.050

(a) $\alpha = 4$; $f_{max} = 18$ cycles/degree

1.000	0.896	0.585	0.323	0.159	0.114	0.035	0.069
0.896	0.725	0.471	0.267	0.134	0.094	0.030	0.056
0.585	0.471	0.316	0.187	0.098	0.066	0.024	0.038
0.323	0.267	0.187	0.116	0.064	0.042	0.016	0.023
0.159	0.134	0.098	0.064	0.037	0.024	0.011	0.012
0.114	0.094	0.066	0.042	0.024	0.016	0.007	0.008
0.035	0.030	0.023	0.016	0.011	0.007	0.003	0.003
0.069	0.056	0.038	0.023	0.012	0.008	0.003	0.005

(b) $\alpha = 7$; $f_{max} = 31.4$ cycles/degree

Figure 4. 2D HVS weighting matrices for the LOT, $N = 512$.

4 Implementation and results

A 256 × 256 pels monochrome image is divided into (8x8) nonoverlapping blocks ($M = 8$) and the LOT is applied to each block. Based on the ac energies, the (8x8) blocks are grouped into eight different equally populated classes ($N_C = 8$). Thus, there are 32 × 32 blocks in the image ($N_B = 1024$). The DC coefficients are quantized with an uniform 7 bits quantizer, and B_{max} is set to be 7. Therefore, the overhead in (??) is 4552 bits and the amount of bits needed to code the DC coefficients is 7168. This yields a total of 11720 bits sent prior to the transmission of the AC coefficients (aproximately 0.18 bpp).

The AC coefficients are well modelled by a Laplacian probability density function (PDF), but the blocks are classified according to their AC activities. If u is the amplitude of an AC coefficient, the actual important function is no longer its density function $p_U(u)$, but one conditional to the estimated standard deviation $p_U(u|\hat{\sigma})$, which depends on the importance of this coefficient in the classification process. If there is just one class ($N_C = 1$), the Laplacian model fits well. As an example for a particular coefficient, suppose its standard deviation is estimated as very large. This indicates that the coefficients on that coordinate $(m, n) \in \Psi$ belonging to the same class, are expected to have high amplitude, and not amplitudes close to zero as in the Laplacian model. Generally these large coefficients have low frequency and have large number of bits allocated for it. Coefficients with one or two bits allocated, generally do not have great influence on the AC energy and are very close to the Laplacian model. In our bit-allocation we assumed quantizers optimized using the same PDF. Therefore, we have chosen the Gaussian PDF model for our Lloyd-Max quantizers, due to its greater robustness against PDF mismatches. Tests carried out (for 8 and 16 classes) using two sets of quantizers (for Laplacian and Gaussian PDFs), showed better berformance for the Gaussian set of quantizers.

The assumption of integer number of bits allocated to each coefficient implies the use of mid-rise quantisers. One and two bit quantizers optimized for a Gaussian input PDF, have high inner reconstruction levels. This leads to annoying granular noise emerging from the amplification of insignificant coefficients, to which a high variance was assigned in the averaging process. For this reason, we decided to apply mid-tread quantizers with three-levels and variable length coding, instead of quantizing with 2 or 4 levels. The standard deviations for quantization and reconstruction of these coefficients would remain the same, but the distortion rule and the average bit-rate would be affected. However, the distortion/rate changes in the 2-bit quantizer, are roughly compensated by the changes in the 1-bit quantizer. In simulations, both schemes yielded roughly the same bit-rates, with the 3-level scheme leading to images with higher signal-to-noise ratios.

The HVS weighted PIT described above is extended to the 2D-DCT. The weighting matrix was found using the method described in [2]. In Table I, a comparison of both methods is carried out, evaluating the SNR of reconstructed images at several stages for images "Lena" and "Girl". In both cases, the HVS model was calculated using $f_{max} = 13.4$ ($\alpha = 6$).

Table I. SNR(dB) resulting from intermediary reconstructed images at several bit-rates for images "Lena" and "Girl".

Rate(bpp)	SNR			
	LOT	DCT	LOT	DCT
	LENA		GIRL	
0.2	16.10	15.18	17.22	16.21
0.3	19.43	18.41	20.27	19.41
0.4	21.00	20.55	22.68	21.93
0.5	22.74	22.39	24.30	23.79
0.6	23.68	23.29	25.21	24.98
0.8	25.35	25.15	26.98	26.76
1.0	26.81	26.67	28.50	28.27

5 Conclusions

A PIT scheme which incorporates adaptive classification in the transform domain and bit-allocation based on the rate-distortion theory is presented. A general technique for developing HVS weighting of the LOT coefficients is developed. The order in which the transform coefficients are transmitted is based on the estimated variances of these coefficients weighted by the human visual system sensitivity, measured in the 2D-LOT domain. Visual comparison of the reconstructed images based on the LOT and DCT shows that the former yields subjectively superior images compared to the DCT for all stages.

References

[1] K.H. Tzou, "Progressive image transmission: A review and comparison of techniques," Optical Engineering, vol. 26, pp. 581-589, July 1987.

[2] B. Chitprasert and K.R. Rao, "Human visual weighted progressive image transmission," IEEE Trans. Commun,. vol. 38, pp. 1040-1044, July 1990.

[3] K.R. Rao and P. Yip, Discrete Cosine Transform; Algorithms, Advantages and Applications, Academic Press, San Diego, 1990.

[4] M. Rabbani and P.W. Jones, "Digital image compression techniques," SPIE Optical Engineering Press, Bellingham, WA, 1991.

[5] K.H. Tzou and S.E. Elnahas, "An optimal progressive transmission and reconstruction scheme for transformed images," ICC 86, pp. 413-418, Toronto, Canada, June 1986.

[6] Final text for ISO/IEC DIs 10918-1, "Info. technology - digital compression and coding of continuous tone still images." Part 1: requirements and guidelines, Jan. 14, 1992. Part 2, compliance testing CD 10918-2, 12/16/91.

[7] S.E. Elnahas et al., "Progressive transmission of digital diagnostic images, "Appl. Digital Image Processing VIII, Proc. SPIE, vol. 575, pp. 48-55, Aug. 1985.

[8] W.H. Chen and W.K. Pratt, "Scene adaptive coder," IEEE Trans. Commun., vol. COM-32, pp. 225-232, March 1984.

[9] W.H. Chen and C.H. Smith, "Adaptive coding of monochrome and color images," IEEE Trans. Commun., vol. COM-25, pp. 1285- 1292, Nov. 1977.

[10] K.N. Ngan, "Adaptive transform coding of video signals," IEE Proc., vol.129, Pt.F, pp. 28-40, Feb. 1982.

[11] Y.S. Ho and A. Gersho, "Classified transform coding of images using vector quantization," Proc, ICASSP '89, pp. 1890- 1893, Apr. 1989.

[12] J.Y. Nam and K.R. Rao, "Image coding using a classified DCT/VQ based on two channel conjugate vector quantization". IEEE Trans. Circuits and Systems for Video Technology. Vol. 1, pp. 327-336, Dec. 1991.

[13] K.H. Tzou, T.R. Hsing, and J.G. Dunham, "Applications of physiological human visual system model to image compression," Proc. SPIE, vol. 504, pp. 419-424, 1984.

[14] S. Ericsson, "Frequency weighted interframe hybrid coding," Rep. TRITA-TTT-8401, Telecommun. Theory, Royal Inst. Technol., Stockholm, Sweden, Jan. 1984.

[15] K.N. Ngan, K.S. Leong, and H. Singh, "Cosine transform coding incorporating human visual system model," presented at SPIE Fiber '86, Cambridge, MA, pp. 165-171, Sept. 1986.

[16] H.S. Malvar and D.H. Staelin, "The LOT: Transform coding without blocking effects," IEEE Trans. ASSP, vol. 37, pp. 553-559, Apr. 1989.

[17] H. S. Malvar, "Signal processing with lapped transforms," Norwood, MA, Artech House, 1992.

[18] E.M. Rubino, H.S. Malvar, and R. L. Queiroz, "Improved Chen-Smith coder," *Proc. of Intl. Symp. Circuits and Systems,* Chicago, IL, *1993.*

[19] M. Vetterli and D. Le Gall, "Perfect reconstruction filter banks : some properties and factorizations", IEEE Trans. Acoust., Speech, Signal Processing, ASSP-37, pp. 1057-1071, July 1989.

[20] R. E. Crochiere, L. R. Rabiner, Multirate Digital Signal Processing. Englewood Cliffs, NJ: Prentice-Hall, 1983.

P.S. - Space limitations preclude the publication of the sequence of images, obtained during the progressive transmission stages, comparing both transforms at each step. However, a complete version of this paper can be found at J. Electronic Imaging, vol. 1, pp 328–338, July 1992.