OPTIMAL ERROR PROTECTION FOR REAL TIME IMAGE AND VIDEO TRANSMISSION

Masoud Farshchian, Sungdae Cho, William A. Pearlman

Electrical, Computer and Systems Engineering
Rensselaer Polytechnic Institute

ABSTRACT

In this paper a novel and computationally inexpensive analytic mean square error (MSE) distortion rate (D-R) estimator for SPIHT which generates a nearly exact distortion rate (D-R) function for 2-D and 3-D SPIHT algorithm is presented. The analytical formula is derived from the observations that for any bit-plane coder, the slope of the D-R curve is constant for each level of the bit plane. Furthermore the slope of D-R curve reduces by a factor proportional to the level of the bit plane. An application of the derived results is presented in the area of 2-D SPIHT transmission employing a binary symmetric channel (BSC) and Reed Solomon (RS) forward error correction (FEC) codes. Utilizing our D-R estimate, we employ unequal error protection (UEP) and equal error protection (EEP) in order to minimize the end to end mean square error (MSE) distortion of the transform domain. UEP yields a significant performance gain relative to EEP only when the average number of parity bits for a group of packets is constrained. When both the source rate and channel code rate varied under a bit budget constraint, optimal UEP yields only a slight improvement over the optimal EEP. A major contribution of this paper is the simple and extremely accurate analytical D-R model which potentially improves upon pre-existing methodologies and applications that rely on an accurate and computationaly inexpensive D-R estimate. Another important contribution is that the optimum EEP, which requires almost no header information and can easily be computed using our method, is only slightly worse than the optimum UEP.

1. INTRODUCTION

Progressive image and video transmission is problematic in the presence of noisy channels. Progressive source coders like Image SPIHT and Video SPIHT [1]use a variable length format where the correct decoding of future bits depend upon the correct transmission of past bits. Decoding after the first single bit error can increase the expected distortion at the receiver and the best strategy is to stop decoding before the first bit error. We assume that the decoder has the capability to detect all block errors. Let us denote by $D(R)$ the mean square error distortion remaining after R bits have been correctly decoded. Due to the progressive nature of the source coder bitstream, we stop decoding prior to the first decoding failure. Since all blocks after an erroneous block are corrupted due to their dependency on the incorrect block, the expected distortion $E(D)$ depends on the location of the first block error. If we successfully decode all blocks up to and not including block m, the distortion is denoted by $D_b(m)$. This probability of first block failure is equal to $P_{bl}(m,t_i)\prod_{j=0}^{m-1}(1-P_{bl}(i,t_i))$, where $P_{bl}(i,t_i)$ is the probability of losing block i which has a total of $2t_i$ parity bits. So the expected end to end distortion $E(D)$ under a bit budget constraint of N equal sized blocks and a total source rate R_s bits is given by:

$$\begin{aligned} E(D) \quad &= \quad \textstyle\sum_{i=0}^{N-1} D_b(i)P_{bl}(i,t_i)\prod_{j=0}^{i-1}(1-P_{bl}(i,t_i)) \\ &\quad + D(R_s)\prod_{i=0}^{N-1}(1-P_{bl}(i,t_i)) \end{aligned} \tag{1}$$

The optimization of Equation (1) forms the objective function of the joint source channel coding scheme analyzed in [2, 3] amongst many other papers. It is clear from Equation (1) that optimal parity allocation across different blocks depends greatly on the D-R characteristics of the source coder. One method to estimate the D-R curve is to decode at certain number of points at the receiver and interpolate the D-R function. The drawback to such a method is that it might not be realizable for a real-time application. Furthermore such a method is not always accurate because the points that are decoded may not accurately capture the slope variation to estimate an accurate D-R function. In [2] Appadwedula *et al.* used an exponential model of the form:

$$D(b) = \sum_{k=1}^{4} c_k e^{-l_k b} \tag{2}$$

Where b is the b^{th} bit and c_k and l_k are parameters for a specific class of images. The major benefit of using such models is that they allow Equation (1) to be solved using optimization techniques. The drawback of such models is that they are not particularized to a specific image and video sequence. Since they are not exact models, a loss in performance relative to an exact model is inevitable. This loss in performance occurs due to the use of the wrong D-R function in Equation (1) that results in allocating a non-optimum parity across the transmission blocks. In this paper by analyzing the SPIHT coder and bit plane coders in general, we offer a more accurate model for image and video coders. Furthermore by using our D-R estimator, other parametric models can be fitted more accurately for a particular image or video sequence.

The discrete two or three dimensional wavelet transforms (DWT) considered here are biorthogonal and separable. As a result of having the biorthogonality condition, the squared norm in the transform domain is not exactly preserved, but is very close for the filters we will be using. In this paper the mean square error (MSE) will refer exclusively to the sum of the square error difference between the quantized wavelet transform and its lossy approximation divided by a constant K. For images K is the number of pixels. For image sequences K is the number of pixels per group of frames(GOF).

2. RATE DISTORTION PROFILE OF SPIHT

Wheeler analyzed the reduction in distortion for each received bit. We will expound on that work and then design an accurate distortion-rate (D-R) curve estimator for the SPIHT coder. Recall from [1] that at each iteration of the SPIHT coder, all coefficients that are greater than the threshold τ at that pass and are less than 2τ are considered significant by the SPIHT coder. All other transformed coefficients which are not significant are deemed insignificant. Once a significant coefficient is found, its position and approximate magnitude, which is about one and half times the threshold level, are inferred from the significance map by one bit of information and its sign is coded using one additional bit of information. So a newly found transformed coefficient $Y(i,j)$ at location (i,j) found to be significant at a threshold $\tau = 2^n$ is assigned a magnitude value of $\tilde{Y}(i,j) = 1.5\tau$. After a coefficient has been found to be significant at τ, then it is put in a special list for further refinement at each subsequent SPIHT pass. Each refinement pass effectively halves the region of uncertainty relative to the previous refinement pass.

Initially before any decoding, each coefficient of the image is assumed to be zero. When a coefficient $Y(i,j)$ is found to be significant at τ, then a sign bit and a significance bit is sent. The mean of the lowest frequency subband is also zero, because the image mean is subrtracted before coding. Assuming that the coefficient is positive and uniformly distributed between $[\tau, 2\tau)$, then expected square error in assuming a zero value for the coefficient is

$$E\{(Y(i,j)-0)^2\} = \int_{\tau}^{2\tau} \frac{1}{\tau} y^2 dy = \frac{7}{3}\tau^2 \quad (3)$$

If we reproduce the coefficient $\tilde{Y}(i,j) = 1.5\tau$ then the expected squared error becomes

$$E\{(Y(i,j)-\tilde{Y}(i,j))^2\} = \int_{\tau}^{2\tau} \frac{1}{\tau}(y-1.5\tau)^2 dy = \frac{1}{12}\tau^2 \quad (4)$$

Finally we consider the case where the coefficient is refined. Given that the coefficient was found to be significant at τ and a 1 is received, the expected distortion is given by

$$E\{(Y(i,j)-\tilde{Y}(i,j))^2\} = \int_{1.5\tau}^{2\tau} \frac{2}{\tau}(y-1.75\tau)^2 dy = \frac{1}{48}\tau^2 \quad (5)$$

If k refinement bits were received for the (i,j) coefficient and the coefficient was found at a significance level τ, the MSE between the two coefficients is

$$E\{(Y(i,j)-\tilde{Y}(i,j))^2\} = \frac{1}{12}(\frac{1}{4})^k\tau^2 \quad (6)$$

We will keep track of the number of the newly found significant bits for each pass of the bit plane coder as well as the total number of bits per each pass. We assume that the bit plane decoding starts at the level $\tau = 2^n$. Let us denote by $N_{SBS}(i)$ as the number of sign bits in pass i and by $Nd_{SBS}(i)$ the number of sign bits decoded in pass i. Note that $N_{SBS}(i)$ is equal to the number of coefficients found significant at pass i. These quantities are easily generated by the SPIHT coder at virtually no cost in the computational complexity of the algorithm.

Since SPIHT finds all the coefficients that are significant relative to a threshold at each pass, then $N_{SBS}(i)$ is equivalent to the number of transformed coefficients whose magnitude is greater than or equal to 2^i and less than 2^{i+1}. Assuming that we stopped decoding during the sorting pass of the significance level $\tau = 2^k$, an approximation for $D(R)$, denoted by $\hat{D}(R)$ is given implicitly by

$$\begin{aligned}\hat{D}\quad(R) = \quad & \tfrac{1}{K}\Big[\textstyle\sum_{i=k+1}^{n} N_{SBS}(i)\tfrac{2^{2i}}{12}(\tfrac{1}{4})^{i-(k+1)} + \\ & \tfrac{2^{2(k)}}{12} Nd_{SBS}(n-k) + \\ & [N_{SBS}(n-k) - Nd_{SBS}(n-k)]\,\tfrac{7}{3}2^{2(k)} + \\ & \textstyle\sum_{m=0}^{k} N_{SBS}(m)\tfrac{7}{3}2^{2m}\Big]\end{aligned} \quad (7)$$

Level	Lena		Goldhill		Barbara	
	$D(R)$	$\hat{D}(R)$	$D(R)$	$\hat{D}(R)$	$D(R)$	$\hat{D}(R)$
12	6241	5905	7671	7638	6615	6482
11	1181	1134	978	1031	1595	1588
10	642	630	511	521	902	896
9	383	380	379	380	641	639
8	237	236	268	267	463	462
7	130	129	177	176	318	315
6	67	67	107	107	155	153
5	33	33	58	57	61	61
4	16	16	26	26	22	22
3	7	7	9	9	8	7

Table 1. Actual and estimated D-R points of coded SPIHT images at the end of each threshold

Level	Flower Garden		Mobile Calendar	
	$D(R)$	$\hat{D}(R)$	$D(R)$	$\hat{D}(R)$
12			13339	14618
11	3126	3066	3119	3164
10	1932	1860	1844	1772
9	1287	1243	1387	1342
8	843	804	921	897
7	487	451	522	511
6	242	220	257	251
5	100	90	102	99
4	34	31	34	31
3	10	9	10	8

Table 2. Actual and estimated D-R points for coded SPIHT video at the end of each threshold

The first component of Equation (7) takes into account the reduction of distortion after $k+1$ passes by decoding all the sign bits found from significance level n all the way down to significance level $k+1$ and is given by Equation (6) per sign bit. The second term is a result of the number of sign bits decoded at the significance level k. The remaining two terms are the mean square error given by Equation (3) for not yet found non-zero coefficients. In order to have $\hat{D}(R)$, we need $N_{SBS}(m)$ for every treshold level.

We have calculated some of the estimated $D(R)$ values for threshold levels 12 down to 3 ($\tau = 2^{12}$ to 2^3) in Table 1 for three popular images. These results verify that $\hat{D}(R)$ is a good approximations for the actual $D(R)$ in all cases. We have also used the same approximation for 3-D SPIHT [2] and Table 2 lists the same results obtained for the luminance Y components of two video sequences. Based on these results the model we propose for a fast computation of the SPIHT D-R characteristic is a linear interpolation between the $\hat{D}(R)$'s at each pass of the SPIHT algorithm.

3. APPLICATION TOWARDS JOINT SOURCE CHANNEL CODING

Using $\hat{D}(R)$ and Equation (1), we have solved for the optimal unequal error protection using a gradient based algorithm. The error correction code we have used are Reed Solomon (RS) blocks of size 255 bytes denoted by $(255, k)$, where the number of information symbols k varies per block. We have assumed that we are operating over a binary symmetric channel with a cross over probability of 0.01. We have solved the algorithm for different transmission rate constraints including that of .1089 bpp which equates exactly to a total of 14 RS blocks. In Figure 1 we have plotted in the transform domain $PSNRT = 10 \log 10(255^2/E(D(R)))$ based on the expected MSE given by Equations(1) for the UEP and EEP case. At the optimal trade-off point between the source rate and the channel rate, the UEP performance at the transmission rate of .1089 bpp (bits per pixel) is 0.17 dB better than the best value obtained using EEP. Only when the average error correction capability per block is constrained, then UEP yields a significant improvement over EEP. Figure 1 illustrates that when the source rate and channel rate are fixed and only the parity allocation per block is allowed to vary, UEP allows for graceful degradation when the average number of parity bits per block is reduced. In Table 2 and Table 3 we have listed the $PSNRT$ for the optimal UEP and the optimal EEP for the Lena and Goldhill images at various rates. In these two tables the entries UEP and EEP signify the optimal UEP and EEP when the exact D-R function is used and the entries UEP-$\hat{D}(R)$ and EEP-$\hat{D}(R)$ signify the optimal parity allocation achieved via Equation (7) and subsequently then applied to Equation (1) with the exact D-R curve. From the table we observe that the maximum difference between the optimal UEP and the optimal EEP is 0.24 dB and the minimum difference is .03 dB. Also using $\hat{D}(R)$ to solve for the optimal parity allocation is as good as having the actual D-R function.

Chande and Farvardin [3] used rate compatible convolutional codes and a spatial domain PSNR criterion. They noticed that for any transmission rate, one of their EEP schemes which is not necessarily the optimal EEP has a small performance loss relative to the optimal UEP. Our results not only confirm this fact, but additionally using our simple D-R estimator, we can always obtain the optimal EEP at any rate. All we have to do is to evaluate Equation (1) via Equation (7) at several equal code rates and get the maximum value. The computation cost of this procedure is negligible relative to any optimization technique that employs procedures such as gradient based method or dynamic programming. This leads us to believe that the optimal EEP will be close to the optimal UEP for a wide range of channel coders and for both PSNR and MSE distortion measures.

Rate	Normal	UEP	EEP	U$\hat{D}(R)$	E$\hat{D}(R)$
.1089	30.52	29	28.83	29	28.82
.249	33.91	32.44	32.2	32.44	32.2
.498	36.94	35.24	35.01	35.23	35.01
.755	38.49	37.06	36.78	37.05	36.78

Table 3. Expected PSNRT for Lena over a memoryless BSC with BER 0.01

Rate	Normal	UEP	EEP	U$\hat{D}(R)$	E$\hat{D}(R)$
.1089	28.15	27.03	26.97	27.03	26.97
.249	30.34	29.25	29.08	29.25	29.07
.498	32.74	31.51	31.49	31.50	31.48
.755	34.70	32.84	32.70	32.84	32.69

Table 4. Expected PSNRT for Goldhill over a memoryless BSC with BER 0.01

Despite the slight performance loss, the optimal EEP has some advantages over the optimal UEP. First, it does not require any extra header information pertaining to the D-R estimation for the receiver. For the optimal UEP, there exist two options for the receiver to have the necessary header information. Option one is to code and transmit the parity allocation per block. This can potentially be a large amount of information at high transmission rates. Option two entails sending the side information that is needed to calculate $\hat{D}(R)$ at the receiver. This is the number of total bits and sign bits at the end of each SPIHT pass. Considering that without the header information the decoding of the image can not correctly occur, the header information must be extremely well protected. On the other hand, for the optimal EEP the receiver just needs to know a single number that achieves the optimal EEP for a large group of blocks. The second advantage of the optimal EEP is that it does not require any optimization techniques like those that employ gradient based methods or dynamic programming to obtain the optimal or near optimal UEP. For real time applications the time delay to run such programs for every image or video sequence may be intolerable. For systems with power constraints like mobile phones, the use of an optimization program for every image or image sequences could potentially be an important factor for the system designer. Finally the optimal EEP is simpler to implement since the code rate is the same per each block over a large group of blocks whereas for an UEP scheme the parity per each block may vary.

4. CONCLUSION AND DISCUSSION

We have introduced a new method to estimate the D-R characteristics of image and video SPIHT accurately at a very small computational cost. Our method is shown to be superior to other methods because it is particularized for any image and video dataset. Although the source coding algorithm used is SPIHT, the D-R estimation method is easily generalizable to any modern progressive coder that employs progressive bit-plane coding. The match between the estimated D-R function and the actual D-R function for both 2-D and 3-D SPIHT verifies that our estimate is an excellent approximation. For the optimal UEP we used a gradient based method to solve for the optimal parity allocation. For the EEP case we did not need any optimization technique and evaluating Equation (1) at several points was sufficient to obtain the optimal EEP. The new accurate D-R estimation proposed in this paper can bridge the gap between theory and actual real time implementation of joint source channel coding for image and video transmission systems. Finally another major result is that the optimum UEP is only slightly superior to the optimum EEP. It was also mentioned that the optimum EEP offers some substantial practical advantages for real-time applications over the optimum UEP. The major advantages are simpler implementation, a significantly smaller header information and independence from any type of optimization procedure as a consequence of our fast D-R estimation.

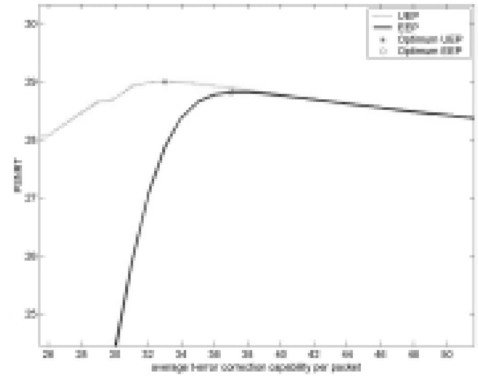

Fig. 1. EEP vs. UEP for Lenna at transmission rate of .1089 bpp

5. REFERENCES

[1] A. Said and W. A. Pearlman. ”A new, fast, and efficient image coded based on SPIHT.” *IEEE CSVT,* 6(3):243-250, June 1996.

[2] S. Appadwedula, D. L. Jones, K. Ramchandran, and I. Konzintsev, ”Joint source-channel matching for wireless communications link,” in *ICIP*, 1998.

[3] V. Chande and N. Farvardin, ”Progressive transmission of images over memoryless noisy channels,” *IEEE J. Select. Areas Commun.,* vol. 18, pp 850-860, July 2000.