Optimal Error Protection for Real-Time Image and Video Transmission

Masoud Farshchian, Sungdae Cho, and William A. Pearlman

***Abstract*—In this letter, a novel and computationally inexpensive analytic mean square error (mse) distortion rate (D-R) estimator for SPIHT which generates a nearly exact D-R function for the two- and three-dimensional SPIHT algorithm is presented. Utilizing our D-R estimate, we employ unequal error protection and equal error protection in order to minimize the end to end mse distortion of the transform domain. A major contribution of this letter 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 computationally inexpensive D-R estimate.**

***Index Terms*—Image transmission, source-channel coding, SPIHT, video transmission, wireless transmission.**

I. Introduction

PROGRESSIVE image and video transmission is problematic in the presence of noisy channels. Progressive source coders like Image SPIHT [1] and Video SPIHT [2] use a variable-length format where the correct decoding of future bits depends 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 (mse) distortion per sample 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 per sample is denoted by $D_b(m)$. This probability of first block failure is equal to $P_{bl}(1)$ for $\mathrm{m} = 1$ and $P_{bl}(m) \prod_{j=1}^{m-1} (1 - P_{bl}(j))$ for $m = 2, 3 \ldots N$, where $P_{bl}(j)$ is the probability of losing block j. 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

$$E(D) = D(0)P_{bl}(1) + \sum_{m=2}^{N} D_b(m)P_{bl}(m) \cdot \prod_{j=1}^{m-1}(1 - P_{bl}(j)) + D(R_s)\prod_{j=1}^{N}(1 - P_{bl}(j)). \tag{1}$$

The optimization of (1) forms the objective function of the joint source channel coding scheme analyzed in [3]–[5], [7] amongst many other papers. The distortion in decoding up to block m, $D_b(m)$, depends on the number of bits received for $m-1$ blocks. Therefore the optimal parity allocation across different blocks depends greatly on the distortion rate (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 [3], Appadwedula *et al.* used a piecewise exponential model with different decay parameters. The major benefit of using such models is that they allow (1) to be solved using optimization techniques. Their parametric model was proposed for a class of images. The drawback of such models is that they are not particularized to a specific image and video sequence. Charfi *et al.* [6] proposed a more accurate parametric Weibull model where the parameters are estimated from the particular image. In order to fit their D-R estimator to the actual D-R curve, they required decoding four and sometimes eight exact points on the actual D-R curve. In this letter, by analyzing the SPIHT coder and bit plane coders in general, we offer an accurate model for individual 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 advantage of our model is that no decoding is required in order to estimate the D-R function. This makes it well suited for real-time applications relative to parametric models that require actual D-R points.

II. D-R Profile of Spiht

Wheeler [8] analyzed the reduction in distortion for each received bit. We will expound on that work and then design an

Manuscript received December 5, 2003; revised March 4, 2004. This work was supported by the Center of Image Transmission and Exploitation (CITE) and by the Information Directorate of the Air Force Research Laboratory, Rome, NY. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Xiang-Gen Xia.

The authors are with the Center for Image Processing Research, Department of Electrical, Computer and Systems Engineering, Rensselaer Polytechnic Institute, Troy, NY 12180-3590 USA (e-mail: pearlw@ecse.rpi.edu).

Digital Object Identifier 10.1109/LSP.2004.835470

accurate D-R curve estimator for the SPIHT coder. Recall from [1] that at each iteration of the SPIHT coder, all coefficients whose magnitudes 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 1.5τ. After a coefficient has been found to be significant at threshold 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 are sent. The mean of the lowest frequency subband is also zero, because the image mean is subtracted before coding. Assuming that the coefficient is positive and uniformly distributed between $[\tau, 2\tau)$[1], then the expected square error in assuming a zero value for the coefficient is

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

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

$$E\{(Y(i,j) - \tilde{Y}(i,j))^2\} = \int_{\tau}^{2\tau} \frac{1}{\tau}(y - 1.5\tau)^2 dy = \frac{1}{12}\tau^2. \tag{3}$$

Since the quantization interval reduces by a factor of 2 and the mse by a factor of 4 at each lower bit plane, then 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 the actual and estimated coefficient value is

$$E\{(Y(i,j) - \tilde{Y}(i,j))^2\} = \frac{1}{12}\left(\frac{1}{4}\right)^k \tau^2. \tag{4}$$

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 $N_{\mathrm{SBS}}(i)$ as the number of sign bits in pass i and by $Nd_{\mathrm{SBS}}(i)$ the number of sign bits decoded in pass i. Note that $N_{\mathrm{SBS}}(i)$ 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 $N_{\mathrm{SBS}}(i)$ 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 by

$$\begin{aligned}\hat{D}(R) = \frac{1}{K}\Bigg[&\sum_{i=k+1}^{n} N_{\mathrm{SBS}}(i)\frac{2^{2i}}{12}\left(\frac{1}{4}\right)^{i-(k+1)} \\ &+ \frac{2^{2(n-k)}}{12} Nd_{\mathrm{SBS}}(n-k) \\ &+ \left[N_{\mathrm{SBS}}(n-k) - Nd_{\mathrm{SBS}}(n-k)\right]\frac{7}{3}2^{2(n-k)} \\ &+ \sum_{m=0}^{n-k-1} N_{\mathrm{SBS}}(m)\frac{7}{3}2^{2m}\Bigg].\end{aligned} \tag{5}$$

TABLE I
ACTUAL AND ESTIMATED D-R POINTS OF SPIHT IMAGES AND VIDEO AT THE END OF EACH THRESHOLD

Level	Lena		Goldhill		Football		Susie	
	$\hat{D}(R)$	$D(R)$	$\hat{D}(R)$	$D(R)$	$\hat{D}(R)$	$D(R)$	$\hat{D}(R)$	$D(R)$
12	5594	6241	7185	7671	19225	16883	15804	12293
11	1222	1181	1102	978	1547	1163	1750	1639
10	724	642	636	511	900	957	366	363
9	457	383	468	379	631	605	174	173
8	284	237	329	268	429	395	116	107
7	157	130	218	177	285	260	80	72
6	83	67	132	107	165	155	51	46
5	42	33	72	58	80	79	29	27
4	22	16	34	26	34	34	14	14
3	11	7	12	9	11	12	6	6

R, the rate, is a sum of location, sign, and refinement bits. The first component of (5) takes into account the reduction of distortion after $n-k$ 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 (4) 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 mse given by (2) for not yet found nonzero coefficients. In order to have $\hat{D}(R)$ for every obtainable R, we need $N_{\mathrm{SBS}}(m)$ for every threshold 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 I for two popular images, Lena and Goldhill. These results verify that $\hat{D}(R)$ is a good approximation for the actual $D(R)$ in all cases. We have also used the same approximation for 3-D SPIHT [2] and Table I lists the same results obtained for the luminance Y components of two video sequences, Football and Susie. Based on these results of 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. A more exact model could have been obtained had we considered the decrease in distortion due to each refinement bit. But this would make our model too complex to be suitable for real-time application.

III. APPLICATION TOWARDS JOINT SOURCE CHANNEL CODING

Using $\hat{D}(R)$ and (1), we have solved for the optimal unequal error protection using a gradient based algorithm. The error correction code is Reed Solomon (RS) [9] (255, k), blocks of size 255 bytes, where the number of information symbols k varies per block. Assuming independent symbol errors, where

[1] For most images, a probability distribution biased toward the smaller values in the interval would be more accurate, but the uniform distribution proves to be accurate enough and satisfies the minimax criterion.

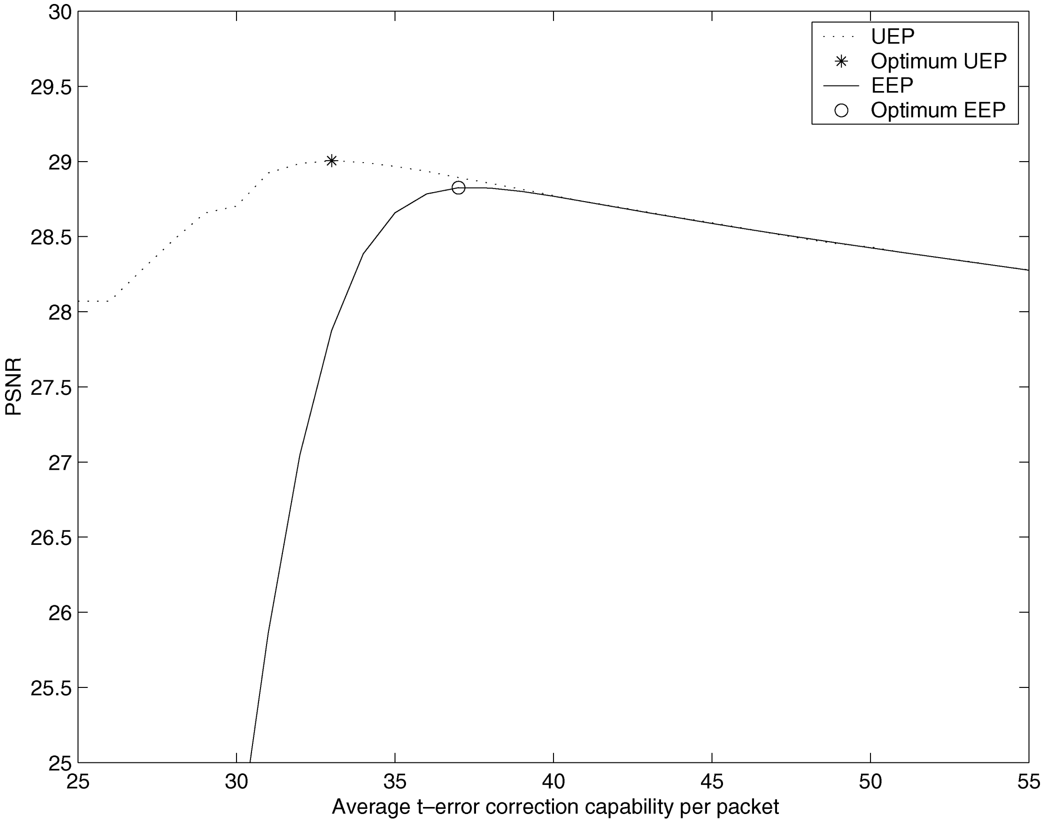

Fig. 1. EEP versus UEP for Lenna at transmission rate of .1089 bpp.

TABLE II
PSNR FOR AVERAGE MSE FOR LENA AND GOLDHILL IMAGES OVER A MEMORYLESS BSC WITH BER 0.01

	Lena				Goldhill			
	0.1089 (14 blocks)	0.249 (32 blocks)	0.498 (64 blocks)	0.755 (97 blocks)	0.1089 (14 blocks)	0.249 (32 blocks)	0.498 (64 blocks)	0.755 (97 blocks)
Noiseless	30.52	33.91	36.94	38.49	28.15	30.34	32.74	34.70
UEP	29.00	32.44	35.24	37.06	27.03	29.25	31.51	32.84
EEP	28.83	32.20	35.01	36.78	26.97	29.08	31.49	32.70
UEP $\hat{D}(R)$	29.00	32.44	35.23	37.05	27.03	29.25	31.50	32.84
EEP $\hat{D}(R)$	28.82	32.20	35.01	36.78	26.97	29.07	31.48	32.69

TABLE III
PSNR FOR AVERAGE MSE FOR SUSIE SEQUENCE OVER A MEMORYLESS BSC WITH BER 0.01 USING 3-D SPIHT

Rate	Noiseless-channel	UEP	EEP	UEP $\hat{D}(R)$	EEP $\hat{D}(R)$
0.0226 (120 blocks)	35.95	34.81	34.71	34.81	34.71
0.0245 (130 blocks)	36.18	35.01	34.89	35.01	34.89

a symbol is a byte, and $2t_j$ parity bytes for block j, the number of symbol errors is binomially distributed and the block's error probability is equal to the probability of more than t_j byte errors.

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 from 0.1089 bpp with 14 RS blocks to 0.755 bpp with 97 RS blocks. In Fig. 1, we have plotted in the transform-domain peak signal-to-noise ratio (PSNR) based on the expected mse given by (1) for the UEP and EEP case. At the optimal tradeoff point between the source rate and the channel rate, the UEP performance at the transmission rate of 0.1089 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. Fig. 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 Tables II and III, we have listed the PSNR based on the expected mse for the optimal UEP and the optimal EEP for the Lena and Goldhill images and Susie image sequence 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 (5) and subsequently then applied to (1) with the estimated D-R curve. Note that there is no more than 0.01 dB difference between PSNRs of the exact and estimated $D(R)$ functions. Furthermore, the maximum difference between the optimal UEP and the optimal EEP is 0.24 dB and the minimum difference is .03 dB. We conclude that $\hat{D}(R)$ is as good as the estimated $D(R)$ in solving for the optimal parity allocation and that optimal EEP attains almost as good performance as the optimal UEP.

Using our simple D-R estimator, we can always obtain the optimal EEP at any transmission rate. Chande and Farvardin [7] used rate-compatible convolutional codes. They noticed that for some transmission rates, one of their EEP schemes, which may not necessarily be the optimal EEP, has a small performance loss relative to the optimal UEP. Our results not only confirm this fact, but our method provides the optimal EEP at every transmission rate. All we have to do is to evaluate (1) via (5) at several equal code rates and get the minimum value.

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 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 obstacle for the system designer. Finally, the optimal EEP is simpler to implement since the code rate is the same for each block over a large group of blocks whereas for an UEP scheme the parity per each block may vary.

IV. 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, which is particularized for individual image and image sequences does not require any decoding at the receiver in order to estimate the D-R. 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 an optimization technique, and evaluating (1) at several points was sufficient to obtain the optimal EEP. The new accurate D-R estimation proposed in this letter 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.

References

[1] A. Said and W. A. Pearlman, "A new, fast, and efficient image coded based on set partitioning in hierarchical trees," *IEEE Trans. Circuits Syst, Video Technol,*, vol. 6, pp. 243–250, June 1996.

[2] S. Cho and W. A. Pearlman, "A full-featured, error resilient, scalable wavelet video codec based on the set partitioning in hierarchical trees (SPIHT) algorithm," *IEEE Trans. Circuits Syst, Video Technol,*, vol. 12, pp. 157–171, 2002.

[3] S. Appadwedula, D. L. Jones, K. Ramchandran, and I. Konzintsev, "Joint source-channel matching for wireless communications link," in *Proc. IEEE Int. Conf. Communications (ICC 98)*, vol. 1, 1998, pp. 482–486.

[4] A. Nosratinia, J. Lu, and B. Aazhang, "Source-channel rate allocation for progressive transmission of images," *IEEE Trans. Commun.*, vol. 51, pp. 186–196, Feb. 2003.

[5] L. Qian, D. L. Jones, K. Ramchandran, and S. Appadwedula, "A general joint source-channel matching method for wireless video transmission," in *Proc. Data Compression Conf.*, Snowbird, UT, 1999, pp. 414–423.

[6] Y. Charfi, R. Hamzaoui, and D. Saupe, "Model-based real-time progressive transmission of images over noisy channels," in *Proc. IEEE Conf. Wireless Communications and Networking*, vol. 2, Mar. 2003, pp. 784–789.

[7] 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.

[8] F. W. Wheeler, "Trellis source coding and memory constrained image coding," Ph.D. dissertation, Rensselaer Polytechnic Institute, Troy, NY, 2000.

[9] S. B. Wicker, *Error Control Systems*. Englewood Cliffs, NJ: Prentice-Hall, 1995.