Quality Evaluation for Robust Wavelet Zerotree Image Coders

Navid Serrano Dirck Schilling Pamela C. Cosman
Dept. of Electrical and Comp. Engineering, Univ. of California at San Diego
La Jolla, CA 92093-0407 {dschilli,pcosman}@code.ucsd.edu

Abstract

The output of wavelet zerotree-style coders is typically very sensitive to channel errors, and several strategies have been proposed in the literature for combatting this problem. In particular, the two solutions of forward error correction and error-resilient packetization, produce images with very different types of degradations (global blurrines versus more localized distortions), making comparison difficult. In this work, we compare the output of these coders by having human observers use the images in a recognition task, and also rate the images subjectively. The results indicate that the localized distortions produced by the error-resilient packetization allow recognition of the image content at lower PSNR, and are also more subjectively pleasing at the same PSNR.

1 Introduction

Wavelet zerotree image coding techniques developed by Shapiro (EZW) [1], and further refinements by Said and Pearlman (SPIHT) [2] provide excellent image compression in terms of distortion-rate performance. However, these coders have a significant vulnerability to channel errors, since a single bit in error can potentially cause the decoder and encoder to lose synchronization for the remainder of the bit stream. This sensitivity to errors has been addressed in a number of different ways. In [3], forward error correction (FEC) is added to the SPIHT bit stream. In [4, 5], the SPIHT output is manipulated into fixed-length packets which are independently decodable. This algorithm is called the packetized zerotree wavelet (PZW) coder. In [6], these approaches are combined into a hybrid coder in which the output is both packetized and protected by FEC, and is then capable of withstanding both packet erasures and bit errors. Quality evaluation for these coders is a complex problem. The decoded PSNR depends on the source coding rate, the channel coding rate, and the pattern of bit errors and packet erasures on the simulated bursty error channel and packet erasure channel. Furthermore, the decoded PSNR is itself only a mediocre measure of the image quality, since the packetized and unpacketized (original) zerotree coding approaches produce distortions with very different visual appearances, in one case producing local distortion (blotches) and in the other a more global blurriness. Figure 1 shows an example of a test image compressed by PZW (and subjected to packet loss) and compressed by SPIHT. The two images have the same PSNR of 23.97.

The current work is concerned with evaluating the perceptual impact of the distortion for these two types of coders. We show that human observers are able to recognize the contents of images at a lower PSNR, for one particular recognition task, for the images with the local distortion than with the global distortion, and they also prefer those images subjectively. The paper is organized as follows. In section 2, we discuss the various robust image coders. In section 3, we discuss the methods and results of our objective recognition study, and in section 4 we discuss the subjective quality evaluation.

2 Robust wavelet zerotree coding

The transmission of images across noisy channels is fundamentally important in many applications and is still an active research problem. One basic approach has been to start with a high-performance source coder, and protect its output from errors by adding redundancy. Another method has been to

Figure 1: Example images of PZW (left) and SPIHT (right) at PSNR=23.97

design resilience into the source coder so the effect of channel errors is reduced and less channel coding is necessary.

The excellent compression performance of the SPIHT [2] source coder comes at the expense of a significant sensitivity to errors. Errors often lead to a complete loss of synchronization in the decoder due to the use of variable length coding. In [3], SPIHT is followed by a strong concatenated channel code (RCPC/CRC) which lowers the probability of decoding errors, thereby providing protection against synchronization loss. Also, the CRC allows detection of uncorrected packets so the source decoder can stop decoding before errors propagate and corrupt the image. Since the underlying source coder (SPIHT) outputs an embedded bit stream, simply stopping the decoding at the first uncorrectable packet and discarding the remainder of the image transmission means the decoder can still reconstruct the full-size image corresponding to that lower source coding rate. For a binary symmetric channel of known error rate, this method often produces acceptable image quality due to the progressive nature of the source coder.

Alternatively, source coding can be designed to provide noise robustness without explicit error-correction coding. The PZW coder [5] provides robustness by producing a compressed image datastream consisting of independently decodable packets. PZW is an error-resilient variation on the EZW and SPIHT coders [2, 1]. Groups of wavelet coefficient trees are placed together into fixed length packets (typically hundreds of bits) with a 16-bit CRC for error detection. At the receiver, packets received with detected errors are discarded; others are decodable independent of any other packet. Missing trees of wavelet coefficients are concealed by interpolating missing low-low band wavelet coefficients; missing higher band coefficients are set to zero prior to inverse wavelet transforming the array.

The growing and pruning of coefficient trees in order to fit fixed-length packets, as well as the addition of a small header, induce some performance loss, but they provide robustness against packet loss. Errors cannot propagate beyond packet boundaries. Synchronization is not lost if packets are dropped. Packets are of equal importance; given a certain packet loss rate; it matters little to the final PSNR which packets were lost.

The hybrid coder [6] combines the FEC and error

resilience approaches. It uses the PZW algorithm for source coding, and each packet is protected by the RCPC/CRC code from [3]. The RCPC/CRC code is designed for channel conditions in the middle of the expected range. Some bit interleaving is also used to improve performance of the RCPC codes on the bursty channels. The resulting datastream is better suited to handle a larger range of channel conditions. Where a packet erasure would truncate the bitstream using the SPIHT+RCPC/CRC coder, the hybrid system can use all received packets. PZW on its own cannot handle arriving packets with errors, but the hybrid has the RCPC/CRC code to correct bit errors making received packets useful to the source decoder. In essence, the two approaches used together are intended to help fix the weaknesses of each other.

In comparing these various coders for noisy channels, it was apparent that channel errors provoked very different types of degradations in them, and the results could not be adequately summarized by PSNR.

3 Objective recognition experiment

At best, evaluation of image coders using PSNR is of questionable perceptual validity, and considerable research has been devoted towards developing computable metrics of image quality that have a higher correlation with perceptual quality than does PSNR. Another approach to evaluating image quality is to have human observers look at the images and provide either a subjective rating of the quality, or else an objective decision of some sort, e.g, a decision on the image contents. This latter approach was taken in [7, 8] and we follow in large part the evaluation methods described in these works. A database of 68 grayscale images was collected. Half of the images showed men, and half showed women. All images were of size 512×512 pixels. Some images showed a single person; others showed a group of people. In the case where a group of people was shown, all the people in the picture were of the same sex. The people were not always located in the center of the picture, and they appeared at different sizes. One evaluation experiment involved objective decisions about image contents, and a second experiment involved subjective ratings.

Each of the 68 test images was compressed using the PZW algorithm to a target bit rate of 0.23 bits per pixel. The actual bit rate might depart slightly from the target bit rate because of the adjustments in wavelet coefficient quantization precision as the coefficient trees are grown or pruned to fit exactly into fixed length packets. The target rate of 0.23 bpp led to a high quality decoded image (typically about 40 dB) and required about 180 packets. The channel-degraded versions of these images were produced by dropping some packets and decoding the remainder. Some packets cause more damage than others to the PSNR when dropped. By trying many different random combinations of different packets to be dropped, we created a sequence of (typically) 20 degraded versions of each of the test images. The sequence of degraded versions had PSNRs ranging between 10 dB and 40 dB, with increments of at least 1 dB between successive images in the sequence.

Each test image was also compressed by the SPIHT algorithm at different bit rates logarithmically increasing from 0.001 bpp to 0.5 bpp. Twenty versions of the image were saved for each image, and PSNRs for these images also corresponded to a range from 10 dB to 40 dB.

The basic idea of the objective recognition portion of the experiment was to show an observer a sequence of degraded versions of an image at successively increasing PSNRs. The observer cannot tell initially whether the image is of a man or a woman. As frames are shown every 500 msec at successively higher PSNRs, at some point the observer can recognize the image contents. The observer is asked to click with the mouse or hit a key on the keyboard as soon as he or she is reasonably confident that the image contents have been recognized. Clicking with the mouse or hitting a key on the keyboard causes the image to disappear from the screen; the observer is then queried as to whether the image showed a man or a woman. After answering, the observer continues on to the next image in the test set.

This portion of the experiment involved a total of 15 observers, who were drawn from the general

university population, had normal or corrected-to-normal vision, and were paid for their participation. Each observer saw each of the 68 images in exactly one sequence, either with the PZW compression or with the SPIHT compression. The selection of PZW or SPIHT was randomized, as was the order in which the images were displayed. The general setup of this experiment is very similar to that described in [7, 8], in which the progression of successive images was aimed more specifically at evaluating increasing bit rates, rather than increasing PSNRs.

Figure 2 shows the number of recognitions (observer responses) that occurred at each PSNR, versus the PSNR, for both SPIHT and PZW. The data look approximately normal. Figure 3 shows the cumulative distribution for these responses as a function of PSNR.

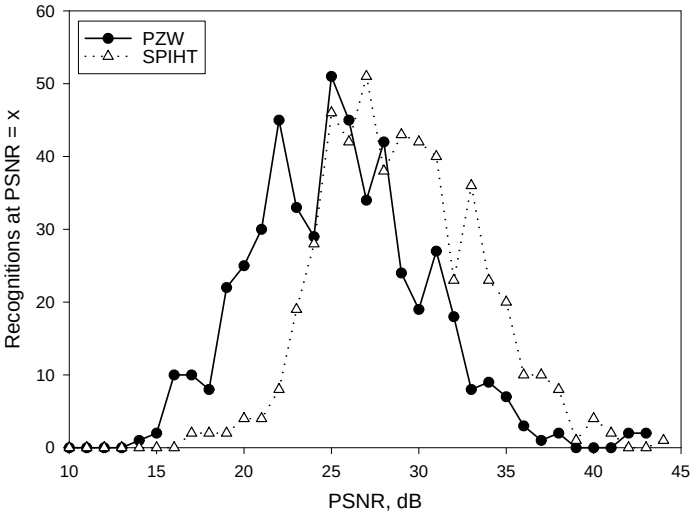

Figure 2: Percentage of observer responses as a function of PSNR for SPIHT and PZW

For a given image, the 20 frames compressed by SPIHT were not matched in PSNR, frame-by-frame, to the frames generated by PZW. The SPIHT sequences tended to run at slightly higher PSNRs, as shown in Figure 4 for one particular image in the test set. For all test images, the SPIHT sequence started out initially at a higher PSNR. Because of this, one could wonder whether the results displayed

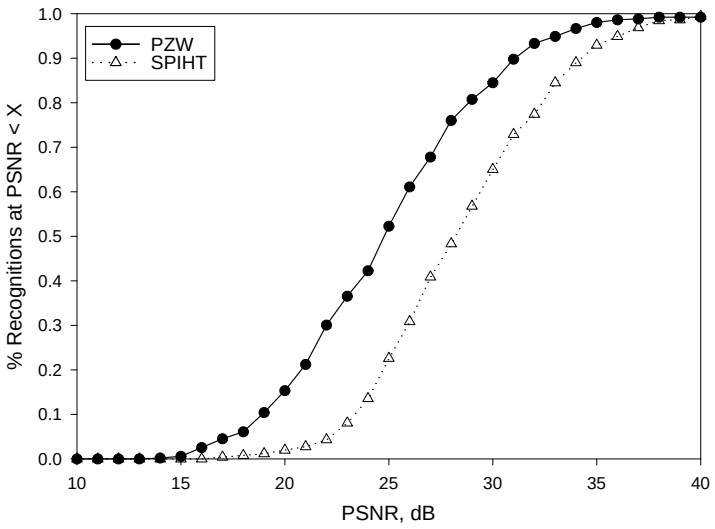

Figure 3: Cumulative distribution plot of observer responses as a function of PSNR for SPIHT and PZW

in Figures 2 and 3 might merely be reflecting a situation in which observers take a certain more-or-less fixed amount of time to recognize a given image, or to respond to its display by clicking a mouse button, and that the PZW sequences allow recognition at a lower PSNR simply because those sequences have lower PSNRs initially. That this is not the case is shown by Figure 5, in which the cumulative distribution plot of observer responses is shown as a function of time. It shows that people responded *sooner* in time for the PZW sequences, despite the fact that they were observing *lower* PSNR values during that time.

Statistical analysis: Denote the PSNR at which observer i answers a question for image j compressed by algorithm A as r_{Aij} and the corresponding PSNR for algorithm B as r_{Bij}, and let $i = 1, \ldots, I$ and $j = 1, \ldots, J$ be indices for observers and images. In this case, $I = 15$ and $J = 68$. Since each observer saw each image sequence compressed with only one of the two algorithms, we have for a given i and j a value for either r_{Aij} or r_{Bij} but not both. For comparing algorithms A and B, we would like to know mean values, r_A and r_B, and whether any difference in these values should be deemed statistically significant. Similarly, one can look at the

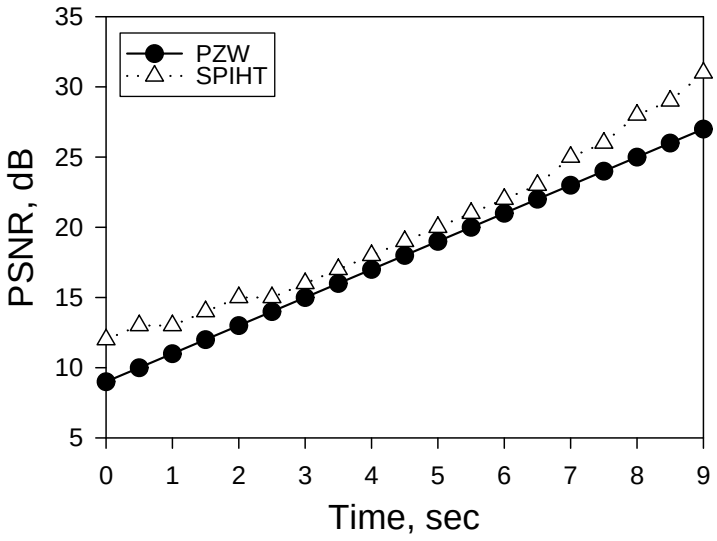

Figure 4: PSNR versus time for SPIHT and PZW for one particular image

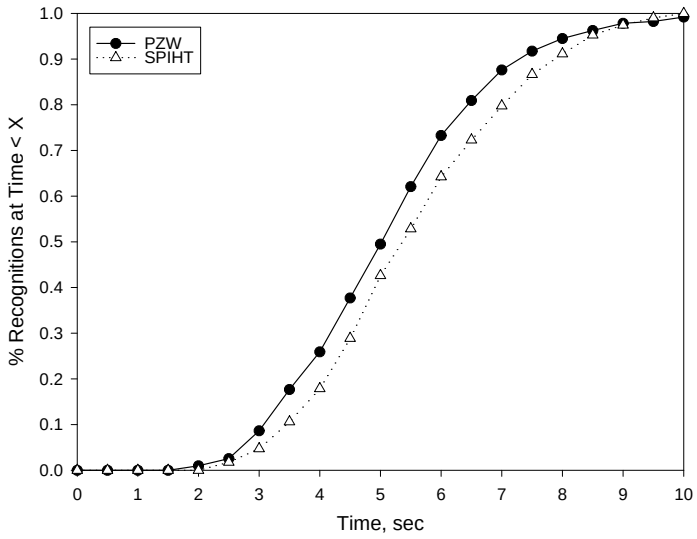

Figure 5: Cumulative distribution plot of observer responses as a function of time for SPIHT and PZW

time t_{Aij} at which observer i answers a question for image j compressed by algorithm A, and whether there is a significant difference in mean time t_A and t_B.

These analyses are carried out by fitting the data to a mixed effects linear model in which the compression algorithm is treated as a fixed effect, and the observers and images are treated as random effects. Model fitting is performed using restricted maximum likelihood (REML) in the *Splus* function `varcomp` [9]. Under this analysis, the mean PSNR for PZW responses was found to be 25.43 dB, whereas it was 28.97 dB for SPIHT. The 95% confidence interval for the difference of means extended from -3.83 to -3.24. Since the confidence interval does not include zero, we can conclude that the PSNR required for observers to answer the question for the PZW images is significantly less than that required for SPIHT images at the 95% confidence level. When applied to time, the t_{Aij} values were taken to be frame numbers, where frames were shown 500 msec apart. The mean time (frame number) for PZW was 10.72, and was 11.59 for SPIHT. The 95% confidence interval for the difference of mean time extended from -1.17 to -0.58, and again does not include zero. So we can conclude that observers answered the questions at significantly faster time with PZW, despite the fact that they were answering them at significantly lower PSNR.

Error analysis: In addition to the bit rates, the responses of the observers to the questions were recorded and examined for correctness. It would be possible, in theory, that one algorithm might lead people to make rapid yet incorrect decisions. A paired-sample Wilcoxon signed rank test was used to examine the number of errors each observer made under each algorithm. Denoting observer i's error count for PZW as x_i and that for SPIHT as y_i, the test statistic T^+ is the rank sum of those $abs(x_i - y_i)$ with $x_i - y_i > 0$. For 15 or more observers, the distribution of T^+ is reasonably approximated as normal. For individual observers, the number of errors ranged from 0 to 5 for both PZW and SPIHT. The Wilcoxon 2-sided signed-rank test had a p-value of 0.749 for the comparison of the observer errors, and thus was not significant at the 5% significance level. The overall error rates for each algorithm – 5.3% for PZW and 5.1% for SPIHT – support the conclusion that neither algorithm was more likely than the other to lead to incorrect responses.

4 Subjective evaluation

In the subjective evaluation, 5 observers were asked to provide subjective ratings of images on

a 5-point scale. Each image was compressed both by SPIHT and by PZW. For a given image, the SPIHT and PZW versions were at the same PSNR (within a maximum deviation of 1 dB), but for different images, the PSNRs were different, ranging from 12 to 40 dB. As before, the low PSNR values for PZW were achieved by dropping packets, rather than just by encoding at a lower source coding rate. Each observer saw a total of 320 images, which came from encoding 40 different test images at 4 different PSNRs, using the 2 different methods. At each PSNR, and for each compression method, the ratings were averaged across all observers, and across all images that were shown at that PSNR. These averages are shown in Figure 6. As one would expect, there is a trend towards higher subjective ratings for higher PSNRs, for both compression methods. For the low quality images, the average ratings for PZW images are higher than those for SPIHT, but for high quality images this distinction seems to disappear.

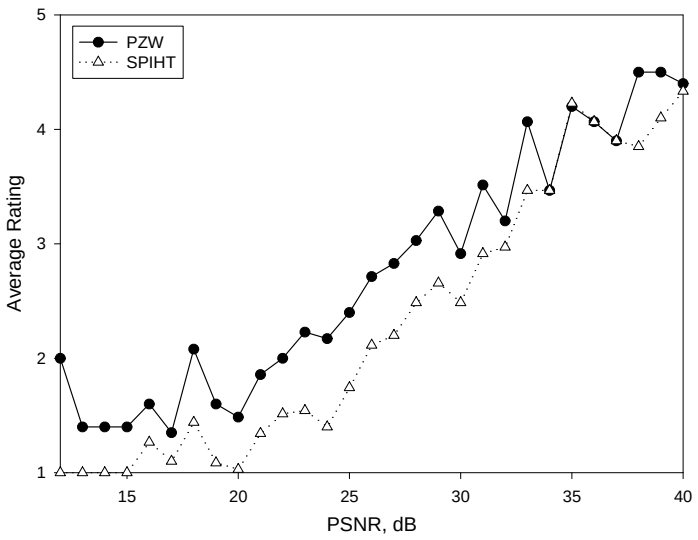

Figure 6: Average subjective score versus PSNR for SPIHT and PZW

The data were analyzed by using paired differences. For a given observer, image, and PSNR, the score for PZW was paired with the score for SPIHT for the same observer, image, and PSNR. There was a total of 790 such pairs. The score for SPIHT was subtracted from the score for PZW, and these differences were analyzed as before using restricted maximum likelihood (REML) in the *Splus* function `varcomp`. When all pairs were included in the analysis, the mean difference was 0.448, and the 95% confidence interval went from 0.305 to 0.591. Since the value 0 was not included in the confidence interval, we can conclude that, at the same PSNR, PZW is significantly subjectively superior to SPIHT, with a mean subjective score nearly a half point higher on a 5-point scale. The PSNR range was also broken down in separate groups, for separate analyses of the data. Taking those pairs with PSNRs in the ranges 0–20 dB, 0–35 dB, 20–35 dB, and 20–40 dB produced similar results to the case where all pairs were included. That is, in all cases PZW was significantly superior in subjective score, with a mean difference ranging from 0.42 to 0.50 over SPIHT. Only for the highest PSNRs (35–40 dB) which comprised 130 data pairs was the difference between PZW and SPIHT not statistically significant.

5 Conclusions

On one level, the conclusion of this paper is that all PSNR is not created equal. In comparing two algorithms that produce images at increasing PSNRs, people are able to recognize images at a lower PSNR with one algorithm than with another, in a statistically significant way. A second conclusion of the paper is the suggestion that localized distortion may perhaps be preferred both objectively and subjectively over global distortion.

More concretely, this paper provides a visual comparison of some different coders which have been proposed for use on noisy channels. The PZW coder uses independently decodable fixed-length packets, and the output of this coder over a noisy channel could be simulated by coding to a given bit rate and then dropping packets and decoding. The SPIHT+RCPC/CRC coder uses channel coding to protect the output of a SPIHT coder. When errors occur, the channel coder may be able to correct them entirely, in which case the output of the system is simply the SPIHT-coded image at the underlying source coding rate. If the errors cannot be corrected, the decoder stops decoding at the first block with uncorrectable errors, in which case the

output of the system is again the SPIHT-coded image at the lower source coding rate corresponding to the error-free initial blocks. Either way, the decoded image is simply a SPIHT-coded image, and so a collection of output images from this joint source-channel coder could be obtained simply by using the SPIHT coder at various bit rates. Similarly, sample output images from the hybrid coder [6] can be gotten by simply using the output PZW images after packet dropping; the effect of the RCRP/CRC would be to lower the rate accorded to the source code, but the type of visual distortions would not be changed. Thus the comparison in this paper of PZW (with dropped packets) versus SPIHT (at lower source coding rates) directly provides insight into the comparison of PZW versus SPIHT+FEC versus PZW+FEC coders for noisy channels. One can conclude that the comparative PSNR plots and numerical results given in [6, 4, 5] are quite conservative on the side of underestimating the benefits of the PZW and hybrid coders.

Acknowledgements: The authors thank Prof. Charles Berry of the Department of Family and Preventive Medicine at UCSD for advice on the statistical analysis. This work was supported by the National Science Foundation and by the Center for Wireless Communications at UCSD.

References

[1] J. Shapiro. Embedded image coding using zerotrees of wavelet coefficients. *IEEE Transactions on Signal Processing*, 41(12):3445–3462, December 1993.

[2] A. Said and W.A. Pearlman. A new, fast, and efficient image codec based on set partitioning in hierarchical trees. *IEEE Trans. on Circuits and Systems for Video Technology*, 6(3):243–250, June 1996.

[3] P.G. Sherwood and K. Zeger. Progressive image coding on noisy channels. In J.A. Storer and M. Cohn, editors, *Proceedings of the 1997 IEEE Data Compression Conference (DCC)*, pages 72–81, Snowbird, Utah, March 1997. IEEE Computer Society Press.

[4] J.K. Rogers and P.C. Cosman. Wavelet zerotree image compression with packetization. *IEEE Signal Processing Letters*, 5(5):105–107, May 1998.

[5] J.K. Rogers and P.C. Cosman. Robust wavelet zerotree image compression with fixed-length packetization. In J.A. Storer and M. Cohn, editors, *Proceedings of the 1998 IEEE Data Compression Conference (DCC)*, pages 418–427, Snowbird, Utah, March 1998. IEEE Computer Society Press.

[6] P. Cosman, J. Rogers, P.G. Sherwood, and K. Zeger. Image transmission over channels with bit errors and packet erasures. In *Proceedings of the 32nd Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, November 1998. IEEE Comput. Soc. Press.

[7] S. Cen, H. Persson, D. Schilling, P. Cosman, and C. Berry. Human observer responses to progressively compressed images. In *Proceedings of the 31st Asilomar Conference on Signals, Systems and Computers*, volume 1, pages 657–661, Pacific Grove, CA, November 1997. IEEE Comput. Soc. Press.

[8] D. Schilling, P. Cosman, and C. Berry. Image recognition in single-scale and multiscale decoders. In *Proceedings of the 32nd Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, November 1998. IEEE Comput. Soc. Press.

[9] W. Venables and B. Ripley. *Modern Applied Statistics with S-Plus*. Springer Verlag, New York, 1994.