

# Lagrangian Optimization of a Group Testing for ENO Wavelets Algorithm<sup>1</sup>

Diego Dugatkin

Dept. of Electrical Eng.  
Caltech, Pasadena CA 91125  
diego@caltech.edu

Hao-Min Zhou

Dept. of App. & Comp. Math.  
Caltech, Pasadena CA 91125  
hmzhou@acm.caltech.edu

Tony F. Chan

Dept. of Math. and IPAM  
UC Los Angeles CA 90095.7121  
chan@math.ucla.edu

Michelle Effros

Dept. of Electrical Eng.  
Caltech, Pasadena CA 91125  
effros@caltech.edu

**Abstract — We consider the generalization of multi-resolution optimization techniques to the Group Testing for Wavelets (GTW) image compression algorithm. The resulting algorithm yields a principled technique for combining standard wavelet coefficients with adaptive Essentially Non-Oscillatory wavelet coefficients in a single coding framework and allows rate-distortion optimization of the quantizer’s wavelet coefficient approximations according to a Lagrangian performance measure. At the resolutions of highest interest, the resulting algorithm achieves gains of approximately 0.7 dB over the GTW algorithm.**

## I. INTRODUCTION

Multi-resolution (MR) data compression systems, also known as progressive transmission or successive refinement codes, yield embedded binary source descriptions with a reproduction quality that improves as the decoded rate increases.

In general, a single MR code cannot achieve the same performance at all resolutions as the best single-resolution codes with the corresponding rates. As a result, MR code design involves prioritization over the resolutions; we achieve better performance at the resolutions of highest interest by allowing performance degradation at resolutions of lower importance. MR optimization is accomplished through the use of Lagrangian performance measures that combine the rates and distortions at all resolutions into a single weighted sum [1, 2]. MR Vector Quantization (MRVQ) [3] and Opt-SPIHT [4, 5] are examples of MR codes optimized for the Lagrangian performance measure. These algorithms perform the MR Lagrangian optimization in the VQ and the Set Partitioning in Hierarchical Trees (SPIHT) [6] frameworks respectively.

In this paper, we apply the MR Lagrangian within a group-testing framework based on the Group-Testing for Wavelets (GTW) [7] algorithm. Two main goals underlie this approach. First, we allow MR optimization within the GTW framework. Second, we introduce a principled approach for choosing between wavelet and Essentially Non-Oscillatory (ENO) wavelet [8, 9] representations. The first goal is motivated by the desire to generalize the techniques developed in [4] from the SPIHT coding framework to the GTW approach to bit-plane coding, which achieves consistently good performance without requiring an arithmetic code. The generalization of the optimization is non-trivial since Opt-SPIHT’s low complexity, globally

optimal dynamic programming solution to the Lagrangian optimization has no natural analog in the GTW framework. Instead, we develop a low-complexity algorithm for estimating the Lagrangian and base our decisions on the resulting estimates.

The second goal is motivated by the hypothesis that ENO coefficients, which reduce the magnitude of high-frequency coefficients in a wavelet decomposition while preserving the total energy of the image, are often better suited to bit-plane coding than standard (STD) wavelets. While ENO’s energy compaction properties generally reduce the rate required for the bit-plane description, the use of ENO coefficients requires the description of an ENO edge map in addition to the bit-plane description. We therefore use the MR Lagrangian to determine when the trade-off between the rate used to describe the ENO map and the rate saved in the bit-plane description makes using the ENO coefficients worthwhile.

The paper is organized as follows. Section II provides an overview of ENO and GTW. Section III describes the proposed optimization algorithm. Section IV contains experimental results comparing the performance of GTW and the new method. Experimental results on the cameraman image show up to 0.7 dB improvement over the GTW coding performance at the resolutions of highest priority.

## II. BACKGROUND

### MR Lagrangian

For any MR code with performance on the lower convex hull of achievable rate-distortion vectors, there exist non-negative constants  $\{\alpha_\ell, \lambda_\ell\}_{\ell=1}^L$  such that the given code’s performance minimizes the Lagrangian performance measure  $J = \sum_{\ell=1}^L \alpha_\ell (D_\ell + \lambda_\ell R_\ell)$  over all MR codes. Here  $D_\ell$  and  $R_\ell$  denote the code’s expected distortion and expected total rate at resolution  $\ell$ . We therefore use  $J$  as our performance criterion in MR code design. Intuitively,  $\alpha_\ell$  describes the priority on the  $\ell^{\text{th}}$  resolution and  $D_\ell + \lambda_\ell R_\ell$  is the familiar rate-distortion Lagrangian for resolution  $\ell$ . Here the vector  $(\alpha^L, \lambda^L)$  characterizes the direction of a hyper-plane tangential to the lower convex hull of the achievable rate-distortion region at a single point.

### ENO Wavelets

The adaptive ENO wavelet transform [8, 9] reduces the ringing effects of image edges by modifying the STD wavelet coefficients near discontinuities. The key idea is that the ENO wavelet transform does not allow filtering across discontinuities. Instead, the ENO wavelet transform uses independent transformations on the two sides of each discontinuity. The associated overhead in computational complexity is low, and the required number of ENO wavelet coefficients is identical

<sup>1</sup>The work done in the Electrical Engineering Department at Caltech was supported in part by an Oringer Fellowship, an H. Bauer Fellowship, the Lee Center for Advanced Networking at Caltech, Intel, and NSF Award No. CCR-9909026. Hao-Min Zhou is supported in part by grants ARO-DAAD19-99-1-0141 and NSF-DMS-0073916. Tony Chan is supported in part by grants NSF-DMS-9973341 and ONR N00014-96-1-0277.

to the required number of STD wavelet coefficients; thus the ENO transform retains many of the advantages of the STD transform while removing edge artifacts. In particular, the ENO wavelet transform provides an MR decomposition with good energy compaction and uniform high order accuracy with reduced ringing effects at region boundaries. The cost of this improvement is the rate needed to describe the boundary locations.

For smooth regions of an image, the ENO transform generates the same wavelet coefficients as the STD transform. For edge regions, the ENO wavelet transform uses extrapolation techniques to perform independent transforms on the two sides of each discontinuity. Therefore, the result is a reduction in high-frequency coefficients relative to the STD wavelet transform. This approach also preserves the sharpness of the discontinuity without smearing. This property is useful for embedded zero-tree coding algorithms. Since it reduces the high-frequency coefficients and concentrates most of the larger coefficients at the top of the wavelet sub-band pyramid, it allows for a rate-efficient description of the coefficients.

In the ENO wavelet transform, a thresholding parameter is used to identify edges in the image data. In adjusting the minimum threshold, we have the flexibility to go from no edge detection (and therefore a STD wavelet transform) to maximum edge sensitivity. A map of edge locations is necessary to make the transform reversible. Increasing the edge sensitivity increases the number of non-zero values in the ENO edge map and, in most cases, also the corresponding description cost. If no edges are detected, the map contains only zeros and can be described at almost no cost.

#### GTW Algorithm

GTW, like SPIHT and other recent MR bit-plane coders, encodes the coefficients of the wavelet decomposition of an image bit-plane by bit-plane. The encoding requires two consecutive passes per bit-plane. First, the *significance pass* codes all coefficients that were not identified as significant in a prior bit-plane. (A coefficient is declared to be significant at bit-plane  $i$  if the first non-zero bit of its binary description occurs in the  $i^{\text{th}}$  most significant position.) Second, the *refinement pass* for a given bit-plane encodes the next bit in the description of those coefficients that were already declared significant in prior bit-planes. GTW's refinement pass is the same as SPIHT's, but its significance pass is different.

In GTW the significance pass involves dividing the bits within each bit-plane into classes and then performing a group-testing algorithm on groups within each class. A class is defined by the combination of three of the coefficient's properties: the *sub-band level*, the *pattern type*, and the *significant neighbor metric* (SNM). The level describes the coefficient's position in the wavelet decomposition. The pattern is a function of the coefficient's location, with neighboring coefficients given different pattern values. The SNM describes the number of neighbors that were declared significant at previous bit-planes or in the same bit-plane but before the given coefficient. The classes are ordered first by SNM, then by pattern type, and lastly by level, and the algorithm encodes all classes in order.

The encoding of each class involves sequentially breaking off a group from the class, describing the contents of that group to the decoder, and using those contents to determine the number of bits in the next group.

The first group in each class contains one bit. The number of bits per group doubles with each subsequent group until the encoder describes a group that has at least one non-zero bit. Subsequent group sizes are based on the empirical probability of a zero in the preceding bits in the given class. In particular, the size of each subsequent group equals the unique integer  $k$  that satisfies the inequality

$$q^k + q^{k+1} \leq 1 < q^k + q^{k-1}, \quad (1)$$

where  $q$  is the empirical probability of a zero when the group size is chosen. The first bit of any group description describes whether or not the group is significant (i.e., whether or not it contains any 'one's'). If the group is not significant, then the group description is complete. If the group is significant, then it is split into two parts. Each part is subsequently tested and, if found significant, split again. The process repeats until the first significant element is found within that group. Once a significant element is found, the code updates  $q$  and  $k$  and defines the next group to be the  $k$  elements immediately following that significant bit. The process repeats on the subsequent groups within the class until every bit in the class has been described or the number of bits remaining is smaller than the current value of  $k$ . Remaining bits join a list of untested items and the algorithm proceeds with bits from the next class.

### III. ALGORITHM

The goal of the proposed scheme is to alter an input image's wavelet coefficients to the values for which GTW yields the best Lagrangian performance. Allowed alterations include both replacement of STD wavelet coefficients with ENO coefficients and modifications of either STD or ENO coefficient values.

In deciding between ENO and STD wavelet coefficients, we calculate the optimal MR Lagrangian for each and choose the representation with the better Lagrangian performance. Each ENO decision affects a collection of coefficients and thus both Lagrangians sum up the expected performance values over all affected coefficients. The Lagrangian for the ENO coefficients also includes the cost of describing the ENO map (compressed using a simple entropy coder). The process is repeated for all ENO decisions being considered.

To calculate the *optimal* MR Lagrangian for a collection of STD or ENO coefficients, we consider possible modifications of those coefficients like those considered in [4]. The coefficient modifications are motivated by the observation that accurate coefficient description sometimes yields MR Lagrangian performance that is inferior to that of the best inaccurate description. In particular, accurate bit-plane descriptions are necessary to guarantee lossless performance at high enough rates, but requiring eventually lossless performance may actually harm the performance at the resolutions of greatest interest to the system user.

While optimal encoding in MRVQ and Opt-SPIHT is computationally feasible, true MR-Lagrangian optimization is computationally prohibitive in GTW. Roughly, the problem arises from a combination of the class definitions and the empirical distributions used in GTW. In particular, neither of these items can be easily calculated without running the algorithm in full, and both can vary enormously with small changes in the coefficients. As a result, we cannot afford to do global optimization of the MR Lagrangian over the set of allowed coefficient modifications. Instead, we estimate the class

index and corresponding empirical probability used to code each coefficient, and we make the coefficient modifications that minimize the resulting approximation of the expected value of the Lagrangian. In calculating these approximations, we trade off performance and complexity. The resulting algorithm does not guarantee optimal rate-distortion performance within the GTW framework but does provide a principled method for improving on the existing performance at reasonable computational expense. Details of the Lagrangian approximation techniques follow.

### Probability and Group-size Estimation

Recall that GTW codes symbols in groups of size  $k$ , where  $k$  is a function of the current empirical probability  $q$  as described in (1). To estimate the expected rate and distortion of our code, we first estimate the members of each group and the corresponding  $q$  and  $k$  values. Since calculating  $q$  and  $k$  effectively requires running the full GTW algorithm and since those values can change significantly with coefficient modifications, precise calculation of the  $q$  and  $k$  values for every possible combination of coefficient modifications is impractical. We therefore estimate  $q$  and  $k$  for each class using a procedure designed to do well on average across many coefficients.

The significance probability in a given class depends on the proportion of ones and zeros in that class. Recall that each bit of each coefficient is assigned a class defined by its level, pattern, and SNM. While the first two parameters are fixed, the SNM depends on the order in which classes are coded and the modified values chosen for neighboring coefficients. Thus we begin the estimation procedure by estimating SNM.

Let  $C(i, j)$  denote the STD or ENO wavelet coefficient at spatial coordinates  $(i, j)$ . We use a two step procedure to estimate the SNM of  $C(i, j)$  at resolution  $\ell$ . The first step is a coarse estimate. If a neighbor of coefficient  $C(i, j)$  is significant at or before resolution  $\ell$ , then that neighbor is counted as significant for the SNM of  $C(i, j)$  at resolution  $\ell$ . The second step refines the initial coarse estimate by decreasing the SNM estimate by the number of  $C(i, j)$ 's significant neighbors that are likely to be declared significant *after* the description of  $C(i, j)$ . The procedure for refining the SNM estimates relies on the order in which the classes are described, which in turn relies on the first-pass SNM estimate.<sup>2</sup> Only neighbors that have already become significant can be counted in the SNM during GTW encoding since the encoder's SNM calculation must be mirrored at the decoder. While the first SNM estimate for the coefficients can be done in any order, the second step takes the coefficients one by one in zig-zag order through each sub-band of the wavelet decomposition. This approach mimics GTW's adaptive estimation of the SNM.

Given the SNM estimates, we can estimate the class of every bit in the wavelet decomposition. Using these estimates, we calculate the insignificance probability  $q$  as the empirical probability of a zero in the estimated class. As in GTW, the estimated group size  $k$  is computed using (1).

<sup>2</sup>For neighboring coefficients that become significant in the same bit-plane, only those with larger SNM, or those with equal SNM but lower pattern index, or with equal SNM, equal pattern index, and lower level, are taken into account for the calculation. In cases where all three parameters match (which could happen with sibling coefficients), only those already declared significant (in zig-zag order through each level and sub-band) are counted.

### Lagrangian Calculation

Given the estimated class values and the corresponding estimates of  $q$  and  $k$ , we next estimate the MR Lagrangian. While the MR Lagrangian can be calculated for an arbitrary number of resolutions, for simplicity, we define  $D_\ell$  and  $R_\ell$  to be the distortion and total rate at the end of the description of bit-plane  $\ell$ . Thus the number of resolutions in our code design equals the number of bit-planes in our wavelet calculation. The number of resolutions can be greatly expanded, increasing the flexibility from the user's point of view as well as the coding complexity from the designer's perspective. For instance, we could define the resolutions to correspond to the class descriptions or even the group descriptions.

We calculate an MR Lagrangian for each coefficient  $C(i, j)$  at every resolution  $s \in \{1, \dots, L\} \cup \{\infty\}$  at which  $C(i, j)$  might be declared significant. Here  $s = \infty$  refers to the scenario when  $C(i, j)$  is never declared significant. The resolution  $s$  at which we declare  $C(i, j)$  significant can be the true significance resolution, a prior resolution, or a later resolution. Finding the Lagrangian performance  $\hat{J}_s(i, j) = \sum_{\ell=1}^L [\alpha_\ell D_{s,\ell}(i, j) + \lambda_\ell R_{s,\ell}(i, j)]$  associated with declaring  $C(i, j)$  significant at resolution  $s$  requires the calculation of  $L$  distortions and rates.

For any  $\ell \in \{1, \dots, L\}$ , the distortion  $D_{s,\ell}(i, j)$  equals the squared error between  $C(i, j)$  and the resolution- $\ell$  reconstruction of the optimal reproduction for coefficient  $C(i, j)$  and significance level  $s$ . If the true significance level of coefficient  $C(i, j)$  equals  $s^*$  and  $s < s^*$ , (that is,  $C(i, j)$  is declared to be significant before it actually becomes significant), then the optimal reproduction is the smallest value with significance level  $s$ . If  $s > s^*$ , then the optimal reproduction is the largest value with significance level  $s$ . If  $s = s^*$ , then the optimal reproduction equals  $C(i, j)$ .

Similarly,  $R_{s,\ell}(i, j)$  equals the expected total rate in the resolution- $\ell$  description of  $C(i, j)$ . The expected rate is given by

$$R_{s,\ell}(i, j) = \sum_{m=1}^{\ell} r_{s,m}(i, j),$$

where  $r_{s,m}(i, j)$  is the incremental rate used in the resolution- $m$  description of the coefficient  $C(i, j)$ . The value of  $r_{s,m}(i, j)$  depends on the  $m$ th bit (denoted by  $C_m(i, j)$ ) of the chosen reproduction, on the estimated probability  $q$  of zeros (insignificant elements) in the class, and on the position of that bit within its group of size  $k$  and within its class of size  $N$ . In particular, if we assume that the bit  $C_m(i, j)$  is the  $v$ th element in its class, the incremental rate is determined by:

$$r_{s,m}(i, j) = \begin{cases} \text{Rate}('0', k, v, q, N) & \text{if } C_m(i, j) = 0 \\ \text{Rate}('1', k, v, q) & \text{if } C_m(i, j) = 1. \end{cases}$$

To see why  $r_{s,m}(i, j)$  varies with all of these parameters, first consider the following example. Table 1 shows the expected description lengths for each item of a group of size  $k = 8$ , where the first 7 elements are zeros and the last element is a one. The first column of Table 1 shows the GTW description of this group. The first bit of the description indicates that at least one element in the group is a '1'. The next bit is a zero because the first half of the group (which contains the first 4 items) is insignificant. The following zero specifies that the first half of the remainder is also insignificant. Finally the last zero establishes that of the two items that still remain unidentified, the first is insignificant; therefore the other item

must be the ‘1’. The cost of each output bit is amortized over the items that “benefit” from that expenditure of rate. Therefore, bits that only discriminate between later elements are not counted in the rate calculation for earlier items of the group. Since early output bits provide information required for all elements, we split that rate cost across the full group. The total rate for each member of the original group appears in the last row of the table. Given a group of  $k$  elements for which the first ‘1’ appears in position  $b$ , we use  $\{U_0(a, b, k)\}_{a=1}^{b-1}$  to denote the rate used to describe each of the first  $b-1$  zeros and  $U_1(b, k)$  to denote the rate used to describe the first ‘1’ at position  $b$ . Thus the final row of Table 1 contains entries  $U_0(1, 8, 8), \dots, U_0(7, 8, 8), U_1(8, 8)$ . Table 2 shows  $U_1(b, 8)$  for  $b \in \{1, \dots, 8\}$ . Notice that  $U_1(b, k)$ , is not monotonic in  $b$ .

To estimate  $Rate('1', k, v, q)$ , we find the expected value of  $U_1(b, k)$  with respect to the distribution on  $b$  imposed by  $q$ . (We assume that the symbols in the group are drawn i.i.d. according to distribution  $q$ .) Thus we estimate the probability of a one, for each possible value of  $b$ , as  $(1-q)q^{b-1}$ .

Recall that GTW divides every class into groups of size  $k$  and tests each group for significance. Given a one at position  $v$  within its class, we distinguish two scenarios. Either the one is the first significant item within its class or it is not. The first event occurs at position  $v$  with probability  $q^{(v-1)}$ . In this case the one falls in position  $((v-1)\%k+1)$  in its group, where  $\%$  is the integer mod operator, and  $\lfloor (v-1)/k \rfloor$  all-zero groups precede the group of the current one. In the second case, suppose that the number of zeros observed since the most recent one is  $t$ , where  $0 \leq t \leq v-2$ . In this situation, the position of the current ‘1’ within its group is  $(t\%k+1)$ , and  $\lfloor t/k \rfloor$  all-zero groups immediately precede the given group.

Therefore, by combining all possible scenarios, the expected rate cost of a ‘1’ can be estimated as

$$Rate('1', k, v, q) = q^{v-1}U_1((v-1)\%k+1, k) + (1-q) \left( \sum_{t=0}^{v-2} q^t U_1(t\%k+1, k) \right).$$

In this estimation, the computation of the summation containing  $(v-1)$  terms can be simplified to

$$Rate('1', k, v, q) = q^{v-1}U_1((v-1)\%k+1, k) + (1-q) \sum_{t=kM}^{v-2} q^t U_1(t\%k+1, k) + (1-q) \sum_{t=0}^{k-1} q^t U_1(t+1, k) \frac{(1-q^{kM})}{1-q^k},$$

where the summation index is at most  $k$ , the estimated group size, and  $M = \lfloor (v-2)/k \rfloor$  is the largest integer less than or equal to  $(v-2)/k$ . Similarly, the estimation of the expected rate needed for the description of a ‘0’ in a class having  $N$

total elements is

$$Rate('0', k, v, q, N) = (1-q) \sum_{b=0}^G q^{b+v-1} U_0((v-1)\%k+1, (v-1)\%k+b+2, k) + (1-q)^2 \sum_{a=0}^L \sum_{b=0}^F q^{a+b} U_0(a\%k+1, a\%k+b+2, k) + (1-q) \left( q^{k-1} \frac{1-q^{kM}}{1-q^k} + q^{kM+k-1} \frac{v-kM-1}{k} \right) + q^{kQ+k-1} \frac{1}{k},$$

where  $G = \min\{k-2-(v-1)\%k, N-v-1\}$ ,  $L = \min\{k-2, v-2\}$ ,  $F = \min\{k-2-a\%k, N-v-1\}$  and  $Q = \lfloor (v-1)/k \rfloor$ .

The above rate and distortion calculations allow us to estimate the Lagrangian  $\hat{J}_s(i, j)$  for each coefficient and each value of  $s$ . We perform these estimates for both STD coefficients and, at the position of an edge, ENO coefficients. (In the latter case, the cost of describing the ENO edge map is included in the estimated rate.) We then choose the coefficient values that yield the lowest Lagrangian.

After all of the modifications for all of the coefficients are done, we encode the modified coefficients using the standard GTW encoder. The decoder is identical to the GTW decoder with the exception of the reverse transform, which uses a reverse ENO transform at all edge locations noted in the edge map.

#### IV. EXPERIMENTAL RESULTS

GTW and our algorithm use the same 9-7 tap filters [10] in their wavelet decomposition of gray scale images. The optimization is applied on a 7 level decomposition of STD- as well as ENO-wavelet coefficients. We use the standard gray-scale image “cameraman” of size  $256 \times 256$ .

The plot of Figure 1 compares the peak signal-to-noise ratio (PSNR) as a function of rate, where  $PSNR = 10 \log_{10}(255^2/MSE)$  dB, for both the optimized code and the original GTW algorithm. The graph shows a performance improvement of approximately 0.7 dB.

Figures 2a and 2b show the reconstructed images of the gray-scale “cameraman” at  $R = 0.1$  bits per pixel, using standard GTW with standard 9-7 Antonini wavelets, and the optimized method Opt-GTW with 9-7 ENO wavelets, respectively. The optimization using ENO wavelets allows a perceptual quality improvement that is particularly noticeable at sharp edges in the image. The combination of the Lagrangian optimization with the application of the ENO wavelet transform proposed in this work yields performance improvements both in PSNR and in visual quality of the reconstructed images.

#### ACKNOWLEDGMENTS

We want to thank Ed Hong and R. E. Ladner for providing us with the code for the standard GTW, to which we added the implementation of our algorithm.

#### REFERENCES

- [1] M. Effros, “Practical multi-resolution source coding: TSVQ revisited,” *IEEE Proceedings of the Data Compression Conference*, pp. 53–62. Snowbird, UT, March 1998.

Grp	0	0	0	0	0	0	0	1
Out	Rate cost per item							
1	1/8	1/8	1/8	1/8	1/8	1/8	1/8	1/8
0	1/8	1/8	1/8	1/8	1/8	1/8	1/8	1/8
0	-	-	-	-	1/4	1/4	1/4	1/4
0	-	-	-	-	-	-	1/2	1/2
Tot	2/8	2/8	2/8	2/8	4/8	4/8	8/8	8/8

Tab. 1: Rate costs per item,  $U_0(a, b, k)$  and  $U_1(b, k)$ , for a group of size  $k = 8$  items with its first '1' at position  $b = 8$ . The output of GTW for the given group appears in the Out column.

$b$	Group items	$U_1(b, 8)$
1	1 ? ? ? ? ? ? ? ?	4
2	0 1 ? ? ? ? ? ? ? ?	2
3	0 0 1 ? ? ? ? ? ? ?	2
4	0 0 0 1 ? ? ? ? ? ?	1.25
5	0 0 0 0 1 ? ? ? ? ?	2.4
6	0 0 0 0 0 1 ? ? ? ?	1.3
7	0 0 0 0 0 0 1 ? ? ?	1.619
8	0 0 0 0 0 0 0 1	1

Tab. 2: Expected description lengths of a '1',  $U_1(b, k)$ , for groups of size  $k = 8$  and  $b - 1 = 0, \dots, 7$  zeros before the '1'.

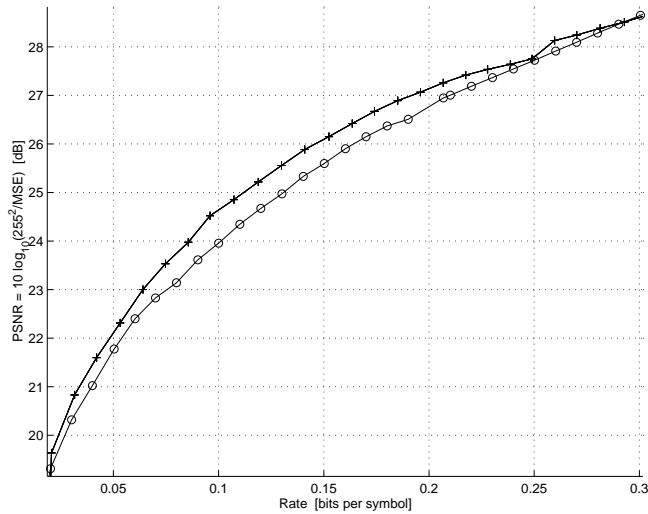


Fig. 1: Comparison of the PSNR as a function of rate for GTW (circles), and the new optimization method with ENO wavelets. The performance improvement of the new method is approximately 0.7 dB.



(a)



(b)

Fig. 2: Reconstructed images of the gray-scale "camera-man" at  $R = 0.1$  bpp, using (a) standard GTW with standard 9-7 Antonini wavelets, and (b) the optimized method Opt-GTW with 9-7 ENO wavelets.

- [2] D. Dugatkin and M. Effros, "Multi-resolution VQ: parameter meaning and choice." *IEEE Conference Record of Thirty-Second Asilomar Conference on Signals, Systems and Computers*, vol. 1, pp. 159–162. November 1998.
- [3] M. Effros and D. Dugatkin, "Multi-Resolution Vector Quantization," *Submitted to IEEE Transactions on Information Theory*, October 2000.
- [4] D. Dugatkin and M. Effros, "Setting Priorities: a New SPIHT-Compatible Algorithm for Image Compression," *SPIE Proceedings of the International Symposium on Optical Science and Technology*, vol. 4119, pp. 799–805. San Diego, CA, July 2000.
- [5] K. K. Lin and R. M. Gray, "Video residual coding using SPIHT and dependent optimization," *IEEE Proceedings of the Data Compression Conference*, pp. 113–122. Snowbird, UT, March 2001.
- [6] A. Said and W. A. Pearlman, "A new, fast, and efficient image codec based on set partitioning in hierarchical trees," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 6, num. 3, pp. 243–250. June 1996.
- [7] E. S. Hong and R. E. Ladner, "Group testing for image compression," *IEEE Proceedings of the Data Compression Conference*, pp. 3–12. Snowbird, UT, March 2000.
- [8] T. F. Chan and H. M. Zhou, "Adaptive ENO-wavelet Transforms for Discontinuous Functions," *Dept. of Math, UCLA, CAM Report, No. 99-21*. June 1999.
- [9] T. F. Chan and H. M. Zhou, "ENO-wavelet Transforms for Piecewise Smooth Functions," *to appear in SIAM J. Numerical Analysis*.
- [10] M. Antonini, M. Barlaud, P. Mathieu and I. Daubechies, "Image Coding Using Wavelet Transform," *IEEE Transactions on Image Processing*, vol. 1, num. 2, pp. 205–220. April 1992.