

# Rate Distortion Optimal Bit Allocation Methods for Volumetric Data Using JPEG 2000

Olga M. Kosheleva, *Member, IEEE*, Bryan E. Usevitch, *Member, IEEE*, Sergio D. Cabrera, *Member, IEEE*, and Edward Vidal, Jr.

**Abstract**—Computer modeling programs that generate three-dimensional (3-D) data on fine grids are capable of generating very large amounts of information. These data sets, as well as 3-D sensor/measured data sets, are prime candidates for the application of data compression algorithms. A very flexible and powerful compression algorithm for imagery data is the newly released JPEG 2000 standard. JPEG 2000 also has the capability to compress volumetric data, as described in Part 2 of the standard, by treating the 3-D data as separate slices. As a decoder standard, JPEG 2000 does not describe any specific method to allocate bits among the separate slices. This paper proposes two new bit allocation algorithms for accomplishing this task. The first procedure is rate distortion optimal (for mean squared error), and is conceptually similar to postcompression rate distortion optimization used for coding codeblocks within JPEG 2000. The disadvantage of this approach is its high computational complexity. The second bit allocation algorithm, here called the mixed model (MM) approach, mathematically models each slice's rate distortion curve using two distinct regions to get more accurate modeling at low bit rates. These two bit allocation algorithms are applied to a 3-D Meteorological data set. Test results show that the MM approach gives distortion results that are nearly identical to the optimal approach, while significantly reducing computational complexity.

**Index Terms**—Bit-rate allocation, data compression, image coding, JPEG2000, multidimensional coding, rate distortion theory.

## I. INTRODUCTION

**I**N THIS paper, we solve the problem of optimal bit-rate allocation for three-dimensional (3-D) JPEG 2000 compression. JPEG 2000 Part 2 [1] has the capability to compress 3-D data by treating data as separate two-dimensional (2-D) slices. The separate slices could be taken directly from the data or from the data that has undergone a decorrelation transformation in one direction. The compression setup that involves doing pre- and postprocessing with the Karhunen–Loève Transform before and after being processed by the JPEG 2000 coder in the other two directions is shown in the Fig. 1. The question that arises is how to optimally allocate bits to the separate slices since

Manuscript received January 24, 2004; revised August 8, 2005. This work was supported in part by the Texas Instruments Foundation and in part by the National Imagery and Mapping Agency under Contract NMA401-02-1-2017. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Trac D. Tran.

O. M. Kosheleva, B. E. Usevitch, and S. D. Cabrera are with the Department of Electrical and Computer Engineering, University of Texas at El Paso, El Paso, TX 79968-0523 USA (e-mail: olga@ece.utep.edu; cabrera@ece.utep.edu; usevitch@ece.utep.edu).

E. Vidal, Jr., is with the United States Army Research Laboratory, White Sands Missile Range, NM 88002-5501 USA (e-mail: develone@sbcglobal.net).

Digital Object Identifier 10.1109/TIP.2006.875216

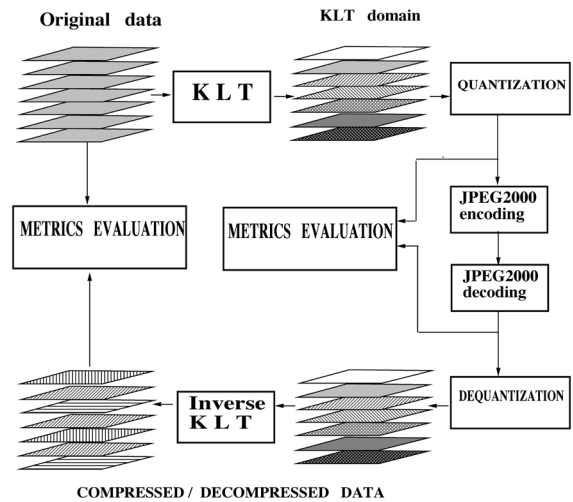


Fig. 1. Diagram illustrating 3-D compression setup.

JPEG 2000 does not require that any specific method be used. In more precise terms, we solve the problem: Given a desired average bit rate, assign bit rates to the individual slices so that the mean-squared error (MSE) distortion metric is minimized. It can be shown that the solution of this (direct) problem also gives a solution to the inverse problem: Given the upper bound on the MSE distortion metric, find the optimal bit-rate allocation for the slices such that the average bit rate is minimized. Prior to this paper, the most generally accepted approach to interslice bit allocation applied a Lagrange multiplier assuming a high bit-rate quantization approximation for each slice [7], [9]. This traditional approach in effect leads to a bit-rate allocation based on the logarithms of variances of the corresponding slices [6], [12], [15]. We propose two new methods. The first approach is here called the rate distortion optimal (RDO) method and is based on postcompression rate-distortion (PCRD) concept. In JPEG 2000, PCRD optimization is used for the problem of selecting the optimal truncation points for the bit streams of the code blocks [14]. The RDO approach is valuable because it gives an optimal MSE lower bound for the bit allocation problem. Since this approach makes use of experimentally obtained rate distortion curves, its main disadvantage is implementation complexity.

The second approach is called the mixed model (MM) approach and consists of extending the traditional high-resolution model with a region that is accurate for low bit rates. The low bit-rate part of the model follows the results of Mallat and Falzon [5], [10] for a general transform coder. Advantages of

the MM approach are that it can give performance nearly identical to the RDO approach with a much lower implementation complexity.

The proposed bit allocation methods are tested by applying them to Meteorological (Met) data. The specific data used was generated by the battlescale forecast model (BFM) [8], which is the analytical model developed by the U.S. Army Research Laboratory. Test results show that the MM approach gives distortion results that are nearly identical to the RDO approach on this data, while significantly reducing complexity of implementation.

The remainder of the paper proceeds as follows. In Section II, we review the JPEG 2000 compression method. Section III describes the KLT preprocessing and contains general descriptions of the relevant optimization problems and Lagrange multiplier techniques. Section IV describes how rate distortion curves are obtained for each slice (using the RDO and MM approaches). Sections V and VI describe the algorithms used in solving for the optimal bit-rate allocations for both approaches.

Experimental results on Met data and conclusions are provided in Section VII.

## II. JPEG 2000: EMBEDDING AND SCALABILITY, PCRD

The JPEG 2000 Part 1 baseline or simply JPEG 2000 [14] brings a new paradigm to image compression standards. It provides among other advantages superior low bit-rate performance, bit-rate scalability and progressive transmission by quality or resolution. Quality scalability is achieved by dividing the wavelet transformed image into codeblocks  $B_i$ . Each codeblock is encoded into embedded representation yielding distortion  $D_i(n_i)$  as a function of bit-rate  $R_i(n_i)$  for each given truncation point  $n_i$ . After having encoded all codeblocks, a postprocessing operation determines where each code-block's embedded stream should be truncated in order to achieve a predefined bit-rate or distortion bound for the whole image. This bitstream rescheduling module is referred to as the Tier 2. It establishes a multilayered representation of the final bitstream, guaranteeing an optimal performance at several bit rates or resolutions. The Tier 2 component optimizes the truncation process, and tries to reach the desired bit-rate while minimizing the introduced distortion, utilizing Lagrangian rate allocation principles. The following procedure is known as PCRD optimization [13], [14] and the basic principles behind it are extensively discussed in [4].

Assuming that the overall distortion metric is additive, i.e.,

$$D = \sum_{i=1}^N D_i(n_i) \quad (1)$$

it is desired to find the optimal selection of bit stream truncation points  $n_i^\lambda$  such that the overall distortion metric is minimized subject to a constraint  $R = \sum_i R_i(n_i) \leq R^{\max}$ . To solve the problem, the Lagrange multiplier method is used. It leads to the unconstrained optimization problem

$$Q = (D(\lambda) + \lambda \cdot R(\lambda)) \rightarrow \min. \quad (2)$$

The resulting objective function  $Q$  depends on  $N$  variables  $n_i$ , but can be represented as a sum of  $N$  terms  $Q_i = D_i(n_i) + \lambda \cdot R_i(n_i)$ . Therefore, to minimize the sum, we must find, for each code-block  $i$ , a truncation point  $n_i^\lambda$  that minimizes the corresponding term  $Q_i$ . This approach follows from the general result proven in [4]. The determination of the  $N$  optimal truncation points for any given  $\lambda$  is performed efficiently based on the experimental information about rate distortion dependence collected during generation of each code-block's embedded bit stream. Only convex hull rate distortion points (i.e., the largest set of truncation points for which the corresponding distortion-rate slopes are strictly decreasing) are used in the optimization. Basically, the algorithm finds the truncation points  $n_i^\lambda$ , where each rate distortion slope  $S_i(n_i^\lambda) = \Delta D_i(n_i^\lambda) / \Delta R_i(n_i^\lambda)$  is closest to the fixed  $\lambda$ .

Since this algorithm has to be repeated for several values of  $\lambda$ , the distortion-rate slopes are precalculated, and only the set of acceptable truncation points  $N_i$  (defining the convex hull) is stored together with the corresponding rates and distortion-rate slopes.

The PCRD optimization approach was the main motivation for the proposed RDO method described in this paper for finding the optimal bit-rate allocation for the slices of data or transform coefficients in JPEG 2000 Part 2 compression.

## III. INTERSLICE BIT ALLOCATION PROBLEM

### A. KLT Preprocessing

In the simplest (2-D) approach, the 3-D volume of data is considered as a set of independent slices, which are successively compressed. The disadvantage of this approach is that the compression uses the intra-slice redundancies to improve performance, while the interslice redundancies (in the third dimension) are not used. In order to take advantage of these interslice correlations as well, the JPEG 2000 Part 2 allows for a cross-component (slice) transform. In this paper, we use KLT preprocessing to decorrelate data.

To establish notation we consider the vertical direction vectors at each  $(x, y)$ ,  $\vec{I}(x, y) = (I(x, y, 1), \dots, I(x, y, N))$ . In terms of vector space representation, these values are the coefficients in an  $N$ -dimensional space using the standard basis. In the KLT transform representation we are selecting a different orthonormal basis  $\vec{e}_1, \dots, \vec{e}_N$  and finding the following representation after subtracting the mean vector:

$$\vec{I}(x, y) - \vec{I}_0 = a_1(x, y) \cdot \vec{e}_1 + \dots + a_N(x, y) \cdot \vec{e}_N \quad (3)$$

where vectors  $\vec{e}_1, \dots, \vec{e}_N$  are the eigenvectors of the covariance matrix [11] of the ensemble of  $N$ -vectors with mean  $\vec{I}_0$ .

### B. Formulation of Optimization Problem

The problem of optimal bit-rate allocation becomes most important when we apply KLT preprocessing in JPEG 2000 Part 2. The diagram in Fig. 2 shows that the KLT transformed data will result in widely varying bit rates as compared to the non-transformed data. Because the KLT concentrates most of the energy into relatively few slices, (shown in the diagram as more bright), those slices will be allocated significantly more bits than the others with lesser energy. In the *direct* optimization problem,

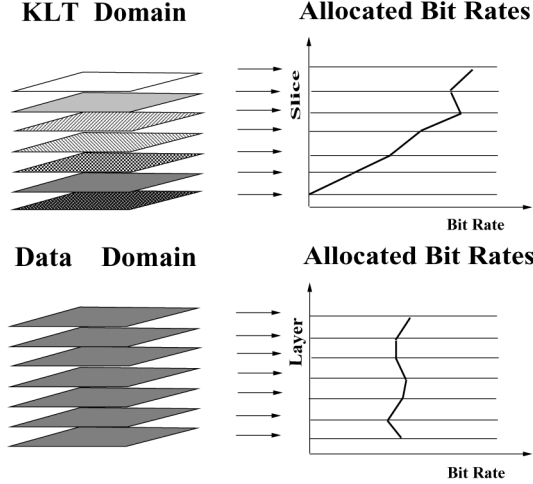


Fig. 2. Diagram illustrating the expected nature of optimal bit-rate allocation schemes for the KLT and data domains.

we are given a target average bit rate of the collection of  $N$  slices as  $R = 1/N \cdot \sum_{z=1}^N R_z$ , and we want to find, among all possible feasible bit-rate combinations  $\{R_1, \dots, R_N\}$ , a specific choice (allocation) for which the MSE distortion attains the smallest possible value.

In the *inverse* optimization problem, we are given the required value of MSE, and we want to find, among all bit-rate allocations that guarantee this compression quality, the allocation  $\{R_1, \dots, R_N\}$  for which the average bit rate is the smallest. From the mathematical point of view, the solution of the inverse problem is easy to deduce from the direct problem, and, therefore, we will focus on the direct optimization problem.

### C. Solving the MSE-Related Problems

Defining  $\text{MSE}_z(R_z)$  as the rate distortion curve (RDC) for slices, solving the direct constrained optimization problem using the Lagrange multiplier method requires the solution of an unconstrained optimization problem with objective function (a function of  $N$  bit rates)

$$Q \triangleq \sum_{z=1}^N \text{MSE}_z(R_z) + \lambda \cdot \sum_{z=1}^N R_z = \sum_{z=1}^N (\text{MSE}_z(R_z) + \lambda \cdot R_z). \quad (4)$$

As shown in [4], minimization of  $Q$  can be accomplished through individual minimization of each entry  $(\text{MSE}_z(R_z) + \lambda \cdot R_z)$  of the sum (4). Differentiating and setting to zero, each desired value  $R_z^o$  can be obtained as a solution of  $\text{MSE}'_z(R_z) + \lambda = 0$ . Thus, the optimal values  $R_z^o$  are such points where all the slopes of the rate distortion curves are equal  $-\text{MSE}'_1(R_1^o) = \dots = -\text{MSE}'_N(R_N^o) = \lambda$ . In order to deal with nonnegative slopes and  $\lambda$ , the slope is redefined to be the negation of the standard slope.

Since MSE is preserved under the (orthonormal) KLT transform, we can use the same approach in the KLT or in the data domain. Thus, we need to have RDCs or reasonable approxi-

mations to them for each slice; once they are determined it is straightforward to find the optimal values  $R_z^o$ .

## IV. RATE DISTORTION CURVES: EXPERIMENTAL AND PROPOSED MM APPROXIMATION

The first proposed approach to generating RDCs makes use of experimentally acquired pairs  $(R_z, \text{MSE}_z(R_z))$ . Then, using the RDO approach analogous to PCRD, we generate a true optimal bit-rate allocation. To generate experimental data, we select for each slice  $z$  several different increasing bit rates  $R_z(1), \dots, R_z(T_z)$  in the acceptable range (from  $R_{MIN_z}$  that strictly depends on the overhead to  $R_{MAX_z}$  which is the highest rate achievable for a given slice) and obtain the corresponding  $\text{MSE}_z(R_z)$  at each point using a JPEG 2000 coder and decoder. Once the RDCs are determined, the optimal bit rates  $R_z^o$  are determined by finding on all the curves points of the same slope such that the corresponding bit-rate average is the desired target rate.

The use of experimentally composed RDCs helps us to find the true MSE optimal bit-rate allocation; however, this is a computationally expensive process that should be avoided if possible. With this in mind we propose a second method for producing rate distortion curves called the MM approach. The MM method is based in part on the traditional high-resolution quantizer model given in [6]. According to this model, distortion depends on rate as  $\text{MSE}(R) = \alpha \cdot \sigma^2 \cdot 2^{-2R}$ , where  $\alpha$  is a constant depending on the probability distribution of the data, and  $\sigma^2$  is the data variance.

Previous research by Mallat and Falzon [5], [10] shows that the high-resolution model is inaccurate at low bit rates for transform coding. They describe (and theoretically justify) a new model which basically adds a new rate distortion relationship  $A \cdot 1/R^\alpha$ , that is valid at low bit rates. Combining this with the high-resolution model gives the following Mallat–Falzon (M–F) model:

$$\text{MSE}(R) = \begin{cases} A \cdot \frac{1}{R^\alpha}, & \text{if } R \leq \tilde{R} \\ B \cdot 2^{-2R}, & \text{if } R > \tilde{R} \end{cases} \quad (5)$$

which gives better agreement with actual RDCs than the high-resolution model.

The M–F approach has one technical difficulty when used to model RDCs of KLT data slices. Due to the optimal energy compaction, the first few slices carry practically all the information contained in the data set, and the last slices contain practically no significant information at all. It is, therefore, reasonable to expect that in the optimal bit-rate allocation, we will ignore the last noninformative slices, i.e. we will set the corresponding bit rates to 0. However, substituting  $R = 0$  into M–F model results in an infinite MSE value. To overcome this difficulty, we propose the following modified M–F model called the MM

$$\text{MSE}(R) = \begin{cases} A \cdot \frac{1}{(R+R_0)^\alpha}, & \text{if } R \leq \tilde{R} \\ B \cdot 2^{-2R}, & \text{if } R > \tilde{R} \end{cases} \quad (6)$$

where the new parameter satisfies  $R_0 > 0$ . This term is chosen such that at zero rate  $R = 0$  the MSE corresponds to the average

power of the data, thus agreeing with standard rate-distortion results [2]. The determination of  $R_0$  is further explained below.

## V. BIT-RATE ALLOCATION BY USING FULLY EXPERIMENTAL MSE DATA

To find the optimal bit-rate allocation using the RDO approach, we first have to select a set of feasible points that define the vertices of the convex hull for each rate distortion curve. For the convex hull the slope varies monotonically with respect to the bit rate.

As discussed in Section III-C, the optimal bit-rate allocation  $R_1^o, \dots, R_N^o$  (for a given average target bit rate  $R_{\text{targ}}$ ) will be such that the corresponding slopes on the rate distortion curves will have the same value  $\lambda$ . Therefore, we can solve the problem by searching through different  $\lambda$  and finding the one that leads to the bit rates whose average equals the desired value  $R_{\text{targ}}$ .

Since we only have finitely many values of the bit rate, there are only finitely many possible values of the slopes. Thus, for a given  $\lambda$ , generally, we will not be able to find a bit rate for which the slope is exactly equal to  $\lambda$ . The best that can be done is to assign the optimal bit rate  $R_i^o$  for slice  $i$  as that bit rate corresponding to the minimum slope in  $\lambda_i^{\min}$  that satisfies  $\lambda_i^{\min} \geq \lambda$  [14]. We use the bisection algorithm (see, e.g., [3]) to find the optimal bit rate  $R_i^o$  for each slice  $i$ .

To complete the description of this algorithm, we must describe how to select the optimal value of the Lagrange multiplier  $\lambda^o$ . We must select it in such a way that the corresponding average bit rate  $R(\lambda^o) = 1/N \cdot \sum_{z=1}^N R_z(\lambda^o)$  is equal to the given average bit rate  $R_{\text{targ}}$ .

As we have mentioned earlier,  $\lambda^o$  is the value of the slope on the  $z$ th rate distortion curve corresponding to the optimal value  $R_z^o$ . The slope decreases as the bit rate  $R_z$  increases; thus, when  $\lambda$  decreases, the corresponding values  $R_z(\lambda)$  increase and, therefore, the average bit rate  $R(\lambda)$  increases, as well. So, the dependence  $R(\lambda)$  is monotonically decreasing and again, we use bisection to find the value  $\lambda^o$  for which  $R(\lambda^o) = R_{\text{targ}}$ . We start with  $[\underline{\lambda}, \bar{\lambda}]$ , where  $\underline{\lambda} = 0$  and  $\bar{\lambda} = \min_z S_z(1)$ , where  $S_z(1)$  is the first RDC slope for the largest MSE (this interval is guaranteed to contain the optimal value of  $\lambda$ ).

At each stage of the bisection algorithm, the interval  $[\underline{\lambda}, \bar{\lambda}]$  is halved as follows:

- first, we find a midpoint

$$\lambda_m \stackrel{\text{def}}{=} \frac{\underline{\lambda} + \bar{\lambda}}{2} \quad (7)$$

- for this  $\lambda_m$ , we find the bit rates for each slice  $R_1(\lambda_m), \dots, R_N(\lambda_m)$  (also using bisection to match slopes on all slices), and then compute the average bit rate  $R(\lambda_m)$ ;
- if  $R(\lambda_m) > R_{\text{targ}}$ , this means that the desired value of  $\lambda$  is larger than  $\lambda_m$ ; thus, we can take a half-interval  $[\lambda_m, \bar{\lambda}]$  as the new interval that is guaranteed to contain  $\lambda$ ;
- if  $R(\lambda_m) < R_{\text{targ}}$ , this means that the desired value of  $\lambda$  is smaller than  $\lambda_m$ ; thus, we can take a half-interval  $[\underline{\lambda}, \lambda_m]$  as the new interval that is guaranteed to contain  $\lambda$ .

We stop iterations when for selected accuracy  $\epsilon$ ,  $|R(\lambda_m) - R_{\text{targ}}| \leq \epsilon$ .

## VI. BIT-RATE ALLOCATION USING THE MM

### A. How to Determine Parameters of the MM

The use of the MM requires for each slice  $z$  the determination of the model parameters  $A_z, R_{0z}, \alpha_z, B_z$ , and the crossover point  $\tilde{R}_z$  given in (6). A straightforward way to compute the model parameters is to pick several bit rates, a high bit rate,  $R_H$ , and two low bit rates  $R_{L1}, R_{L2}$ , and experimentally determine the MSE at these bit rates.

The parameter  $B$  is then computed as  $B_z = \text{MSE}_z(R_H) \cdot 2^{2R_H}$ . To determine  $A_z$  and  $\alpha_z$ , we need to consider the low bit-rate part of the new model. If we choose low bit rates  $R_{L1}, R_{L2}$  such that they are significantly greater than  $R_{0z}$ , then we can use the simplified approximation  $\text{MSE}_z(R) = A_z \cdot 1/R^{\alpha_z}$  for the curve.

Substituting  $R_{L1}, R_{L2}$  into this equation and solving gives the following closed form expressions for  $A_z$  and  $\alpha_z$ :

$$\alpha_z = \frac{\log_2(\text{MSE}_z(R_{L1})) - \log_2(\text{MSE}_z(R_{L2}))}{\log_2(R_{L2}) - \log_2(R_{L1})} \quad (8)$$

$$\log_2(A_z) = \frac{\log_2(\text{MSE}_z(R_{L1})) \cdot \log_2(R_{L2})}{\log_2(R_{L2}) - \log_2(R_{L1})} - \frac{\log_2(\text{MSE}_z(R_{L2})) \cdot \log_2(R_{L1})}{\log_2(R_{L2}) - \log_2(R_{L1})}. \quad (9)$$

To find  $R_{0z}$ , we use the fact that the  $\text{MSE}_z$  equals the average power  $E[I^2]$  for slice  $I$  at  $R_{0z} = 0$  (or  $\sigma^2$  for the zero-mean data), and, thus,  $R_{0z} = (A_z/E[I^2])^{1/\alpha_z}$ .

The crossover point  $\tilde{R}_z$  determines when to switch from the low bit rate to the high bit-rate model. Note that to properly model a rate distortion curve, the combined MM curve must satisfy the following:

- it should agree with the traditional model for high bit rates and with the exponential model for low bit rates;
- it must be convex with slopes increasing as the rate decreases.

The only possible way to satisfy these two criteria is to select the crossover point  $\tilde{R}_z$  as the rate corresponding to the intersection point between the two curves (see Fig. 4 and lower curves of Fig. 3). In certain pathological cases, such as low variance slices, the curves do not intersect. In these cases one of the two curves must be chosen and used for all bit rates. For low variance slices, the low bit-rate and high bit-rate models agree most closely at high bit rates, resulting in the low bit-rate model being the best overall choice for all bit rates.

### B. Analytical Solution to the Unconstrained Optimization Problem

Following the Lagrange multiplier method, we arrive at an unconstrained optimization problem (4). For a given  $\lambda$ , the optimal values  $R_1^o, \dots, R_N^o$  are attained when the derivatives of the objective function (4) with respect to all the variables  $R_1, \dots, R_N$  are equal to 0.

For each variable  $R_z$ , only two terms in the sum (4) depend on this variable:  $\text{MSE}_z(R_z)$  and  $\lambda \cdot R_z$ . For convenience, let us denote the dependence of  $-\text{MSE}_z$  on  $R_z$  as  $f_z(R_z)$ . In this

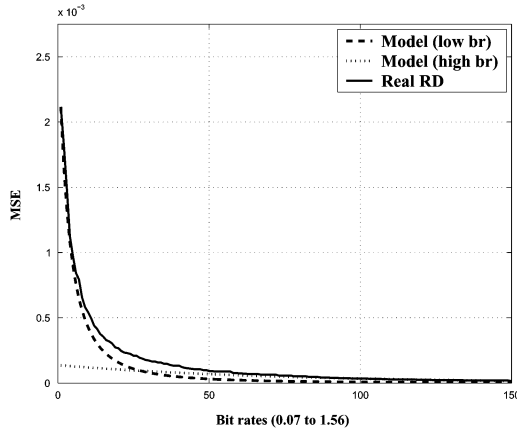


Fig. 3. Comparison of experimental and MM component rate distortion curves for MSE of U variable.

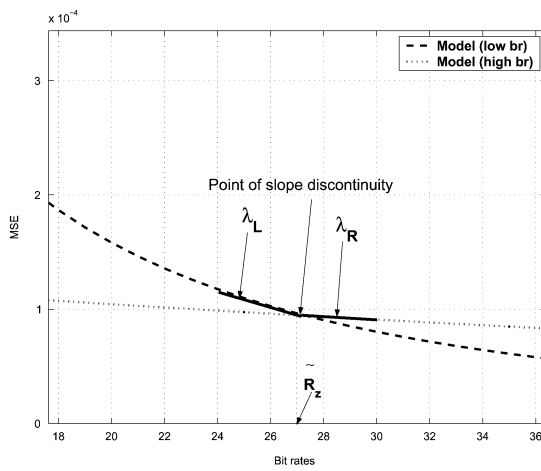


Fig. 4. Plot illustrating the slope discontinuity for MM.

notation, the optimal bit rate  $R_z^o$  satisfies  $f'_z(R_z^o) = \lambda$ . Differentiating  $f_z(R_z)$  in (6) with respect to  $R_z$ , we get

$$f'_z(R) = \begin{cases} A_z \cdot \alpha_z \cdot \frac{1}{(R+R_{0z})^{\alpha_z+1}}, & \text{if } R \leq \tilde{R}_z \\ 2B_z \cdot \ln 2 \cdot 2^{-2 \cdot R_z}, & \text{if } R > \tilde{R}_z. \end{cases} \quad (10)$$

If  $R_z \leq \tilde{R}_z$ , then  $f'_z(R_z) = \lambda$  leads to

$$R_z = \left( \frac{A_z \cdot \alpha_z}{\lambda} \right)^{1/(\alpha_z+1)} - R_{0z}. \quad (11)$$

If  $R_z > \tilde{R}_z$ , it leads to

$$R_z = \frac{1}{2} \cdot \log_2 \left( \frac{2 \cdot \ln(2) \cdot B_z}{\lambda} \right). \quad (12)$$

Since the MM leads in general to a slope discontinuity in the RD curve, one special case for the slope/rate calculation needs to be explained. As seen in Fig. 4, the RD curve slope discontinuity leads to a range of slopes  $\lambda \in (\lambda_R, \lambda_L)$  that have no achievable bit rates associated with them. As mentioned above, the best rate for such  $\lambda$ s is that rate associated with the minimal  $\lambda$  satisfying  $\lambda^{\min} \geq \lambda$ . As can be seen from Fig. 4, this corresponds to slope  $\lambda_L$ , and, thus, the optimal rate is  $\tilde{R}_z$ .

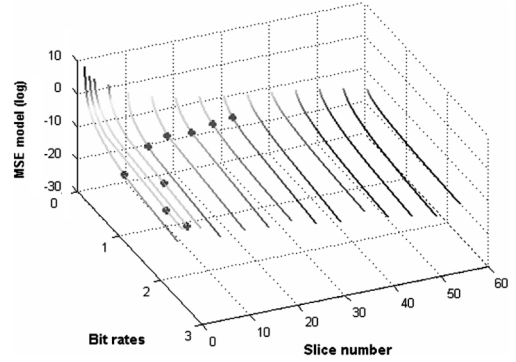


Fig. 5. Optimal bit-rate allocation for U variable.

### C. Algorithm for Optimal Bit-Rate Allocation

Once we have searched and found the optimal value of the Lagrange multiplier  $\lambda^o$ , we can determine the optimal values of the bit rates by using the above explicit formulas. Since possible values of the bit rate can only be between  $R_{\min_z}$  and  $R_{\max_z}$  (see Section IV), once we get the values  $R_z^o$  from the above formulas, we replace the values  $R_z^o < R_{\min_z}$  with 0, and we replace the values  $R_z^o > R_{\max_z}$  with  $R_{\max_z}$ .

As we can see from the MM formulas (11) and (12), when  $\lambda$  is small, we get large bit rates  $R_z$  (hence, large average bit rate  $R$  and small MSE). When  $\lambda$  increases, the average bit rates decrease and the MSE increases. Again we solve the search problem using the *bisection* method.

A typical bit-rate allocation for the Met variable U is shown on Fig. 5.

## VII. RESULTS AND CONCLUSION

A model being used today at the U.S. Army Research Laboratory to generate Met data is the Battlescale Forecast Model (BFM) [8].

The BFM uses physics models and actual measured data as boundary conditions to produce floating point data values on a user specified 3-D spatial grid. The BFM data set available for use in this study consists of a cube of data for each of six physical variables. For a specific variable Met we will call the cube  $Met(Z, X, Y)$  which is of dimensions  $64 \times 129 \times 129$ . The first dimension is the vertical height  $Z$ , and the other two are  $X$  and  $Y$  for the two horizontal spatial variables.

The six available Met variables are: potential temperature  $T$ , pressure  $P$ , water vapor mixing ratio  $WV$  and the  $U$ ,  $V$ , and  $W$  components of the wind speed vector. Each slice of dimension  $129 \times 129$  is available on a uniform 2-D grid, but the specific spatial altitude or height  $z$  is nonuniformly spaced. The original floating point BFM data is first converted to fixed point (16 bits) before any compression is done.

All the results are shown in percentages of the total amplitude range of the specific data cube (see Table I describing the ranges of the Met data). The experiments were performed for several approaches: 1) traditional approach, based on the logarithms of variances (LogVar), 2) RDO approach, providing optimal bit-rate allocation, 3) MM approach, and 4) uniform bit-rate approach. The uniform approach is included for completeness purposes and to further justify the need to optimize bit allocation. The log-variance approach is used without the standard constraint

TABLE I  
SIX VARIABLES USED TO TEST BIT ALLOCATION

Component	Max	Min	Range	Units
U	32.9	-18.5	51.4	m/sec
V	25.9	-18.5	44.4	m/sec
W	0.37	-0.45	0.82	m/sec
T	333.26	270.91	62.35	deg K
WV	13.81	0.027	13.783	$\frac{\text{gr of water}}{\text{Kg of air}}$
P	1019.1	247	772.1	millibars

TABLE II  
RMSE (%) FOR THE PRESSURE VARIABLE

BR	P LogVar	P UBR	P MM	P RDO
0.3	0.56	0.33	0.002	0.0014
0.5	0.12	0.26	0.0007	0.0007
0.7	0.05	0.21	0.0006	0.00062
0.9	0.04	0.18	0.0006	0.0006
1.1	0.03	0.17	0.0006	0.00059
1.3	0.03	0.15	0.0006	0.00059
1.5	0.03	0.13	0.0006	0.00059

TABLE III  
RMSE (%) FOR THE TEMPERATURE VARIABLE

BR	T LogVar	T UBR	T MM	T RDO
0.3	0.26	0.29	0.08	0.082
0.5	0.14	0.22	0.06	0.054
0.7	0.1	0.19	0.06	0.04
0.9	0.08	0.16	0.05	0.032
1.1	0.06	0.14	0.05	0.026
1.3	0.05	0.12	0.04	0.021
1.5	0.04	0.11	0.04	0.018

which forces bit rates to be integers. Instead, any nonnegative bit rate in acceptable range can be achieved on each slice.

Tables II–VII give results comparing the log-variance approach with a uniform bit-rate approach (UBR), the MM approach, and the RDO approach for all Met variables. All tables show that the MM and RDO approaches achieve better results than the other approaches. The log-variance approach shows poor RMSE results at low bit rates and acceptable performance at high bit rates, thus illustrating the need for the MM approach.

One of the major purposes of the experiments was to show that by using the MM approach, we can achieve results close to the RDO results (which provides the overall best possible bit-rate allocation). Tables II–VII show that, for all six components, the results for MM approach and RDO approach are very similar (as expected, the RDO results are slightly better). Fig. 6 shows the results for the U component comparing the MM with the RDO approach and we can see that the results are very similar. Thus, to find a nearly optimal bit-rate allocation, it is not

TABLE IV  
RMSE (%) FOR THE WATER VAPOR VARIABLE.

BR	WV LogVar	WV UBR	WV MM	WV RDO
0.3	0.49	0.23	0.07	0.071
0.5	0.19	0.16	0.05	0.046
0.7	0.13	0.13	0.03	0.033
0.9	0.08	0.10	0.03	0.025
1.1	0.06	0.08	0.02	0.02
1.3	0.05	0.07	0.02	0.016
1.5	0.04	0.06	0.01	0.014

TABLE V  
RMSE (%) FOR THE U WIND SPEED COMPONENT VARIABLE

BR	U LogVar	U UBR	U MM	U RDO
0.3	0.31	0.09	0.04	0.039
0.5	0.16	0.06	0.03	0.025
0.7	0.09	0.04	0.02	0.018
0.9	0.07	0.04	0.01	0.014
1.1	0.05	0.03	0.01	0.011
1.3	0.03	0.02	0.01	0.009
1.5	0.03	0.02	0.01	0.0072

TABLE VI  
RMSE (%) FOR THE V WIND SPEED COMPONENT VARIABLE

BR	V LogVar	V UBR	V MM	V RDO
0.3	0.51	0.17	0.08	0.061
0.5	0.24	0.11	0.05	0.039
0.7	0.13	0.08	0.04	0.028
0.9	0.09	0.07	0.03	0.021
1.1	0.06	0.05	0.02	0.016
1.3	0.05	0.05	0.02	0.013
1.5	0.03	0.04	0.01	0.011

TABLE VII  
RMSE (%) FOR THE W WIND SPEED COMPONENT VARIABLE

BR	W LogVar	W UBR	W MM	W RDO
0.3	0.22	1.36	0.14	0.14
0.5	0.1	1.09	0.08	0.077
0.7	0.06	0.91	0.05	0.048
0.9	0.04	0.79	0.04	0.033
1.1	0.03	0.7	0.03	0.023
1.3	0.02	0.62	0.02	0.018
1.5	0.02	0.55	0.02	0.014

necessary to go through the full rate distortion computations needed for the RDO approach, since the MM approach gives nearly identical results.

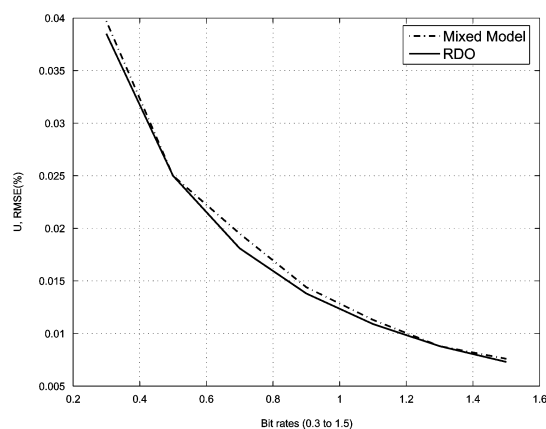


Fig. 6. Detail plots of RMSE(%) of proposed and RDO for the U variable.

The computational savings can be seen by noting that all points for the RDO and MM approach can be generated with only one coding of each slice. However, the RDO approach and MM approach differ in the number of decodings. The MM approach requires only three decodings corresponding to the three RDC points  $R_H$ ,  $R_{L_1}$ ,  $R_{L_2}$  for each slice. In contrast, the RDO approach requires a significant increase in the number of decodings in order to generate RDC information. Our simulations used on the order of 80–100 decodings. As pointed out by one reviewer, near optimal rate-distortion curves can be generated using 20–30 decodings. In either case, there is at least a factor 5 reduction in the computational complexity required. We also note that a JPEG 2000 encoder could be modified to output the image rate distortion curve in one encoding pass. This process would have to be done at the encoder since it requires knowing the rate distortion information from each of the codeblocks, and this information is not stored in the encoded data stream [14]. However, this approach is nontrivial and less practical since it requires access to the JPEG 2000 implementation source code and detailed knowledge of the coding, so that such modifications could be made. Because of these difficulties, this approach was not pursued in this paper.

#### ACKNOWLEDGMENT

The authors would like to thank A. Aguirre for the multidimensional compression software tools. They would also like to thank Dr. T. Henmi of the U. S. Army Research Laboratory for generating and providing the meteorological data used in this study.

#### REFERENCES

- [1] *Information technology—JPEG 2000 Image Coding System—Part 2: Extensions*, ITU-T Recommendation T.801 ISO/IEC 15444-1:2001, 2000.
- [2] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
- [3] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*. New York: McGraw-Hill, 2001.
- [4] H. E. Everett III, "Generalized lagrange multiplier method for solving problems of optimum allocation of resources," *Oper. Res.*, vol. 11, pp. 399–417, 1963.
- [5] F. Falzon and S. Mallat, "Low bit rate image coding over bases," in *Proc. 14th Int. Conf. Pattern Recognition*, Aug. 1998, vol. 2, pp. 692–695.
- [6] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. Boston, MA: Kluwer, 1992.
- [7] R. M. Gray and D. Neuhoff, "Quantization," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 1–63, Oct. 1998.

- [8] T. Henmi and J. R. Dumais, Description of the Battlescale Forecast Model Army Res. Lab. Tech. Rep., May 1998.
- [9] R. Jayant and P. Noll, *Digital Coding of Waveforms*. Englewood Cliffs, NJ: Prentice-Hall, 1984.
- [10] S. Mallat and F. Falzon, "Analysis of low bit rate image transform coding," *IEEE Trans. Signal Process.*, vol. 46, no. 4, pp. 1027–1042, Apr. 1998.
- [11] S. G. Mallat, *A wavelet Tour of Signal Processing*. New York: Academic, 1997.
- [12] M. Pal, C. Brislawn, and S. Brumby, "Feature extraction from hyperspectral images compressed using the JPEG 2000 standard," *Proc. IEEE Southwest Symp. Image Analysis and Interpretation*, pp. 168–172, Apr. 2002.
- [13] D. Taubman, "High performance scalable image compression with EBCOT," *IEEE Trans. Image Process.*, vol. 9, no. 7, pp. 1158–1170, Jul., 2000.
- [14] D. Taubman and M. Marcellin, *JPEG 2000 Image Compression Fundamentals, Standards and Practice*. Boston, MA: Kluwer, 2002.
- [15] P. Topiwala, P. Topiwala, Ed., "Wavelet-based embedded multispectral image compression," in *Wavelet Image and Video Compression*. Norwell, MA: Kluwer, 1998, pp. 261–270.

**Olga M. Kosheleva** (S'89–M'04) received the M.S. degree in mathematics and applied mathematics (summa cum laude) from the Novosibirsk State University, Novosibirsk, Russia, in 1978, and the M.S. degree in computer science and the Ph.D. degree in computer engineering from The University of Texas at El Paso (UTEP) in 1994 and 2003, respectively.

She is currently an Assistant Professor at the UTEP. Her research interests include image and data compression focusing on wavelet image compression and quality-assessment metrics applied in both image processing and educational assessment in the STEM disciplines.

**Bryan E. Usevitch** (S'92–M'93) received the B.S. degree in electrical engineer (magna cum laude) from Brigham Young University, Provo, UT, in 1986, and the M.S. and Ph.D. degrees from the University of Illinois at Urbana-Champaign, Urbana, in 1989 and 1993, respectively.

From 1986 to 1987 and 1993 to 1995, he was a Staff Engineer at TRW designing satellite communication systems. In 1995, he joined the Department of Electrical Engineering, University of Texas at El Paso, where he is currently an Associate Professor. His research interests are in signal processing, focusing on wavelet-based image compression and multicarrier modulation.

**Sergio D. Cabrera** (S'76–M'85) received the B.S. and M.S. degrees in electrical engineering from the Massachusetts Institute of Technology, Cambridge, and the University of Arizona, Tucson, respectively, in 1977 and 1979, respectively, and the Ph.D. degree in electrical engineering from Rice University, Houston, TX, in 1985, specializing in the area of digital signal processing.

From 1979 to 1980, he was with the Digital Image Analysis Laboratory, University of Arizona, College of Engineering. From 1985 to 1987, he was a Senior Engineer at Anadroll/Schlumberger, and from 1987 to 1991, he was an Assistant Professor with the Department of Electrical and Computer Engineering, Pennsylvania State University, University Park, PA. He joined The University of Texas at El Paso in January 1992, where he is currently an Associate Professor of electrical and computer engineering and the current holder of the Texas Instruments Foundation Professorship in Digital Signal Processing. Throughout his career, he has supervised over 30 graduate theses and dissertations and has coauthored more than 100 publications. His research interests are in the areas of signal and image processing, including multirate and time-frequency signal processing; image and volumetric data compression; superresolution applied to SAR imaging; and applications of computer vision and DSP in the areas of manufacturing and biomedical engineering.

**Edward Vidal, Jr.**, received the B.S. degree in electrical engineering from The University of Texas at El Paso (UTEP) in 1977.

He has been with the Army Research Laboratory, White Sands Missile Range, NM, for 35 years. He is currently the Chief Engineer at the Atmospheric Effects Branch, Battlefield Environment Division, Computational and Information Sciences Directorate, Army Research Laboratory. In addition to his formal training, he has extended his capabilities to include computer and software engineering and currently has the responsibility for the development of new methods for data processing and transfer. He initiated and leads a collaborative effort with UTEP in the area of compression of large meteorological data sets for communication over long distances and between distributed processors.