Model Based Optimal Bit Allocation

Nasir M. Rajpoot

Department of Computer Science
University of Warwick
Coventry CV4 7AL
United Kingdom
email: nasir@dcs.warwick.ac.uk

January, 2004

Abstract

Modeling of the operational rate-distortion characteristics of a signal can significantly reduce the computational complexity of an optimal bit allocation algorithm. In this report, such models are studied.

1 Introduction

In recent years, the problem of bit allocation has been a subject of investigation by many researchers working in the areas of audio, image and video coding. The problem is conventionally defined as follows: given a signal $\mathbf{x}$ of size N divided into n non-overlapping subsignals $\mathbf{x}_i$ of size N_i, where $\sum_{i=1}^{n} N_i = N$, find the most efficient distribution of a given bit budget R among a number of available quantizers for the subsignals. The bit allocation algorithm of Shoham & Gersho [1] was to be the first one in a series of algorithms which provide an optimal solution to the problem for an arbitrary set of quantizers. However, this algorithm and its other derivations such as [2] have high computational complexity due to a need to compute the operational rate-distortion (R-D) characteristics for all available quantizers in order to find an optimal solution. The complexity of such algorithms can be greatly reduced if the operational R-D characteristics could be approximated efficiently and with as less error as possible. Two types of bit allocation algorithms that approximate the R-D behaviour of a source can be found in the literature: algorithms which use polynomial functions (such as splines [3]) to fit the operational R-D curve, and algorithms which use analytical models (such as the average distortion-rate function for high bit rates [4, 5]) to approximate the empirical R-D curve. The scope of former class of algorithms is somewhat limited by their requirement to specify the control points on the empirical R-D curve. Moreover, their ability to extrapolate the R-D characteristics beyond the range of control points remains questionable. The latter class of algorithms overcome both of these problems by estimating the parameters of an anlytical model and using it for generate an arbitrary point on the operational rate-distortion curve. An attractive feature of such algorithms is that some of the models used may lend themselves to a convenient mathematical analysis of the problem.

In this report, we investigate further the problem of optimal bit allocation using analytical models for operational R-D characteristics. The average distortion-rate function $D(R) = \sigma^2 2^{-2R}$ has been used in [4] and [5] for optimal bit allocation in video and image coding applications respectively. However, as pointed in [6], there are two problems with using this function: first, it satisfies the so-called high resolution hypothesis and so is useful only for bit rates of over 1 bits/pixel in the context of image coding; second, it assumes that the signal is a realization of a Gaussian source, an assumption which often does not hold for *natural* images or their transformations. This should be motivation enough to look for other models for the empirical R-D behaviour.

Two major contributions of this work are: **(1)** a comparative study of different models that can be used to approximate the operational R-D characteristics of a source, and **(2)** proposal of two new models which appear to provide better fit for the empirical R-D curve in most cases, and a study of some of their properties. The remainder of this report is organised as follows. In the next section, two new models are proposed, and comparative results for empirical R-D curve fitting are provided for these models, the aforementioned average distortion-rate function, and the low bit rate model of Mallat & Falzon [6]. Section 3 describes how these models can be used for optimal bit allocation. Experimental results are presented in Section 4 and the report ends with concluding remarks and some directions for future work.

2 The R-D Model

The average distortion-rate function $D^{(1)}(R) = \sigma^2 2^{-2R}$ can also be written as

$$D^{(1)}(R) = e^{aR+b} \tag{1}$$

where $a, b \in \Re$ and $a < 0$, $b > 0$. The above equation also shows there to be a linear relationship between R and $\log D^{(1)}$. Motivated by the limitations of model in (1), Mallat & Falzon analyzed the coefficients of wavelet and block DCT transforms of images and developed the relationship $D^{(2)}(R) = R^{1-2\gamma}$, where γ is of the order of 1 for most images. In general, their model can be written as follows,

$$\log D^{(2)}(R) = a \log R + b \tag{2}$$

One apparent problem with this model is that the distortion does not approach variance (ie, maximum distortion) as R approaches zero. Noting that the distortion varies much more rapidly at low bit rates than at high bit rates, we propose the following two models,

$$D^{(3)}(R) = e^{e^{\alpha R+\beta}} \tag{3}$$

where $\alpha < 0, \beta > 0$, and

$$D^{(4)}(R) = e^{aR^2+bR+c} \tag{4}$$

where $a, b, c \in \Re$.

Property 1 (Convexity) *The distortion-rate functions given by (3) and (4) are convex $\bigcup$ functions for $D \geq 1, \alpha \geq 0$ and $D \geq 0, a > 0$ respectively.*

The proof of this property follows directly from the fact that $\partial^2 D^{(3)}/\partial R^2 = \alpha^2(D^{(3)}\log D^{(3)}) + \alpha \log D^{(3)}$ is non-negative only when $D \geq 1$ and $\alpha \geq 0$, and $\partial^2 D^{(4)}/\partial R^2 = D^{(4)}[(2aR + b)^2 + 2a]$ is non-negative when $D \geq 0, a > 0$. For the mean square error distortion measure, we can safely say that $D \geq 1$ at low bit rates. Also, slope of the straight line obtained by plotting $\log(\log(D^{(3)}))$ against R is always going to be non-negative, assuming the distortion will always go up when the allocated bit rate is reduced.

Property 2 (Lower Bound on the Average Distortion) *For the distortion-rate function given by (3), the average distortion is bounded from below by the distortion at average bit rate.*

The above property is proved by considering the fact that due to its being convex $\bigcup$, the distortion-rate function of (3) satisfies the Jensen's inequality

$$\mathcal{E}\{D(R)\} \;\geq\; D(\mathcal{E}\{R\}) \tag{5}$$

Figure 1 shows empirical R-D curves for three images and their wavelet transforms, with the results of fitting the curves with all four of the above models. From these results, we observe that the empirical R-D model $D^{(4)}(R)$ represented by (4) outperforms all the other models in almost all the cases. It is also to be noted that the model $D^{(1)}(R)$ given in (1) produces the worst fit for both images in the transform domain (whose operating points in our examples belong to the low bit rate regime). On the other hand, Mallat & Falzon's model performs badly in the high bit rate regime (as can be seen from the R-D curves of the original images). The two proposed models seem to be able to adapt to both low and high bit rate regimes.

2.1 Solving the Optimization Problem

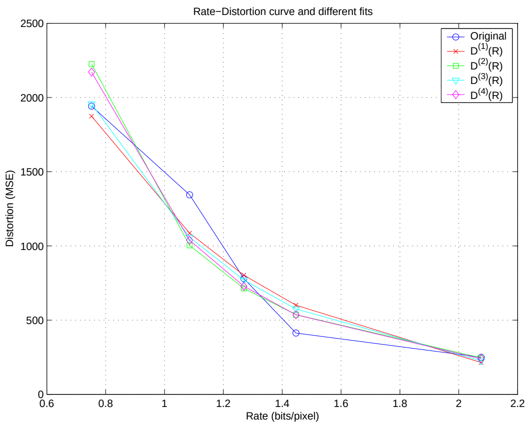
The optimal bit allocation problem can be formulated as an optimization problem which aims to minimize the overall distortion while remaining within an upper limit on the bit budget. It was proved by Everett [7] that an optimal solution to this problem can be found by solving an equivalent Lagrangian optimization problem. For the sake of completeness, the proof is provided below.

Lemma 1 (Unconstrained Problem) *The optimal solution $\mathbf{r}^*$ to the constrained problem*

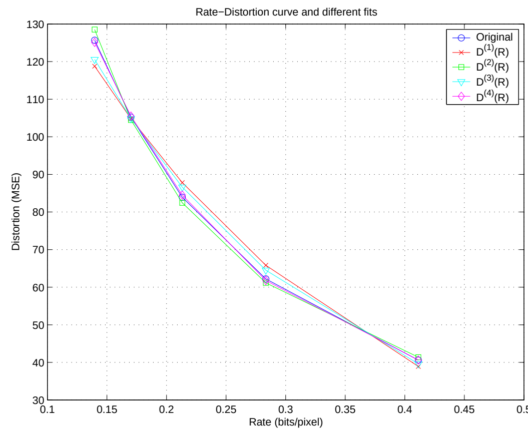
$$min \sum_i d_i(r_i)$$

subject to

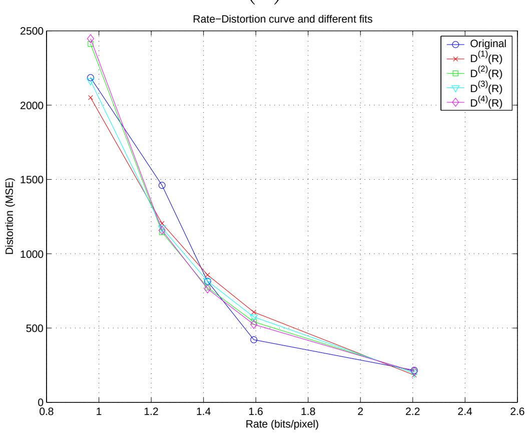
$$\sum_i r_i \leq R$$



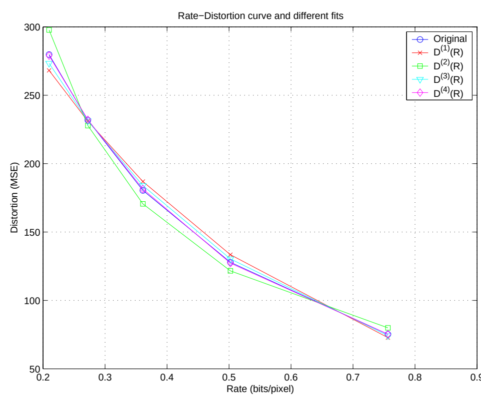
(*a*)



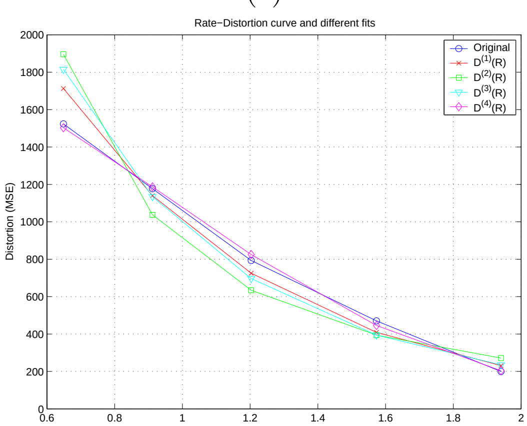
(*b*)



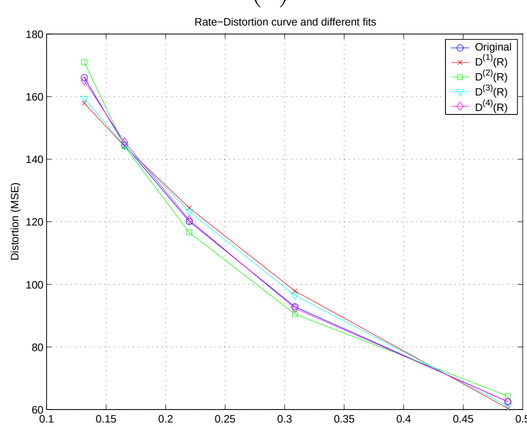
(*c*)



(*d*)



(*e*)



(*f*)

Figure 1: Empirical R-D behaviour and results of fitting
(*a*) 512×512 *Lena* image, and (*b*) 5-level DWT of *Lena*;
(*c*) 512×512 *Barbara* image, and (*d*) 5-level DWT of *Barbara*;
(*e*) 512×512 *Goldhill* image, and (*f*) 5-level DWT of *Goldhill*;

is also optimal solution of the unconstrained problem

$$min \sum_i (d_i(r_i) + \lambda r_i)$$

for the partcilar case where $R = R(\lambda^*) = \sum_{i=1}^{n} r_i^*$.

Proof:
Consider the Lagrangian function given by

$$I(\lambda) = \sum_{i=1}^{n} d_i(r_i) + \lambda \sum_{i=1}^{n} r_i.$$

or

$$I(\lambda) = D(\mathbf{r}) + \lambda \sum_{i=1}^{n} r_i,$$

where $D(\mathbf{r}) = \sum_{i=1}^{n} d_i(r_i)$ denotes the overall distortion. Let λ^* denotes the value of λ corresponding to the optimal solution $\mathbf{r}^*$. Then clearly,

$$I(\lambda^*) \leq I(\lambda), \quad \forall\lambda$$

or

$$D(\mathbf{r}^*) + \lambda^* \sum_{i=1}^{n} r_i^* \ \leq \ D(\mathbf{r}) + \lambda \sum_{i=1}^{n} r_i$$

or

$$D(\mathbf{r}^*) - D(\mathbf{r}) \ \leq \ \lambda(\sum_{i=1}^{n} r_i - \sum_{i=1}^{n} r_i^*)$$

or

$$D(\mathbf{r}^*) - D(\mathbf{r}) \ \leq \ \lambda(\sum_{i=1}^{n} r_i - R)$$

We also know that $\sum_{i=1}^{n} r_i \leq R$. This implies directly that for all values of $\lambda \geq 0$,

$$D(\mathbf{r}^*) \ \leq \ D(\mathbf{r}).$$

Hence the lemma.

3 Bit Allocation

Lemma 2 (Optimality Condition) *Given that the distortion-rate function* $d_i(r_i)$ *is a convex function for each segment* $\mathbf{x}_i$ *of the signal, the optimal solution to the problem*

$$min \sum_i d_i(r_i)$$

subject to

$$\sum_i r_i \leq R$$

is given by $\mathbf{r}^*=(r_1^*, r_2^*, \ldots, r_n^*)$ *which satisfies the following condition*

$$\frac{\partial d_i}{\partial r_i} = \frac{\partial d_j}{\partial r_j} \qquad \forall\ i,\ j \quad i{\neq}j.$$

Proof:
We follow a simplex approach to solve the given optimization problem. Let us denote the sum of distortions $d_i(r_i)$, for all the segments $\mathbf{x}_i$ of the signal, by $D(\mathbf{r}) = D(r_1, r_2, \ldots, r_n)$. The constraint $\sum_i r_i \leq R$ corresponds to an area on and under the surface $\mathcal{S}$ which is given by

$$\sum_i r_i = R \quad \forall\ r_i \geq 0.$$

There are only two possibilites as to where the optimal solution $\mathbf{r}^*$ can lie: either under the surface $\mathcal{S}$, or on its boundary. In the former case (ie, when $\sum_i r_i^* < R$), the optimality condition would be

$$\frac{\partial d_i}{\partial r_i} = 0,\ \forall\ i.$$

But given that the function $d_i(r_i)$ is convex (for all values of i), it can easily be deduced that the optimal distortion $D(\mathbf{r}^*)$ corresponds to a solution $\mathbf{r}^*$ which lies on boundary of the surface $\mathcal{S}$. This implies that the gradient of overall distortion function $D(\mathbf{r})$ should be normal to boundary of the surface $\mathcal{S}$. In other words,

$$\nabla D(\mathbf{r}) = (\frac{\partial d_1}{\partial r_1}, \frac{\partial d_2}{\partial r_2}, \ldots, \frac{\partial d_n}{\partial r_n}) \ \perp\ \partial\mathcal{S}$$

or

$$\nabla D(\mathbf{r}) = (\frac{\partial d_1}{\partial r_1}, \frac{\partial d_2}{\partial r_2}, \ldots, \frac{\partial d_n}{\partial r_n}) = \mu(1, 1, \ldots, 1)$$

for some constant μ. This proves the lemma.

Corollary 1 (Fixed λ) *: From the above lemmas, it is clear that the optimal solution* $\mathbf{r}^*$ *satisfies the condition* $\partial d_i/\partial r_i = -\lambda,\ \forall\ i$*. Therefore, the optimal solution can be found by restricting the value of* λ *to be the same for each segment* $\mathbf{x}_i$ *of the signal.*

From the above result, an algorithm for finding an optimal value of λ can be devised using a gradient descent method, as explained in [1]. The expressions $\frac{\partial D_i}{\partial R_i}$ using all the four models are given by

$$\frac{\partial D_i^{(1)}}{\partial R_i} = aD_i^{(1)},\ \frac{\partial D_i^{(2)}}{\partial R_i} = D_i^{(2)}(b{+}a/R_i),\ \frac{\partial D_i^{(3)}}{\partial R_i} = \alpha D_i^{(3)} \log D_i^{(3)},\ \frac{\partial D_i^{(4)}}{\partial R_i} = D_i^{(4)}(2aR_i{+}b). \tag{6}$$

It may be noted here that expression only for $\frac{\partial D_i^{(1)}}{\partial R_i}$ and $\frac{\partial D_i^{(3)}}{\partial R_i}$ lead to plausible analytical and efficient numerical solutions, as they involve only distortion.

Proposition 1 (Equivalence) *Let* $\mathbf{r}^* = (r_1, r_2, \ldots, r_n)$ *be the optimal budget distribution for a given bit budget* R *bits per pixel, where each* $r_i, \ \forall \ i = 1, 2, \ldots, n$ *corresponds to the ith subsignal having* α_i *and* β_i *as its R-D model parameters. Let* $x_i = \alpha_i r_i + \beta_i, \ \forall \ i = 1, 2, \ldots, n$*. Then* $x_i = A_i R + B_i$ *(where* $A_i, \ B_i \in \Re$ *are constants,* $\forall \ i$*) holds if and only if* $x_1 = x_2 = \ldots = x_n$*.*

Proof:
We know from Theorem 2 that the optimality condition to distribute the given budget R into the n subsignals is

$$\frac{\partial d_i}{\partial r_i} = \frac{\partial d_j}{\partial r_j} \qquad \forall \ i, \ j \quad i \neq j$$

or

$$\alpha_1(d_1 \log d_1) = \alpha_2(d_2 \log d_2) = \ldots = \alpha_n(d_n \log d_n).$$

The above set of equations can be re-written as the following $n-1$ independent equations.

$$\begin{array}{c} x_1 + e^{x_1} = x_2 + e^{x_2} + K_1 \\ x_2 + e^{x_2} = x_3 + e^{x_3} + K_2 \\ \vdots \\ x_{n-1} + e^{x_{n-1}} = x_n + e^{x_n} + K_{n-1} \end{array} \tag{7}$$

where

$$K_i = \log\left(\frac{\alpha_{i+1}}{\alpha_i}\right),$$

and

$$x_i = \alpha_i r_i + \beta_i \tag{8}$$

for all $i = 1, 2, \ldots, n-1$. We also know that

$$r_1 + r_2 + \ldots + r_n = nR$$

Putting the values of r_i from equation (8) into the above equation, we get

$$\sum_{i=1}^{n} a_i x_i = nR + \sum_{i=1}^{n} \frac{\beta i}{\alpha_i} \tag{9}$$

where $a_i = 1/\alpha_i, \ \forall \ i = 1, 2, \ldots, n$. The set of equations 7 can also be written as,

$$\phi(x_i) = \phi(x_{i+1}) + K_i, \qquad \forall \ i = 1, 2, \ldots, n-1 \tag{10}$$

where

$$\phi(x) = x + e^x.$$

Equation (10) implies that given the value of x_i, x_{i+1} can be found by solving this equation. Although it does not have an obvious analytical solution, let us assume that x_{i+1} is directly related to x_i by the following function,

$$x_{i+1} = f_i(x_i) \tag{11}$$

Re-arranging the terms of equation (9), we get the following relation between x_{i+1} and x_i,

$$x_{i+1} = \sum_{\substack{j=1 \\ j \neq i+1}}^{n} b_j x_j + \frac{n}{a_i} R + \frac{1}{a_i} \sum_{j=1}^{n} (\frac{\beta_j}{\alpha_j}) \tag{12}$$

where $b_j = -a_j/a_i$. Now let us suppose that $x_i = A_i R + B_i, \ \forall \ i = 1, 2, \ldots, n$ such that

$$\frac{\partial x_i}{\partial R} = A_i \ \neq 0. \tag{13}$$

Comparing equations (11) and (12), we obtain

$$f_i(x_i) = \sum_{\substack{j=1 \\ j \neq i+1}}^{n} b_j x_j + \frac{n}{a_i} R + \frac{1}{a_i} \sum_{j=1}^{n} (\frac{\beta_j}{\alpha_j})$$

Differentiating above equation with respect to R, we have

$$\frac{\partial f_i}{\partial x_i} \cdot \frac{\partial x_i}{\partial R} = \sum_{\substack{j=1 \\ j \neq i+1}}^{n} b_j \cdot \frac{\partial x_j}{\partial R} + \frac{n}{a_i}$$

or

$$f_i{}'(x_i) = M_i \tag{14}$$

where $M_i = \frac{1}{A_i} \sum_{\substack{j=1 \\ j \neq i+1}}^{n} b_j A_j + \frac{n}{a_i A_i}$ is a constant. It is also clear from (14) that $f_i{}''(x_i) = 0$. From (10) and the implicit function theorem, we obtain

$$f_i'(x_i) = \frac{\phi'(x_i)}{\phi'(f_i(x_i))} = \frac{1 + e^{x_i}}{1 + e^{x_{i+1}}} \tag{15}$$

Differentiating (15) and equating $f_i''(x_i)$ to zero, we obtain

$$e^{-x_{i+1}} + e^{x_{i+1}} = e^{-x_i} + e^{x_i}.$$

The function $(e^{-x} + e^x) = 2cosh(x)$ is an even function and thus the above equation has one of the two possible solutions: either $x_{i+1} = x_i$ or $x_{i+1} = -x_i, \ \forall \ i = 1, 2, \ldots, n-1$. The latter of these solutions can clearly be ruled out. Thus we have

$$x_1 = x_2 = \ldots = x_n$$

Conversely, if $x_{i+1} = x_i, \ \forall \ i = 1, 2, \ldots, n-1$, then $x_i = A_i R + B_i$ holds for some constants A_i and B_i by the definition of x_i (ie, $x_i = \alpha_i r_i + \beta_i, \ \forall \ i = 1, \ldots, n$).

Corollary 2 (Linear Variation) *From the above proposition, it is obvious that the optimal bit budget distributions $r_i, \ \forall \ i = 1, 2, \ldots, n$ vary linearly with the target bit budget R according to the following relation*

$$r_i = \frac{\alpha}{\alpha_i} R + (\frac{\beta - \beta_i}{\alpha_i})$$

Bit Budget (bpp.)	Uniform Allocation		*var*-based Allocation		Optimal Allocation	
	bpp.	PSNR (dB)	bpp.	PSNR (dB)	bpp.	PSNR (dB)
0.5	0.5	37.24	0.68	39.24	0.36	35.25
	0.5	41.50	0.51	41.61	0.26	38.11
	0.5	34.89	0.35	32.73	0.33	32.45
	0.5	31.41	0.37	29.75	0.40	30.10
	0.5	33.13	0.43	32.34	0.32	30.98
	0.5	37.66	0.38	36.05	0.21	32.50
	0.5	28.61	0.47	28.18	0.67	30.32
	0.5	29.22	0.53	29.52	0.70	31.59
	0.5	28.64	0.71	30.48	0.60	29.60
	0.5	33.31	0.53	33.81	0.33	30.88
	0.5	26.04	0.69	27.68	0.81	28.58
	0.5	24.59	0.71	27.00	0.97	27.01
	0.5	38.02	0.22	35.05	0.22	35.05
	0.5	36.96	0.31	35.42	0.23	34.50
	0.5	25.98	0.32	24.35	0.66	27.53
	0.5	26.00	0.79	28.71	0.93	29.89

Table 1: Results of bit allocation for coding sixteen 128×128 blocks of *Barbara*

where

$$\alpha = \frac{n}{\sum_{i=1}^{n} \frac{1}{\alpha_i}}$$

and

$$\beta = \frac{\alpha}{n} \sum_{i=1}^{n} \frac{\beta_i}{\alpha_i}.$$

4 Experimental Results

The optimal bit allocation algorithm outlined in the previous section was tested on block partitions of an image in order to efficiently distribute a given bit budget between the image blocks. The model parameters were determined by quantizing with five different step sizes the wavelet transform coefficients of each image block. For the purposes of experimentation, third model for the empirical R-D curve was used. Results of budget distribution between 16 equal sized partitions of the *Barbara* image are given in Table 1. For comparison, results are also provided for a relatively naive activity based bit allocation algorithm which works by analyzing the variance of each of the blocks, and for a uniform distribution of the bit budget. The bit allocation algorithm assigns more bits to an image block which contains regions of high activity.

5 Conclusions

This report addressed the issue of model based optimal bit allocation. Two new models were proposed for representing the empirical R-D curves. It was demonstrated that while $D^{(4)}$ provides the best fit for operational R-D characteristics, $D^{(3)}$ lends itself to tractable analysis of the bit allocation problem. Future work may include a generalization of the bit allocation algorithm to all models, and an investigation into the relationship of these models with the distribution of transform coefficients for subband image coding.

6 Acknowledgements

The author would like to thank Artur Sowa and François Meyer for their ideas and many motivating discussions. Part of this work was carried out when the author was visiting the department of Mathematics, Yale University, with generous support from Professor Ronald Coifman.

References

[1] Y. Shoham and A. Gersho. Efficient bit allocation for an arbitrary set of quantizers. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 36(9):1445–1453, September 1988.

[2] K. Ramchandran and M. Vetterli. Best wavelet packet bases in a rate-distortion sense. *IEEE Transactions on Image Processing*, 2(2):160–175, April 1993.

[3] L-J. Lin and A. Ortega. Bit-rate control using piecewise approximated rate-distortion characteristics. *IEEE Transactions on Circuits and Systems for Video Technology*, 8(4):446–459, August 1998.

[4] W. Y. Lee and J. B. Ra. Fast algorithm for optimal bit allocation in rate-distortion sense. *IEE Electronic Letters*, 32:1871–1896, October 1996.

[5] J. Li, P-Y. Cheng, and C-C. Kuo. Embedded wavelet packet image coder with fast rate-distortion optimized decomposition. In *SPIE: Visual Communication and Image Processing*, volume 3204, pages 1077–1088, February 1997.

[6] S. Mallat and F. Falzon. Analysis of low bit rate image transform coding. *IEEE Transactions on Signal Processing*, April 1998.

[7] H. Everett. Generalized Lagrange multiplier method for solving problems of optimum allocation of resources. *Operations Research*, 11:399–417, 1963.