International Conference on Intelligent Computing. August 23-26, 2005. HeFei, China

Advanced Real-Time Rate Control for Low Bit Rate Video Communication *

Yoon Kim[1], Inbum Jung[1], Dong-Mahn Seo[1], Jae-Young Pyun[2], and Sang-Hyun Park[3]

[1] Department of Electrical and Computer Engineering, Kangwon National University, Chunchon, 200-701, Korea.
{yooni, ibjung, sarum}@kangwon.ac.kr
[2] Department of Information Communication Engineering, Chosun University, 375 Susuk-dong, Dong-gu, Gwangju, 501-759, Korea.
jypyun@chosun.ac.kr
[3] Department of Multimedia Engineering, Sunchon National University, Sunchon, 540-742, Korea
shark@sunchon.ac.kr

Abstract. In this paper, we propose a novel real-time frame-layer rate control algorithm using sliding window method for low bit rate video coding. The proposed rate control method performs bit allocation at the frame level to minimize the average distortion over an entire sequence as well as variations in distortion between frames. A new frame-layer rate-distortion model is derived, and a non-iterative optimization method is used for low computational complexity. In order to reduce the quality fluctuation, we use a sliding window scheme which does not require the pre-analysis process. Therefore, the proposed algorithm does not produce time delay from encoding, and is suitable for real-time low-complexity video encoder. Experimental results indicate that the proposed control method provides better visual and PSNR performance than the existing TMN8 rate control method.

1 Introduction

As a consequence of the increasing role of video in the rapidly evolving world of multimedia systems, an important evolution in the concept of audiovisual information is taking place. The digital video compression technique plays an important role in development of an audiovisual communication system. A near-term enhancement of H.263 known as H.263+ [1] is suitable for low bit rate visual communications such as video over the Internet. To transmit compressed video efficiently over the Internet, we should consider both the underlying video content and channel conditions, and develop an effective rate control scheme accordingly. Rate control of H.263+ video over the Internet is the main focus of this work.

* This work was supported by the IT Research Center (ITRC) and Kangwon National University.

In block-based video coders such as MPEG and H.26x, the number of bits and distortion for each image block encoding are controlled by the quantization parameter of the block. The objective of the rate control is to select the quantization parameters so that the encoder produces bits at transmission bandwidth and the overall distortion is minimized. Therefore, the rate control not only regulates the output bit-stream to meet certain given conditions, but also enhances the quality of coded video. However, the rate control algorithms are not standardized since they are independent of the decoder structure.

Many rate control schemes have been proposed in [2]-[11]. In general, these schemes can be thought of as having a frame and a macroblock layer. The frame-layer rate control assigns a target number of bits to each video frame and, at a given frame, the block-layer rate control selects the block-quantization parameters to achieve the frame target [3]-[5]. Some frame-layer rate control approaches use simple formulas, but these simple methods generally do not achieve the target number of bits accurately [6]. Other approaches use various rate-distortion strategies to assign a target number of bits to each frame [7]-[8]. However, since they usually use either an iteration method for optimal bit allocation or a pre-analysis method on a group of frames before encoding, they produce time delay or high computational complexity. The frame-layer rate control algorithm in TMN10 [9] is particularly useful for those that use B frames.

In this paper, we propose a real-time frame-layer control method using a sliding window for low bit rate video compression standard, H.263+. In order to achieve accurate frame rate control, a new frame-layer rate-distortion (R-D) model is derived. The proposed R-D model minimizes the average distortion over an entire sequence as well as variations in distortion between frames. Furthermore we use a non-iterative method with a low computational complexity for real-time rate control. The proposed sliding window scheme can reduce the quality fluctuation without the pre-analysis process. It is seen that the proposed rate control algorithm does not produce time delay from encoding.

The rest of this paper is organized as follows. In the next section, the conventional TMN8 [4], [5] rate control algorithm, which was designed for low-delay video communications, is briefly introduced. The proposed frame-layer rate control scheme is presented in Section 3. Section 4 presents and discusses the experimental results. Finally, our conclusions are given in Section 5.

2 Review on Conventional TMN8 Rate Control

In H.263+, the current video frame to be encoded is decomposed into macroblocks of 16×16 pixels per block, and the pixel values for each of the four 8×8 blocks in a macroblock are transformed into a set of coefficients using the DCT. These coefficients are then quantized and encoded with some type of variable-length coding. The number of bits and distortion for a given macroblock depend on the macroblock's quantization parameter used for quantizing the transformed coefficients. For example, in a test model TMN8 for the H.263 standard, the quantization parameter is denoted by QP whose value corresponds to half the quantization step size. The TMN8 rate control uses a frame-layer rate control to select a target number of bits for the current frame and a macroblock-layer rate control to select the values of the quantization step-sizes for the macroblocks. In the following discussions, the following definitions are used:

B : target number of bits for a frame;

R : channel rate in bits per second;

F : frame rate in frames per second;

W : number of bits in the encoder buffer;

M : some maximum value indicating buffer fullness, by default, set R/F;

W_{prev} : previous number of bits in the buffer;

B' : actual number of bits used of encoding the previous frame.

In the frame-layer rate control, a target number of bits for the current frame is determined by

$$B = \frac{R}{F} - \Delta, \tag{1}$$

$$\Delta = \begin{cases} W/F, & W > Z \cdot M, \\ W - Z \cdot M, & \text{otherwise}, \end{cases} \tag{2}$$

$$W = max(W_{prev} + B' - R/F, 0), \tag{3}$$

where $Z = 0.1$ by default. The frame target varies depending on the nature of the video frame, the buffer fullness, and the channel throughput. To achieve low delay, the algorithm tries to maintain the buffer fullness at about 10% of the maximum M. If W is larger than 10% of the maximum M, the frame target B is slightly decreased. Otherwise, B is slightly increased.

The macroblock-layer rate control selects the values of the quantization step-sizes for all the macroblocks in the frame, so that the sum of the bits used in all macroblocks is close to the frame target B in (1). The optimized quantization step size Q_i^* for macroblock i in a frame can be determined by

$$Q_i^* = \sqrt{\frac{AK}{\beta_i - AN_iC} \frac{\sigma_i}{\alpha_i} \sum_{k=i}^{N} \alpha_k \sigma_k}, \tag{4}$$

where

K : model parameter;

A : number of pixels in a macroblock;

N_i : number of macroblocks that remain to be encoded in the frame;

σ_i : standard deviation of the ith macroblock;

α_i : distortion weight of the ith macroblock;

C : overhead rate;

β_i : number of bits left for encoding the frame, where $\beta_1 = B$ at the initialization stage.

3 Proposed Frame-Layer Rate Control

For frame-layer rate control, the sliding window is utilized to analyze the scene characteristics. Then, the optimization problem with bit budget constraint is solved by the R-D model based on Lagrangian method. The previous bit allocation approaches used the jumping window method for the analysis of scene characteristics [12]. However, the jumping window scheme requires a pre-analysis process for analyzing a group of frames before encoding, and thus causes an additional delay. We introduce the sliding window method that can analyze scene characteristics without time delay.

Fig. 1 shows the sliding window method where the window moves one frame at a time to determine the target bit for each frame according to the scene characteristics. P_i is the current sliding window consisting of frames $\{f_{i-N_W+1}, f_{i-N_W+2}, \cdots, f_{i-1}, f_i\}$, where N_W is the number of frames within the sliding window.

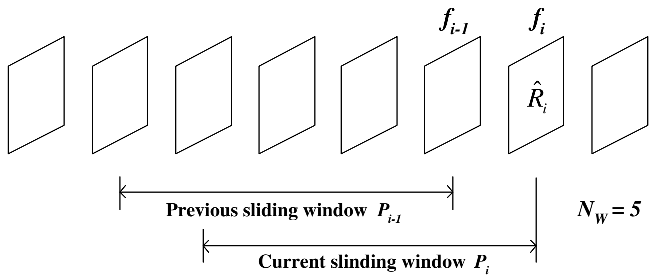


Fig. 1. Concept of the proposed sliding window method.

For the frame-layer rate control, we employ an empirical data-based frame-layer R-D model using the quadratic rate model and the affine distortion model [2] with respect to the average QP in a frame, which is given by

$$\hat{R}(\bar{q}_i) = \left(a \cdot \bar{q}_i^{-1} + b \cdot \bar{q}_i^{-2}\right) \cdot MAD\left(\hat{f}_{ref}, f_{cur}\right), \quad (5)$$

$$\hat{D}(\bar{q}_i) = a^{'} \cdot \bar{q}_i + b^{'}, \quad (6)$$

where a, b, $a^{'}$, and $b^{'}$ are the model coefficients, $\hat{f}_{ref}$ is the reconstructed reference frame at the previous time instant, f_{cur} is the uncompressed image at the current time instant, $MAD(\cdot,\cdot)$ is the mean of absolute difference between two frames, $\bar{q}_i$ is the average QP of all macroblocks in the ith frame, $\hat{R}(\bar{q}_i)$ and $\hat{D}(\bar{q}_i)$ are the rate and distortion models of the ith frame, respectively. The model coefficients can be determined by using the linear regression analysis and the formula consisting of the previous encoding results as follow: From the rate-distortion model in (5) and (6), We first define error functions given by

$$E_R = \sum_{i=1}^{N} \left\{ \frac{R_i \cdot \bar{q}_i}{MAD(\hat{f}_{ref}, f_{cur})} - a - b\bar{q}_i^{-1} \right\}^2, \quad (7)$$

$$E_D = \sum_{i=1}^{N} \{D_i - a\bar{q}_i - b\}^2, \quad (8)$$

where N is the number frames observed in the past, R_i and D_i are the actual bit rate and distortion of the encoded ith frame, respectively. Then, the model coefficients a, b, $a^{'}$, and $b^{'}$ are determined by minimizing the above error functions as follows:

$$a = \frac{\sum_{i=1}^{N} \left(\frac{R_i \cdot \bar{q}_i}{MAD(\hat{f}_{i-1}, f_i)} - b \cdot \bar{q}_i^{-1} \right)}{N},$$

$$b = \frac{N \cdot \left(\sum_{i=1}^{N} \frac{R_i}{MAD(\hat{f}_{i-1}, f_i)} \right)}{N \cdot \left(\sum_{i=1}^{N} \bar{q}_i^{-2} \right) - \left(\sum_{i=1}^{N} \bar{q}_i^{-1} \right)^2} - \frac{\left(\sum_{i=1}^{N} \frac{R_i \cdot \bar{q}_i}{MAD(\hat{f}_{i-1}, f_i)} \right) \left(\sum_{i=1}^{N} \bar{q}_i^{-1} \right)}{N \cdot \left(\sum_{i=1}^{N} \bar{q}_i^{-2} \right) - \left(\sum_{i=1}^{N} \bar{q}_i^{-1} \right)^2},$$

$$a' = \frac{\sum_{i=1}^{N} D_i \cdot \sum_{i=1}^{N} \bar{q}_i - N \cdot \sum_{i=1}^{N} D_i \cdot \bar{q}_i}{\left(\sum_{i=1}^{N} \bar{q}_i \right)^2 - N \cdot \sum_{i=1}^{N} \bar{q}_i^2},$$

$$b' = \frac{\sum_{i=1}^{N} D_i - a' \cdot \sum_{i=1}^{N} \bar{q}_i}{N}.$$

We consider a new formulation of frame-layer rate control based on the R-D model as follows: Determine $\bar{q}_i,\ i = 1, 2, ..., N_W$ to minimize

$$\sum_{i=1}^{N_W} \hat{D}_i(\bar{q}_i) \cdot (\hat{D}_i(\bar{q}_i) - D_{i-1}), \tag{9}$$

subject to

$$\sum_{i=1}^{N_W} R_i \leq BW \cdot T_W, \tag{10}$$

where $\hat{D}_i$ is the estimated distortion of the current frame, D_{i-1} is the actual distortion of the previous frame, BW and T_W are the bandwidth and the time intervals of sliding window, respectively. In (9), we introduce a formulation minimizing the average distortion over an entire sequences as well as variations in distortion between frames.

The optimization task in (9) and (10) can be elegantly solved using Lagrangian optimization where a distortion term is weighted against a rate term. The Lagrangian formulation of the minimization problem is given by

$$J_i(\bar{q}_i) = \hat{D}_i(\bar{q}_i) \cdot (\hat{D}_i(\bar{q}_i) - D_{i-1}) + \lambda_i \cdot max(\hat{B}_i^{res}, 0), \tag{11}$$

$$\hat{B}_i^{res} = \sum_{j=1}^{i-1} R_j + \hat{R}_i(\bar{q}_i) - \sum_{j=1}^{i} \frac{MAD_j}{Ave\_MAD} \cdot \frac{BW \cdot T_W}{N_W}, \tag{12}$$

where $J_i(\bar{q}_i)$ and λ_i is the cost function and the Lagrange multiplier for the ith frame, R_j is the used bit-rate for the jth frame, MAD_j is the MAD between $(j-1)$th and jth frames in the current window, and Ave_MAD is the average of MADs in the current window. Note that $\hat{B}_i^{res}$ denotes the estimated bit based on the R-D model. It was shown in [13] that $J_i(\bar{q}_i)$ is a convex function generally. Thus, we can get its optimal solution by using the gradient method as described in (13).

$$\bar{q}_i^* = arg \min_{\bar{q}_i} J_i(\bar{q}_i). \tag{13}$$

Note that what we finally need is not $\bar{q}_i^*$, but $\hat{R}_i(\bar{q}_i^*)$ which is the target bit budget for the ith frame.

The proposed frame-layer rate control algorithm consists of two steps. The first step is to find the optimal bit-rates with the current Lagrange multiplier, and the second step is to adjust the Lagrange multiplier based on residual bit-rates. The properties of the Lagrange multiplier method are very appealing in terms of computation. Finding the best quantizer for a given λ is easy and can be done independently for each coding unit. In order to achieve the optimal solution at the required rate, an optimal λ must be found. Several approaches including the bisection search algorithm [14] are proposed to find a correct λ. However, the number of iterations required in searching for λ can be kept low as long as we do not seek to have an exact match of the budget rate. Moreover, since we may be performing allocations on successive frames having similar characteristics in video coding, it is possible to adjust λ for a frame using the value achieved for the previous frame. Thus, we employ the adaptive adjustment rule [3] given by

$$\lambda_{i+1} = \lambda_i + \Delta\lambda, \quad \Delta\lambda = \frac{B_i}{B_{target,i}} - 1, \tag{14}$$

where λ_i is the Lagrange multiplier for the ith frame and

$$B_i = \sum_{j=1}^{i} R_j, \quad B_{target,i} = \sum_{j=1}^{i} \frac{MAD_j}{Ave\_MAD} \cdot \frac{BW \cdot T_W}{N_W}. \tag{15}$$

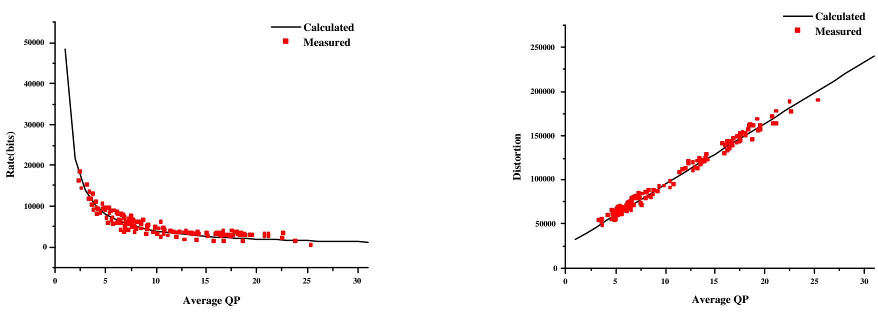
Therefore, the proposed rate control algorithm does not produce encoding time delay. However, a negligible performance loss due to its intrinsic sub-optimality is inevitable in our design.

Once the bit rate is allocated to the frame using the aforementioned frame-layer rate control, the TMN8 macroblock layer rate control algorithm allocates the bit budget to each macroblock with the solution $\hat{R}_i(\bar{q}_i^*)$.

4 Experimental Results and Discussion

Extensive experimental testing and comparison were performed on several sequences with different characteristics: “AKIYO”, “COASTGUARD”, “CONTAINER”, “FOREMAN”, “MOBILE”, and “MOTHER DAUGHTER”. These sequences are in QCIF format (176×144), and the frame rate F is 30 fps. The the bandwidth BW is set to 64 kbps and the size of the sliding window N_W is equal to 12.

Fig. 2 (a) and (b) show the rate and distortion models, respectively, for the



(a) (b)

Fig. 2. Frame layer R-D modeling for the QCIF COASTGUARD sequence: (a) the rate model and (b) the distortion model as a function of the average QP of macroblocks.

Table 1. Performance comparison of the proposed algorithm with TMN8.

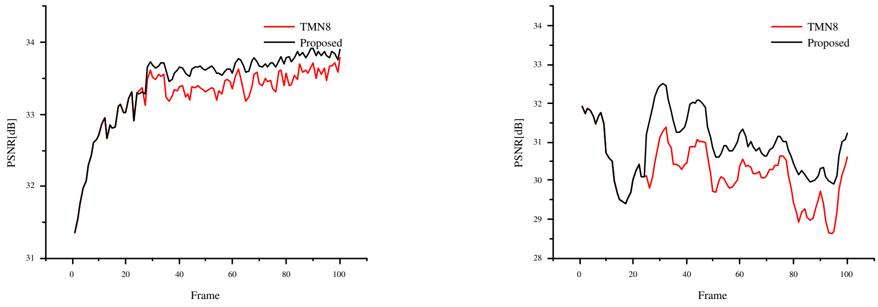
Test sequence	Rate control method	Average PSNR	σ of PSNR	Average number of bits per frame (kbits)
AKIYO	TMN 8	31.781	0.893	2.132
	Proposed algorithm	32.291	0.782	2.133
COASTGUARD	TMN 8	27.953	1.339	2.129
	Proposed algorithm	28.243	1.149	2.133
CONTAINER	TMN 8	33.259	0.454	2.133
	Proposed algorithm	33.437	0.432	2.133
FOREMAN	TMN 8	30.231	0.761	2.135
	Proposed algorithm	30.938	0.670	2.137
MOBILE	TMN 8	27.240	0.825	2.133
	Proposed algorithm	27.837	0.733	2.134
MOTHER DAUGHTER	TMN 8	31.429	1.047	2.132
	Proposed algorithm	32.041	0.929	2.131

"COASTGUARD" sequence, where the dots are the measured data points while the solid curve is the plot of the R-D model obtained by (5) and (6). As shown in these two figures, the R-D modeling works reasonably well. The R-D modeling method in fact provides a very good approximation for all test sequences in our experiment.

Performance was mainly evaluated by visual judgment since there is no standard measure currently available to evaluate subjective quality. And, as an objective measure of the distance between an original image $f(x, y)$ and its reconstructed image $g(x, y)$, peak signal-to-noise ratio (PSNR) is used.

The performance of the proposed frame-layer rate control scheme is compared with that of TMN8. For the performance comparison for six test sequences, we show the average PSNR value, the standard deviation (σ) of PSNR, and the average bits generated per frame in Table 1. It is clearly seen that when compared with TMN8 the proposed frame rate control algorithm can not only improve the average PSNR value, but also reduce the standard deviation of PSNR while the average number of bits per frame is slightly larger than that of TMN8. The PSNR plots associated with the "CONTAINER" and "FOREMAN" sequences as a function of the frame number are shown in Fig. 3 (a) and (b), respectively. This figure shows that the proposed method improves the image quality in scene whose activity is large and reduces the abrupt quality degradation. It is also known that the proposed method reduces the fluctuation of image quality.

Visual comparisons of the proposed algorithm with the TMN8 are also provided in Fig. 4. To make a comparison of the subjective quality more clear, zoomed images are also presented. It is observed in Fig. 4 that edges are well preserved by the proposed technique. Note that the proposed algorithm performs better than TMN8. Similar results were also observed on the other test images. Experimental results indicate that the proposed algorithm is a useful alternative to TMN8 in terms of both the PSNR and subjective quality.



(a) (b)

Fig. 3. PSNR comparison with a target average rate at 64kbps: (a) QCIF CONTAINER. (b) QCIF FOREMAN.

5 Conclusion

In this paper, we presented new real-time frame-layer rate control based on the sliding window for H.263+ over the Internet. In order to reduce the quality fluctuation, the sliding window method has been proposed. And, we introduced a frame-layer rate control to minimize the average distortion over an entire sequences as well as variations in distortion between frames. Since the proposed technique uses fast convergence method and does not require pre-analysis, it is suitable to real-time low-complexity video coding. The proposed algorithm has been tested on several sequences and found to provide better visual and PSNR performance than the existing TMN8 rate control algorithm.

References

1. ITU-T, Video coding for low bitrate communication, ITU-T Recommendation H.263 version 2, Jan. 1998.
2. T. Chiang and Y. Zhang, "A new rate control scheme using quadratic rate distortion model," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, no. 1, pp. 246-250, Feb. 1997.
3. T. Wiegand, M. Lightstone, D. Mukherjee, T.G. Campbell, and S.K. Mitra, "Rate-distortion optimized mode for very low bit rate video coding and emerging H.263 standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, pp. 182-190, Apr. 1996.
4. J. Ribas-Corbera and S. Lei, "Rate control in DCT video coding for low-delay communications," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 1, pp. 172-185, Feb. 1999.
5. ITU-T, *Video codec Test model, near-term, version 8 (TMN8)*, H.263 AdHoc Group, Portland, June 1997.
6. MPEG-2 Test Model 5, Doc. *ISO/IEC JTC1/SC29 WG11/93-400*, Apr. 1993.

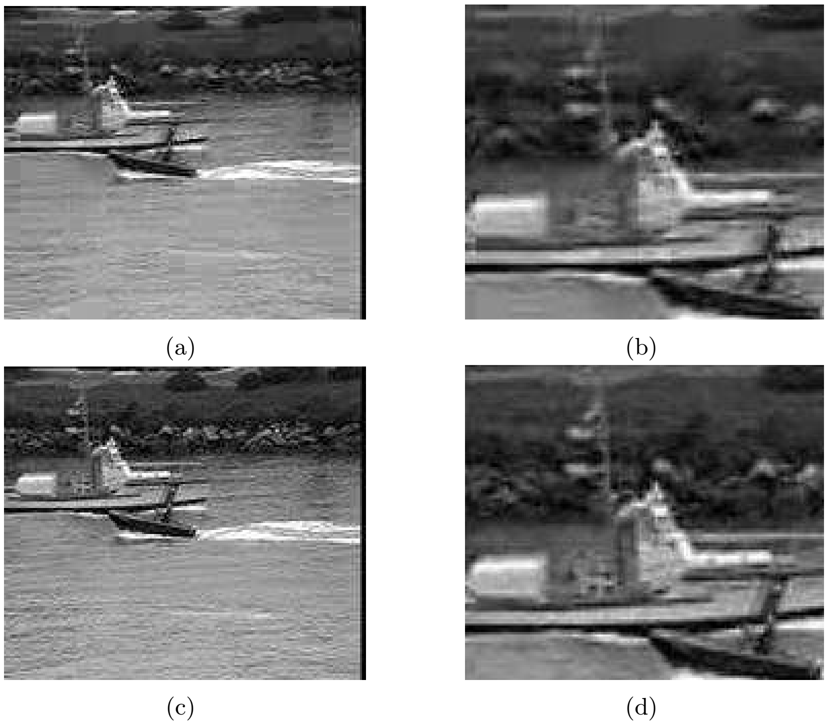


Fig. 4. Visual quality performance between proposed algorithm and TMN8 on the 50th frame of COASTGUARD. (a) TMN8. (b) Zoomed image of (a). (c) Proposed algorithm. (d) Zoomed image of (c).

7. K. Ramchandran, A. Ortega, and M. Vetterli, "Bit allocation for dependent quantization with applications to multiresolution and MPEG video coders," *IEEE Trans. Image Processing*, vol. 3, pp. 533-545, Sept. 1994.
8. P. Y. Cheng, J. Li, and C. C. Kuo, "Rate control for an embedded wavelet video coder," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, pp. 696-702, Aug. 1997.
9. ITU-T SG16/Q15, Video Test Model Number 10 (TMN10), T.Gardos, Ed., Apr. 1998.
10. ITU-T/SG15, Video codec test model, TMN7, Nice, Feb. 1997.
11. ITU-T, SG16/Q15, *Video codec Test model, near-term, version 9 (TMN9)*, Dec. 1997.
12. Y. Yokoyama and S. Nogaki, "A rate control method with pre-analysis for real-time MPEG-2 video coding," *in Proc. Int. Conf. Image Processing*, pp. 514-517, Oct. 2001.
13. L. J. Lin, A. Ortega, and C. C. J. Kuo, "Rate control using spline interpolated R-D characteristics," *in Proc. of SPIE Visual Communication Image Processing*, pp. 111-122, Mar. 1996.
14. K. Ramchandran and M. Vetterli, "Best wavelet packet bases in a rate-distortion sense," *IEEE Trans. Image Processing*, vol. 2, pp. 160-175, Apr. 1993.