MODEL-BASED RATE ALLOCATION IN DISTRIBUTED VIDEO CODING SYSTEMS

Yevgeny Priziment and David Malah

Department of Electrical Engineering
Technion IIT, Haifa 32000, Israel
evgeniyp@techunix.technion.ac.il, malah@ee.technion.ac.il

ABSTRACT

In Distributed Video Coding (DVC) systems the rate control is usually performed at the decoder through a feedback channel. However, a feedback channel incurs a delay and might be impractical for real-time applications. Typically, systems that suppress feedback use a fixed uniform quantization across the entire frame, regardless of its local characteristics. In this work a framework for intraframe rate allocation at the encoder side is suggested as a way to both suppress the feedback channel and adapt the quantization to the frame characteristics. The rate allocation utilizes an approximation model to the rate-distortion function for Laplacian data and correlation channel distributions. This model replaces an analytical solution given in integral form that can be evaluated only numerically. A practical DVC system that is based on this scheme is proposed. The use of intraframe rate allocation is indeed found to improve the performance of the feedback-less system.

Index Terms— Distributed video coding, Wyner–Ziv Rate–Distortion, Rate allocation

1. INTRODUCTION

Distributed Video Coding (DVC) is a novel coding scheme which employs principles of lossy source coding with Side Information (SI) at the decoder, also known as Wyner-Ziv (WZ) coding [1]. The DVC framework enables to shift the computational load of motion estimation from the encoder to the decoder resulting in reversed encoder-decoder complexity. This reversed complexity distribution could be appealing for applications in which the encoder is power and/or complexity constrained. Similar to standard video coding systems, performance enhancement of a DVC system can be obtained by classifying frame regions into different types and allocating the limited bitrate among these regions according to their coding complexity and perceptual importance. However, adopting these ideas for DVC is quite complex, since both the rate and the reconstruction fidelity depend on the quality of the SI, which is known only at the decoder.

This work was partly supported by Elbit Systems Ltd.

In the baseline DVC system [1], the video sequence is split into Key frames and WZ frames. The Key frames are encoded by standard intra coding techniques while the WZ frames are coded using syndromes (channel code) and decoded by combining these syndromes with SI. Actually, SI is a prediction of the WZ frame generated at the decoder by motion compensated interpolation or extrapolation of the previously decoded Key and/or WZ frames. Due to the lack of stationarity and evolving dynamics in video the joint statistics (or the correlation model) between WZ frames and SI frames should be monitored online. As a consequence, the rate control, in the baseline system, is performed by the SI aware decoder through a feedback channel. However, the feedback channel incurs additional delay and might be impractical for real-time applications.

In order to suppress the feedback channel and allow the encoder to perform rate control and bit allocation SI, or its joint statistics with the WZ frame, is needed at the encoder. However, implementing a closed-loop like system and constructing SI at the encoder will eliminate the desired low complexity. Therefore, a 'low cost' estimate of the SI, or its joint statistics with the WZ frame, should be obtained by the encoder for rate management purposes.

Despite the fact that feedback channel suppression and rate allocation are closely related problems most of the works focus only on the first one. In [2] it is proposed to estimate the SI frame by averaging adjacent frames. Furthermore, the rate to be allocated was estimated by calculating the conditional entropy of WZ frame given the estimated SI. The system operates in the pixel domain, advocated by lower complexity in comparison to a transform based one. However, the spatial redundancy is not exploited in any manner and results in inefficient coding. A similar approach, but for a transform domain system is presented in [3]. In addition, in order to obtain a more accurate SI estimate, it is proposed there to perform a fast motion estimation at the encoder for 7% of the blocks having the largest displaced frame difference energy. Nevertheless, optimal selection of quantization modes is not addressed there and, actually, the allocation is performed for some fixed quantizer. A pixel domain rate allocation method is discussed in [4]. The different image regions are treated as independent Gaussian random variables and the correlation

channel is also modeled by a Gaussian PDF. A closed form Rate-Distortion (RD) function for a quadratic Gaussian case is used in the rate allocation module. However, empirical experiments [5] have shown that modeling the data and 'correlation channel' by Laplacian PDFs results in a more accurate representation of the real data relatively to the one obtained by a quadratic Gaussian model.

In this work we describe a transform domain DVC system that performs rate allocation and suppresses the feedback channel by using an empirical WZ RD model. The rate allocation is performed for non-overlapping rectangular slices in each WZ frame. Local statistics of these slices are used to obtain the expected distortion and coding rate for a set of available qunantizers. Since SI frames are unavailable at the encoder, a coarse estimation of SI frame parameters is performed based on adjacent Key frames. A configuration of quantizers giving the minimal overall distortion for a given rate is selected. Additionally, the presented WZ RD model assumes a uniformly quantized source, a Laplacian SI and a Laplacian 'correlation channel'. The model defines the dependencies of rate and distortion on the quantization step size for given parameters of the correlation channel and SI. In Section 2 we show that the empirical model closely approximates the integral expressions in [6] that can be evaluated only numerically. In the case of bitplane by bitplane encoding, the established rate-quantization step dependency is used to split the total allocated rate for a specific subband into a set of rates, one for each of the bitplanes. The DVC system itself is described in Section 3 and the experimental results are presented in Section 4.

2. RATE DISTORTION MODEL

2.1. Analytical Model

Performance of lossy source coding with side information at the decoder can be characterized by the WZ rate-distortion function. Let X and Y denote the source and side information respectively then the RD function is given below:

$$R(d) = \inf_{\mathscr{M}(d)} \left[I(X;Z) - I(Y;Z) \right] \tag{1}$$

where Z is an auxiliary random variable and the minimization is carried over, $\mathscr{M}(d)$, a set of joint probability density functions $p(x,y,z)$ such that $Z \leftrightarrow X \leftrightarrow Y$ form a Markov chain, i.e., $p(x,y,z) = p(z|x)p(x,y)$ and there exists a fixed reconstruction function $\hat{x} = f(z,y)$ for which the average distortion is smaller or equals to d.

$$\mathbb{E}\left[D(X, f(Z,Y)) \right] \leq d \tag{2}$$

Given the joint density function of the source and the side information $p(x,y)$ and a fidelity criterion $D(x,\hat{x})$ the goal of the minimization process is to find the 'test channel' density function $p(z|x)$ and a reconstruction function $\hat{x} = f(z,y)$ which satisfy equations (1) and (2). Generally the minimization process is very complex and a closed form analytical expression for the RD function was found only for the quadratic Gaussian case. In the general case, the RD function can be evaluated numerically by extended Arimoto-Blahut algorithm [7].

Development of the RD function can be simplified by examining only a subset of the minimization domain, $\mathscr{M}(d)$. Of course, at the cost of loosing some optimality. Let $\mathscr{M}^q(d)$ denote all joint density functions $p(z|x)p(x,y)$ for which the 'test channel' is a simple uniform scalar quantizer with quantization step Δ. Let i denote the quantization index of the bin $((i-\varepsilon)\Delta, (i+1-\varepsilon)\Delta]$ where $\varepsilon \in [0,1)$ is the quantizer's offset relatively to the origin ($\varepsilon = 0.5$ corresponds to a midtread quantizer, while $\varepsilon = 0$ corresponds to a midrise quantizer). Then, the 'test channel' is defined as follows

$$p(Z=i|x) = \begin{cases} 1 & x \in ((i-\varepsilon)\Delta, (i+1-\varepsilon)\Delta] \\ 0 & otherwise \end{cases} \tag{3}$$

Furthermore, if the fidelity criterion is defined as the mean squared error $D(x,\hat{x}) = (x-\hat{x})^2$, the optimal reconstruction function is given by

$$\hat{x} = f(z,y) = \mathbb{E}[X|Y,Z] = \int_{(z-\varepsilon)\Delta}^{(z+1-\varepsilon)\Delta} x p(x|y)\, dx \tag{4}$$

Based on (4) the overall expected distortion can be calculated for any quantization step Δ (and offset ε). In addition, for a given Δ (which satisfies the distortion constraint (2)), the rate is given by

$$\begin{aligned} R(d) &= R(\Delta, \varepsilon) \\ &= I\left(X; Z_{(\Delta,\varepsilon)}\right) - I\left(Y; Z_{(\Delta,\varepsilon)}\right) \\ &= H(Z_{(\Delta,\varepsilon)}) - H(Z_{(\Delta,\varepsilon)}|X) \\ &\quad - H(Z_{(\Delta,\varepsilon)}) + H(Z_{(\Delta,\varepsilon)}|Y) = H(Z_{(\Delta,\varepsilon)}|Y) \end{aligned} \tag{5}$$

The last expression can be recognized as the minimal rate needed to losslessly encode Z given side information Y at the encoder. Thus, WZ encoding can be thought of as quantization followed by Slepian Wolf [8] encoding.

2.2. Lapacian Case and Practical Issues

A special case, in which $X = Y + N$, where Y and N are independent zero mean Laplace random variables with scale parameters α_y and α_n respectively, is analyzed in [6]. This special case is of great interest since, as it is known [9], both the AC coefficients of the DCT transformed natural images and video sequences and the difference between side information and the source, in DVC, can be modeled by Laplacian distributions. Integral form expressions for the rate $R(\Delta,\varepsilon)$ and distortion $d(\Delta,\varepsilon)$ functions were presented. However, perfect Slepian Wolf (channel) coding was assumed and the

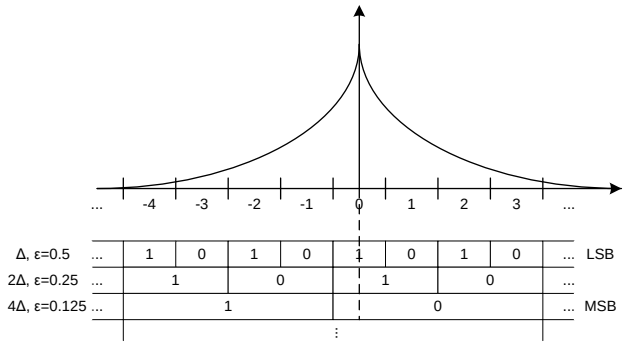

Fig. 1. Example of index assignment enabling evaluation of the number of bitplanes and the coding rate of these bitplanes

conditional entropy $H(Z|Y)$ was used to measure the coding rate, which do not apply to practical systems.

In addition to its primary goal, the RD function can be utilized to determine the number of bits to be used to represent the quantization index. Also, if a binary channel code is used in Slepian Wolf coding it can be used to determine the rate of each bitplane. Assume that $\mu_x = 0$, X is quantized by a midtread quantizer ($\varepsilon = 0.5$), with a quantization step Δ and a quantization index that is represented by k bits, $Z = z_{k-1}z_{k-2}\dots z_1 z_0$ (z_0 is LSB). Using the chain rule, the total rate can be expressed as follows:

$$R(\Delta,\varepsilon) = H(Z|Y) = \sum_{i=k-1}^{0} H(z_i|z_{k-1}\dots z_{i+1},Y)$$
$$= \sum_{i=k-1}^{0} \left[H(z_{k-1}\dots z_i|Y) - H(z_{k-1}\dots z_{i+1}|Y)\right] \quad (6)$$

Each term of the form $H(z_{k-1}\dots z_i|Y)$ represents the minimum rate needed to transmit the bitplanes $k-1$ through i. In case that the binary index representation is such that removal of each of the lower significance bitplanes merges pairs of adjacent bins than each such removal results in a uniform scalar quantizer with doubled quantization step size and twice smaller offset (see Figure 1). Thus, equation (6) can be rewritten in the following form:

$$R(\Delta,\varepsilon) = \sum_{i=k-1}^{0} \left[R(2^i\Delta,\varepsilon/2^i) - R(2^{i+1}\Delta,\varepsilon/2^{i+1})\right] \quad (7)$$

Hence, in practice, the number of bits to represent the indices might be obtained by setting some small threshold $\varsigma > 0$ and evaluating the sum at the righthand side of (7) for increasing values of k till it converges to the lefthand side within ς. Moreover, each term of the sum represents the number of bits needed to encode bitplane i given the preceding bitplnanes $k-1\dots i+1$.

Another important issue is the implementation of an infinite quantizer using a finite number of indices. Fortunately,

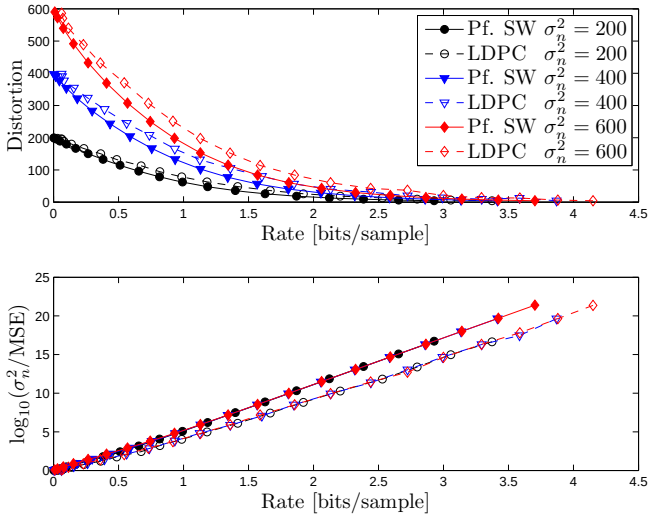

Fig. 2. Rate distortion (top) and cSNR (bottom) performance of a practical WZ encoder using LDPC code compared to an analytical bound derived in [6], derived assuming perfect SW (Pf. SW) encoding (obtained for $\sigma_y^2 = 1000$)

this problem can be resolved by index reuse as it is done in nested quantization [10]. All bins of the infinite quantizer are labeled by one of the 2^k indices in a cyclic fashion by using a modulo function. Concatenation of these 2^k bins form a coarse infinite quantizer with a step size of $2^k\Delta$. Index of the 'coarse' quantizer, i_c, in combination with the modulo relative index , i_m, uniquely identifies the bin of the infinite 'fine' quantizer, $i_f = i_c 2^k + i_m$. In distributed source coding only the modulo relative index is transmitted while the index of the 'coarse' and consequently of the 'fine' quantizer is recovered at the decoder by using the side information and the posterior distribution.

$$\widehat{i_c} = \arg\max_{i_c} \int_{(i_c 2^k + i_m - \varepsilon)\Delta}^{(i_c 2^k + i_m + 1 - \varepsilon)\Delta} p(x|y)dx \quad (8)$$

Rate distortion performance of a practical system in comparison with the analytical bounds is presented in Figure 2. The implemented system consists of the nested scalar quantizer described above and a binary rate-adaptive LDPC encoder [11]. Synthetic Laplacian signals Y and N were generated for combination of $\sigma_y^2 = 1000$ and $\sigma_n^2 = \{200, 400, 600\}$. The resulting source signal X was quantized by a uniform scalar quantizer with varying quantization step. As it can be seen, there is a gap in performance which is due to the gap between the channel rate and channel capacity. In addition, as it can been seen from Figure 2, both analytical and practical correlation Signal to Noise Ratio (cSNR) curves can be considered invariant to the 'correlation' noise variance, σ_n^2. This interesting observation can become handy when modeling the RD function.

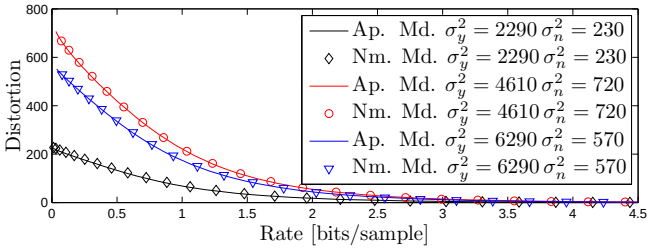

Fig. 3. Comparison of rate distortion functions calculated numerically (denoted by Nm. Md.) and by using approximation equations (9) and (10) (denoted by Ap. Md.)

2.3. Simple RD Model for Laplacian Case

As mentioned earlier, the RD expressions presented in [6] for the Laplacian case are given in integral form. These expressions can be evaluated numerically for given values of α_y, α_n, Δ and ε. However, if rate control is performed by the encoder, and preserving encoder's complexity as low as possible is one of the goals, then numerical integral evaluations are not appropriate. A possible solution is to approximate the RD curves by a computationally less expensive model. Empirically, it was found that equations (9) and (10) give a good approximation to the real rate and distortion as a function of the quantization step Δ.

$$R(\Delta) = \exp\left[a_r e^{-\left(\frac{\Delta}{b_r}\right)^{\gamma_r}} + m_r\Delta + n_r\right] \tag{9}$$

$$D(\Delta) = \exp\left[a_d e^{-\left(\frac{\Delta}{b_d}\right)^{\gamma_d}} + n_d\right], \tag{10}$$

where the parameters $\{a_r, b_r, \gamma_r, m_r, n_r\}$ and $\{a_d, b_d, \gamma_d, n_d\}$ depend on α_y, α_n and ε. For each of the parameters, a table defining their dependency on a range of α_y, α_n and ε values is stored. The performance of the proposed approximation to the RD function, compared to the analytical one, which was evaluated numerically, are presented in Figure 3. The set of parameters was obtained offline by fitting the approximation model to the numerical one for all combinations of the following values: $\varepsilon = 0.5$, $\sigma_y^2 = \{1000, 2000, \ldots, 10000\}$ and $\sigma_n^2 = \{100, 200, \ldots, 900\}$. The approximation model and the parameter tables were used to evaluate the RD function for arbitrary pairs of σ_y^2 and σ_n^2. In case that the tables do not contain an entry for some pair σ_y^2 and σ_n^2 (as is the case in Figure 3), the needed parameters are obtained online using simple linear interpolation between existing entries.

Another option is to utilize the nearly linear slope observed in Figure 2, for rates higher than 1 bit/sample . In this case the distortion (rate) can be deduced from the given rate (distortion) value. Adopting the latter approach may save a significant amount of storage space.

3. RATE ALLOCATION IN DVC SYSTEMS

The system for source coding, with side information at the decoder and the RD model for the Laplacian case, which were discussed in the previous section, can be utilized in a rate allocation module of a DCT domain DVC system. As mentioned earlier, working in the DCT domain enables quite accurate modeling of the transform coefficients and correlation channel by a Laplacian distribution. In addition, similarly to standard video coding, the DCT transform is an efficient tool for exploiting spatial redundancy. Moreover, an integer DCT transform, as used in H.264, can be applied to reduce the transform's computational complexity.

In a practical DVC system, contrary to the synthetic case, the joint PDF parameters α_y and α_n vary temporally and spatially and are not known at the encoder. Their variation depends on the spatio-temporal dynamics of the video sequence and the accuracy of WZ frame prediction at the decoder, i.e., the SI construction. Recalling that the assumed correlation model is such that $X = Y + N$, and X is readily available at the encoder, it remains to obtain Y or its statistics (parameterized by α_y) in order to enable encoder side rate control. Imitating the decoder at the encoder and constructing the SI by using motion estimation will result in a high complexity encoder, similar to the one used in standard hybrid video coders.

A low complexity estimate of the SI can be obtained by averaging the adjacent Key or WZ frames. This method can give a good estimate for low motion sequences where large parts of the imaged scene remain static. However, for sequences with medium to high motion activity this method will produce a 'pessimistic' estimate corresponding to a very noisy 'correlation channel'. This will result in a rate overestimation by the encoder. Unfortunately, the over-allocated bits can not be used at the decoder to improve the quality of the reconstructed image. Consequently, overestimation should be minimized, possibly by using fast motion estimation at the encoder for SI estimation or by employing some model which can describe the relationship between the encoder's SI estimate and the real SI at the decoder.

Once the encoder has obtained an SI estimate, the joint statistics can be evaluated. Partitioning the frame into non-overlapping slices and evaluating this statistics separately for each slice enables capturing, to some extent, the spatial variations. This partitioning facilitates the allocation of a different amount of bits to each slice, according to its statistics. Thus, for example, slices with higher motion activity will get more bits which will assist the decoder to overcome the usually less accurate prediction for such slices. On the other hand, regions with light motion, or with no motion at all, are usually characterized by a high fidelity prediction. Consequently, a relatively small amount of bits will suffice for their decoding and reconstruction.

The rate allocation among different slices, can actually be formulated as an optimization problem in which the goal is to

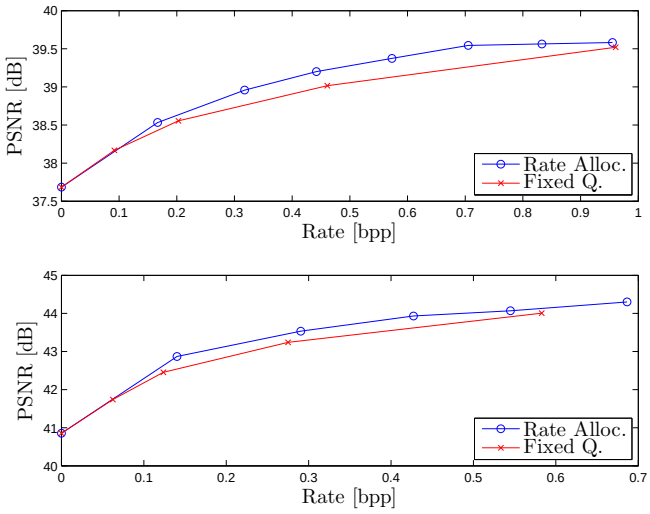

Fig. 4. Performance of a DVC system with proposed rate allocation method (Rate Alloc.) vs. fixed quantization (Fixed Q.) for *Hall Monitor* (top) and *Mother and Daughter* (bottom) sequences

minimize the overall distortion subject to some constraint on the total rate. In practice, the rate of each slice is controlled by its quantizer quality factor, which is selected from a finite predefined set. For each quality factor the cumulative rate and distortion over all bands in each one of the slices are calculated. The goal is to choose for each slice a quality factor such that the total rate is not higher than the maximum allowable rate and the total distortion is as small as possible. An efficient algorithm for solving this problem is given in [12].

4. SIMULATION RESULTS

A DVC system based on the principles described in the previous sections was implemented and tested on *Hall Monitor* and *Mother and Daughter* CIF sequences. The system key settings are as follows: The GOP size is 2 (Key–WZ–Key). The Key frames are assumed to be losslessly recovered at the decoder and all the presented results refer only to WZ frames. At the encoder, the SI estimate is obtained by averaging the adjacent Key frames. Each frame is partitioned into 16 slices such that each slice contains 396 4×4 blocks. Channel (syndrome) coding is implemented by using a binary LDPC code. In case of unsuccessful decoding, due to bitrate under-allocation, the entire band from SI is adopted. Simulation results are presented in Figure 4 and as it can be seen a gain over feedback-less system with fixed quantization is obtained.

5. CONCLUSION

A DVC system with encoder side rate control and rate allocation is proposed. Practical aspects of WZ coders based on scalar quantizer and channel encoding are discussed. An empirical model approximating the RD function for the Laplacian case is presented as an alternative to the analytical model, which can be evaluated only numerically. Integration of the RD function with rate control and bitrate allocation is demonstrated.

Frame partitioning can be further utilized for perceptual quality enhancement by weighting the different regions according to their perceptual importance. This issue will be examined in the future work along with integration of non-binary LDPC codes in order to reduce the practical-theoretical performance gap. Issues arising from coding of video sequences with high motion activity will be treated as well.

6. REFERENCES

[1] B. Girod, A.M. Aaron, S. Rane, and D. Rebollo-Monedero, "Distributed video coding," *Proc. of the IEEE*, vol. 93, no. 1, pp. 71–83, Jan. 2005.

[2] M. Morbee, J. Prades-Nebot, A. Roca, A. Pizurica, and W. Philips, "Improved pixel-based rate allocation for pixel-domain distributed video coders without feedback channel," *LNCS*, vol. 4678, pp. 663, 2007.

[3] C. Brites and F. Pereira, "Encoder rate control for transform domain Wyner-Ziv video coding," *IEEE Intl. Conf. on Image Processing, ICIP*, vol. 2, pp. II –5–II –8, 2007.

[4] P. Wang, J. Wang, S. Yu, and Y. Pang, "Theory and practice of rate division in distributed video coding," *IEICE Trans. on Fundamentals of Electronics, Communications and Computer Sciences*, vol. 91, no. 7, pp. 1806–1811, 2008.

[5] C. Brites and F. Pereira, "Correlation noise modeling for efficient pixel and transform domain Wyner-Ziv video coding," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 18, no. 9, pp. 1177–1190, Sept. 2008.

[6] V. Sheinin, A. Jagmohan, and D. He, "Uniform threshold scalar quantizer performance in Wyner-Ziv coding with memoryless, additive laplacian correlation channel," *Proc., Intl. Conf., Acoustics, Speech and Signal Processing. ICASSP*, vol. 4, May 2006.

[7] F. Dupuis, W. Yu, and F.M.J. Willems, "Blahut-Arimoto algorithms for computing channel capacity and rate-distortion with side information," *Proc., Intl. Symp., ISIT*, July 2004.

[8] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans., Information Theory*, vol. 19, no. 4, pp. 471–480, 1973.

[9] E.Y. Lam, "Analysis of the dct coefficient distributions for document coding," *IEEE Signal Processing Letters*, vol. 11, no. 2, pp. 97–100, Feb. 2004.

[10] R. Zamir, S. Shamai, and U. Erez, "Nested linear/lattice codes for structured multiterminal binning," *IEEE Trans., Information Theory*, vol. 48, no. 6, pp. 1250–1276, 2002.

[11] D. Varodayan, A. Aaron, and B. Girod, "Rate-adaptive codes for distributed source coding," *Signal Process.*, vol. 86, no. 11, pp. 3123–3130, 2006.

[12] Y. Shoham and A. Gersho, "Efficient bit allocation for an arbitrary set of quantizers," *IEEE Trans., Acoustics, Speech, and Signal Processing*, vol. 36, no. 9, pp. 1445–1453, 1988.