Asymptotically Optimal Scalable Coding for Minimum Weighted Mean Square Error

Ashish Aggarwal, Shankar L. Regunathan and Kenneth Rose
Department of Electrical and Computer Engineering
University of California
Santa Barbara, CA 93106
Email:[ashish,otto,rose]@scl.ece.ucsb.edu

Abstract

In this paper, we derive an asymptotically optimal multi-layer coding scheme for entropy-coded scalar quantizers (SQ) that minimizes the weighted mean-squared error (WMSE). The optimal entropy-coded SQ is non-uniform in the case of WMSE. The conventional multi-layer coder quantizes the base-layer reconstruction error at the enhancement-layer, and is sub-optimal for the WMSE criterion. We consider the compander representation of the quantizer, and propose to implement scalability in the compressed domain. We show that such a multi-layer coding system achieves the operational rate-distortion bound given by the non-scalable entropy-coded SQ, at the limit of high resolution. Simulation results for a synthetic memoryless Laplace source with μ-law companding are presented for various values of layer rates. Substantial gains are also achieved on the "real-world" sources of audio signals, when the optimal multi-layer approach is applied to a two-layer scalable MPEG-4 Advanced Audio Coder.

I. Introduction

Bit rate scalability is emerging as a major requirement in wireless and networking world. A scalable bit stream allows the decoder to produce a coarse reconstruction if only a portion of the bit stream is received, and to improve the quality as more of the total bit stream is made available. Scalability is especially important in applications, such as digital audio/video broadcasting and multicast audio, which require simultaneous transmission over multiple channels of differing capacity. Further, a scalable bit stream provides robustness to packet loss for tranmission over packet networks (e.g., over the Internet). However, the major objection to incorporating bit rate scalability

This work is supported in part by the NSF under grants no. MIP-9707764, EIA-9986057, the University of California MICRO Program, Conexant Systems, Inc., Lernout & Hauspie Speech Products, Lucent Technologies, Inc., Medio Stream, Inc., and Qualcomm, Inc.

within existing coders is the loss in performance relative to the non-scalable coding. In particular, the degradation in reconstruction quality, measured by a perceptually motivated weighted mean square error (WMSE) criterion, can be severe when the scalable coder employs entropy-coded scalar quantization (ECSQ). For example, the recent MPEG-4 Advanced Audio Coder (AAC) uses ECSQ, and pays a substantial performance penalty for providing a scalable bitstream.

ECSQ is widely used as it offers ease of design and implementation without sacrificing much performance when compared to vector quantization [1]. The typical ECSQ-based scalable coder consists of coding layers where each layer requantizes the reconstruction error of the preceding layer. This approach is asymptotically optimal *only* if the distortion measure is the standard mean squared error (MSE). For a perceptually motivated WMSE measure, and in contrast to MSE criterion, quantization intervals of the ECSQ do not become uniform even at high resolution. A direct requantization of reconstruction error at the enhancement-layer does not take into account the information conveyed by the base-layer index regarding the quantization interval. The enhancement-layer ECSQ remains mismatched to the task of minimizing the WMSE criterion. As a result, the conventional scalable coder may considerably underperform the non-scalable coder. A brute force method of improving compression performance of a scalable coder is to design a different enhancement-layer ECSQ for each base-layer index. However, this method seems impractical as it requires excessive memory. An alternate approach to efficient multi-layer coding, which performs linear transformation of the reconstruction error based on the base-layer index, has been proposed for fixed rate vector quantizers by Lee, Neuhoff and Paliwal [2].

In this paper, we derive an asymptotically optimal multi-layer coding scheme for ECSQs under the WMSE criterion. We consider the companding representation of the ECSQ: here, optimization of the WMSE of the original signal is equivalent to optimizing the MSE of the compressed signal. We then exploit the fact that it is possible to requantize the reconstruction error without loss of optimality if the distortion measure is MSE. Hence, we propose to achieve optimal multi-layer coding by requantizing *in the compressed domain*. The optimal multi-layer approach is first demonstrated on a synthetic memoryless Laplace source under μ-law companded system. To demonstrate the impact on practical applications, we implement the scheme within the MPEG-4 Advanced Audio Coder [3][4][5]. The optimal companding scheme gives considerable gains over the existing method.

The organization of the paper is as follows: Section II provides a brief overview of ECSQ. The conventional approach to multi-layer coding is outlined in Section III. The proposed asymptotically optimal scalable coding scheme is outlined in Section IV and experimental results are discussed in Section V.

II. Preliminaries

A. Notation

Let $x \in \mathcal{R}$ be a scalar random variable with probability density function (pdf) $f_x(x)$. A (non-uniform) scalar quantizer (SQ), $Q(x)$, partitions $\mathcal{R}$ into disjoint and exhaustive regions, S_i, such that $Q(x) = \hat{x}_i$ if $x \in S_i$. A high resolution quantizer $Q(x)$ can be described by an unnormalized point density function $\Lambda(x)$ [6]. The distortion criterion is the weighted mean-squared error (WMSE) given by,

$$D = E[(x - Q(x))^2 w(x)] = \int_x (x - Q(x))^2 w(x) f_x(x) dx$$

where, $w(x)$ is the weight function and E is the expectation operator. By the high-resolution approximation, D may be rewritten as follows: [7]

$$D \approx \frac{1}{12} \int_x \frac{w(x) f_x(x)}{\Lambda(x)^2} dx. \tag{1}$$

The rate of the quantizer can be approximated by the entropy of the quantized output,

$$R \approx h(X) - E[log(\frac{1}{\Lambda(X)})],$$

where $h(X)$ is the differential entropy of X. All logarithms are to the base 2 and the rate is hence measured in bits.

Let us consider the equivalent companding representation for the quantizer, which consists of a compressor, a uniform SQ and an expander (inverse compressor). Such a scheme is shown in figure 1. The compressor function, $c(x)$, and the uniform SQ stepsize, Δ, are related to Λ via,

$$\frac{\partial c(x)}{\partial x} = c'(x) = \Delta\Lambda(x).$$

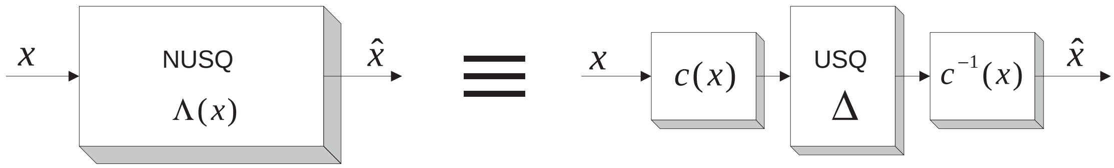


Fig. 1. Block diagram of a non-uniform scalar quantizer and its equivalent companding scheme

B. Non-scalable Entropy-Coded Scalar Quantizer

The best non-scalable entropy-coded SQ (ECSQ), Q_{ns}, is one that minimizes D subject to the rate constraint, $R \leq R_c$, and is given by [7]

$$\Lambda_{ns}(x) = \arg \min_{\Lambda(x):R\leq R_c} D = \sqrt{w(x)}/\Delta, \tag{2}$$

where the normalizing constant, Δ, equals the stepsize of the uniform SQ in the compressed domain and is given by, $log(\Delta) = h(X) - R_c + E[\log(w(x))]/2$.

The operational distortion-rate function of the non-scalable ECSQ, δ_{ns}, is easily derived from (1) and (2) as,

$$\delta_{ns}(R) = \frac{1}{12} 2^{2(h(X)-R)-E[\log(w(x))]}. \tag{3}$$

III. Conventional Scalable Coding with ECSQ

Consider a two-layer scalable coder. In the standard approach, the enhancement-layer simply quantizes the reconstruction error of the base-layer. We refer to this scheme as the conventional scalable (CS) coding scheme. MPEG-4's Advanced Audio Coder (AAC) is an example of CS coding where the base and the enhancement-layer compressor functions are identical.

The block diagram of a CS coder is shown in Figure 2. The base-layer compressor is denoted by $c_b(x)$ and the uniform SQ stepsize by Δ_b. Similarly, at the enhancement-layer, $c_e(x)$ and Δ_e are the compressor function and the uniform SQ stepsize respectively. Let $\hat{x}$ be the overall reconstructed value of x, z be the reconstruction error at the base-layer, and let $\hat{x}_i$ be the reconstructed value of x when $x \in S_i$ at the base-layer, then the distortion for the CS scheme is,

$$\begin{aligned} D_{cs} &= \int (x-\hat{x})^2 w(x) f_x(x) dx \\ &\approx \sum_i \int_{x\in S_i} (x-\hat{x}_i)^2 w(\hat{x}_i) f_x(\hat{x}_i) dx \\ &= \sum_i w(\hat{x}_i) f_x(\hat{x}_i) \int_{-\Delta_b/2c_b'(x_i)}^{\Delta_b/2c_b'(x_i)} (z-\hat{z})^2 dz \end{aligned}$$

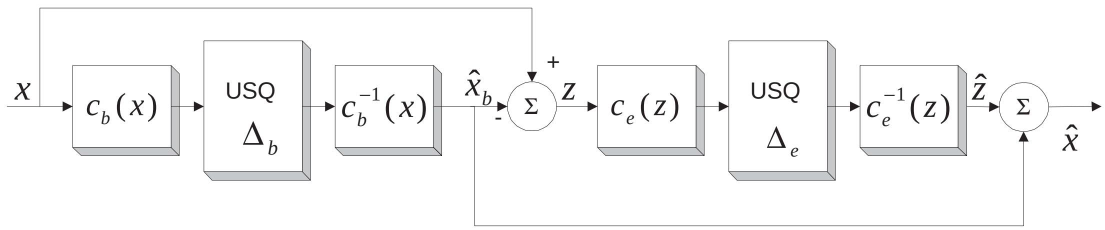


Fig. 2. Block diagram of a conventional scalable coder using non-uniform scalar quantizers

$$
\begin{aligned}
&= \frac{\Delta_e^2}{12}\sum_i w(\hat{x}_i) f_x(\hat{x}_i) \int_{z:2c_b'(\hat{x}_i)|z|\le\Delta_b} \frac{1}{c_e'(z)^2} dz \\
&\approx \frac{\Delta_e^2}{12}\int_z \frac{1}{c_e'(z)^2} \int_{x:2c_b'(x)|z|\le\Delta_b} \frac{w(x)c_b'(x)f_x(x)}{\Delta_b} dx dz \\
&= \frac{\Delta_e^2}{12}\int_z \frac{K(z)}{c_e'(z)^2} dz
\end{aligned}
\tag{4}
$$

where $K(z) = \int_{x:2c_b'(x)|z|\le\Delta_b} w(x)c_b'(x)f_x(x)/\Delta_b dx$.

Note that an optimal base-layer quantizer is one that will satisfy - $w(x) = c_b'(x)^2 = \Delta_b^2\Lambda_b(x)^2$. Therefore, we have

$$
K(z) \;=\; \Delta_b^2 \int_{x:2\Lambda_b(x)|z|\le 1} \Lambda_b(x)^3 f_x(x) dx. \tag{5}
$$

The base and enhancement-layer rates are related to the quantizer stepsize by,

$$
\begin{aligned}
R_b &= h(X) + E[\log(c_b'(x))] - \log(\Delta_b) \\
R_e &= h(Z) + E[\log(c_e'(x))] - \log(\Delta_e).
\end{aligned}
\tag{6}
$$

The performance of CS in (4), is strictly worse than the bound, unless $w(x) = 1$.

IV. Asymptotically Optimal Scalable Coding with ECSQ

An asymptotically optimal scalable (AOS) coder is one whose performance achieves the bound δ_{ns}. For a WMSE measure, the best CS scheme may considerably underperform an AOS scheme. One way to obtain an AOS coder is to exploit all the information from the base-layer by designing a separate enhancement-layer ECSQ per each base-layer index. The obvious drawback of such a scheme is its heavy memory requirement.

We can retain the simplicity of CS scheme and achieve asymptotically optimal performance by taking advantage of the following two observations:

1. *CS is optimal for the MSE criterion ($w(x) = 1$).*

For MSE, the best ECSQ is a uniform [6] at high resolution. The pdf of the reconstruction error is also uniform and hence, the best quantizer at the enhancement-layer is also uniform. The compression function is given by $c_b'(x) = c_e'(x) = \sqrt{w(x)} = 1$. The base and enhancement-layer rates in (6) reduce to,

$$\begin{aligned} R_b|_{w(x)=1} &= h(X) - log(\Delta_b) \\ R_e|_{w(x)=1} &= h(Z) - log(\Delta_e) = log(\Delta_b) - log(\Delta_e). \end{aligned}$$

For MSE, (5) reduces to $K(z) = f_z(z)$ [8], and distortion can be rewritten as

$$\begin{aligned} D_{cs}|_{w(x)=1} &= \frac{1}{12}\Delta_e^2 \\ &= \frac{1}{12}2^{(h(X)-(R_b+R_e))} \\ &= \delta_{ns}(R_b + R_e)|_{w(x)=1}. \end{aligned}$$

2. *For an optimally companded ECSQ, the WMSE of the original signal equals MSE of the compressed signal.*

For the optimal compressor function, Bennett's integral (1) reduces to $D = \Delta^2/12$ (using (2)), which equals the MSE (in compressed domain) of the uniform SQ.

Using the above observations, we build the block diagram of AOS as shown in figure 3. Let the base and enhancement-layer uniform SQs in the compressed domain have stepsizes Δ_b and Δ_e respectively. Let D_{aos} be the distortion of the AOS scheme, and R_b and R_e be the base and enhancement-layer rates. Choose the compressor function to be $c'(x) = \sqrt{w(x)}$ and let $y = c(x)$ be the compressed signal. The rate-distortion performance of the coder is obtained as follows:

$$\begin{aligned} D_{aos} &= \Delta_e^2/12 \\ R_b &= h(Y) - \log(\Delta_b) = h(X) + E[\log(c'(x))] - \log(\Delta_b) \\ R_e &= log(\Delta_b) - log(\Delta_e) \\ \Rightarrow D_{aos} &= \frac{1}{12}2^{2(h(X)-(R_b+R_e))+E[\log(w(x))]} \\ &= \delta_{ns}(R_b + R_e). \end{aligned}$$

We thus achieve asymptotical optimality.

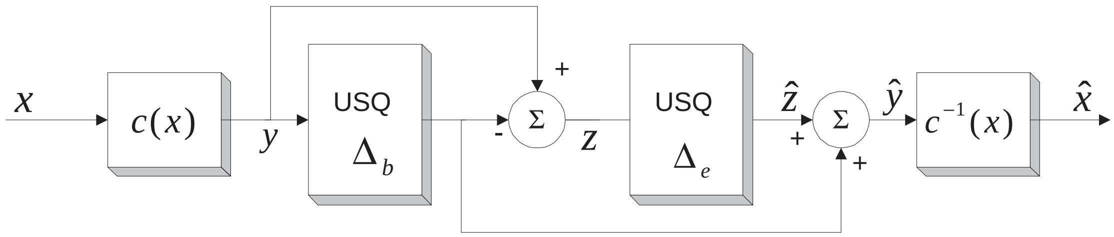


Fig. 3. Block diagram of the asymptotically optimal scalable coder

V. Simulation Results

A. *μ-law companding for a memoryless Laplace source*

The performance of the AOS scheme is compared experimentally to the CS scheme for μ-law companding [9]. The CS and the AOS schemes are implemented as described in Section III and Section IV, respectively. The input is a memoryless Laplace source with zero mean and variance $\sigma^2 = 100$. Consideration of this process is motivated by the observation that speech, image and video signals possess marginal densities that are closely approximated by Laplacian densities [10][11][12]. The x_{max} value of μ-law is set to 7σ and hence, the variance of the source has no effect on the operational RD curves. To implement a WMSE criterion that corresponds to the μ-law, we use a "reverse engineering" approach, i.e., we compute the WMSE distortion measure for which μ-law is the optimal compressor function. The weights derived from the compressor function are,

$$w(x) = \left(\frac{x_{max}}{\ln(1+\mu)} \frac{\mu / x_{max}}{1 + \mu x / x_{max}} \right)^2 .$$

In our implementation of the CS scheme, the compressor functions for the base and the enhancement-layer are identical.

A.1 RD values for different base-layer rates

In figure 4 we compare the operational RD curves of the AOS and the CS schemes for different base-layer rates. The value of μ is kept constant at 255. Figure 4(a) shows the convex-hull of the curves for (a relatively high) base-layer rate of 2.5 bits/sample. At enhancement-layer rates of $\geq$ 2.5 bits/sample (total rate $\geq$ 5 bits/sample), the performance of AOS nearly equals the non-scalable performance, and achieves a gain of $\approx$ 1.5 bits/sample over CS. A limiting (low-rate) case for base-layer rate of 1 bit/sample is shown in the figure 4(b). The AOS curve is somewhat above from the non-scalable curve. However, the CS curve is considerably suboptimal even at high enhancement-layer rates.

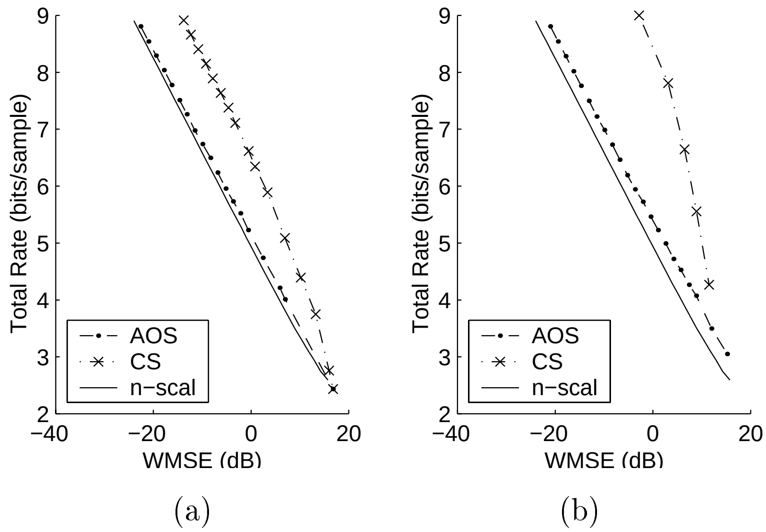


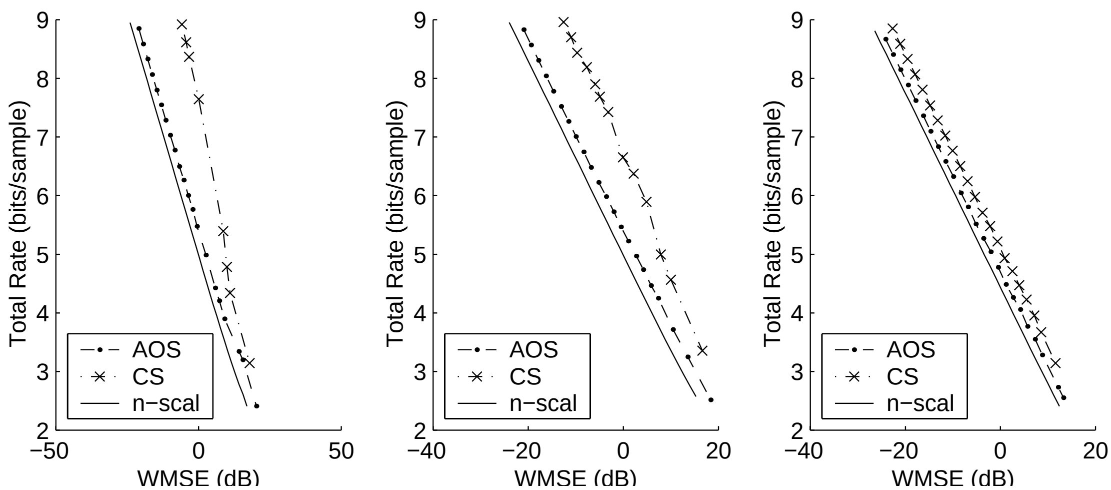
Fig. 4. Performance of a μ-law companding scheme for different base-layer rates. Competing Methods: AOS(proposed) and CS. Performance of non-scalable coder is shown for reference. Memoryless Laplace source (variance $\sigma_x^2 = 100$), ($\mu = 255, xmax = 7\sigma_x$). base-layer rate = (a) *2.5 bits/sample*, (b) *1 bit/sample*

A.2 RD values for different μ

Figure 5 plots the RD curves for AOS and CS schemes as μ is varied. The base-layer rate is kept constant at 1 bit/sample. As we decrease μ we gradually approach the MSE criterion and the CS approaches AOS (identical for $\mu = 0$).

B. Performance of AOS on MPEG-4 Advanced Audio Coding (AAC)

In AAC, audio samples in time domain are transformed using a modified discrete cosine transform. The transform coefficients are then quantized using ECSQ. The compressor function used is $|x|^{0.75}$, and the weights are derived from a psychoacoustic model. Additionally, for a band of coefficients, AAC transmits the stepsize of the uniform SQ used to quantize the compressed transform coefficients in that band. The quantized coefficients are entropy coded using a Huffman code. Scalable coding in AAC is done using the CS approach. We implemented the AOS scheme on AAC by quantizing the coefficients in the compressed (and scaled at base-layer) domain using the same uniform SQ for all the bands. Table I shows the performance of the CS AAC and the AOS AAC schemes for a two-layer coder at different base and



(a) (b) (c)

Fig. 5. Performance of a μ-law companding scheme for different μ. Competing Methods: AOS(proposed) and CS. Performance of non-scalable coder is shown for reference. Memory-less Laplace source - variance $\sigma_x^2 = 100$, μ-law - xmax=$7\sigma_x$, base-layer rate = 1 bit/sample. (a) μ*=150*, (b) μ*=50*, (c) μ*=5*

enhancement-layer rates. Clearly, AOS considerably outperforms CS.

Rate (bits/second) (base + enhancement)	WMSE (dB)	
	CS AAC	AOS AAC
16000 + 16000	8.0941	7.0728
16000 + 32000	5.9514	5.2479
32000 + 32000	4.8689	1.8919
48000 + 48000	-1.9656	-4.3047

TABLE I

PERFORMANCE OF A TWO-LAYER AAC CODER FOR THE CONVENTIONAL SCALABLE AND ASYMPTOTICALLY OPTIMAL SCALABLE (PROPOSED) APPROACHES.

VI. CONCLUSION

In this paper, we derived a asymptotically optimal scalable (AOS) coding scheme for entropy-coded scalar quantizers which optimizes the practically important weighted mean-squared error criterion. Conventional scalable approaches requantize the base-layer reconstruction error at the enhancement-layer. The quantizer intervals at the enhancement-layer are mismatched to the distortion measure resulting in poor performance. The proposed approach performs requantization in the compressed domain where the distortion metric is mean squared error, and the optimal quantizer intervals are uniform. AOS was shown to achieve the rate-distortion performance of a non-scalable coder in the high resolution limit. Simulation results on μ-law companding scheme with a memoryless Laplace source demonstrated that AOS can achieve significant gains over conventional scalable coding, and eliminate a large fraction of the performance penalty due to scalability. In a practical coding scenario, substantial gains were observed for a two-layer MPEG-4 Advanced Audio Coder.

References

[1] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*, ch. 9, pp. 295–302. Kluwer Academic, 1992.

[2] D. H. Lee, D. L. Neuhoff, and K. K. Paliwal, "Cell-conditioned multistage vector quantization," *Proc. of ICASSP*, pp. 653–6, 1991.

[3] ISO/IEC JTC1/SC29, "Information technology - very low bitrate audio-visual coding," *ISO/IEC IS-14496 (Part 3, Audio)*, 1998.

[4] M. Bosi, *et al.*, "ISO/IEC MPEG-2 advanced audio coding," *Journal of Audio Engineering Society*, vol. 45, pp. 789–814, October 1997.

[5] "The MPEG audio web page." http://www.tnt.uni-hannover.de/project/mpeg/audio/.

[6] A. Gersho, "Asymptotically optimal block quantization," *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 373–380, July 1979.

[7] J. Li, N. Chaddha, and R. M. Gray, "Asymptotic performance of vector quantizers with a perceptual distortion measure," *IEEE Trans. Inform. Theory*, vol. 45, pp. 1082–90, May 1999.

[8] D. H. Lee and D. L. Neuhoff, "Asymptotic distribution of the errors in scalar and vector quantizers," *IEEE Trans. Inform. Theory*, vol. 42, pp. 446–60, March 1996.

[9] N. S. Jayant and P. Noll, *Digital Coding of Waveforms: principles and applications to speech and video*, ch. 4, pp. 115–220. Prentice-Hall, 1984.

[10] M. D. Paez and T. H. Glisson, "Minimum mean-squared-error quantization in speech pcm and dpcm systems," *IEEE Trans. on Comm.*, vol. COM-20, pp. 225–30, April 1972.

[11] R. C. Reininger and J. D. Gibson, "Distributions of the two-dimensional dct coefficients for images," *IEEE Trans. on Comm.*, vol. COM-31, pp. 835–9, June 1983.

[12] G. J. Sullivan, "Efficient scalar quantization of exponential and laplacian random variables," *IEEE Trans. Inform. Theory*, vol. 42, pp. 1365–74, September 1996.