Dequantization Bias for JPEG Decompression

Jeffery R. Price

IRIS Lab, Dept. of ECE
University of Tennessee
Knoxville, TN USA
jrp@utk.edu

Majid Rabbani

Imaging Science Division
Eastman Kodak Company
Rochester, NY USA
majid.rabbani@kodak.com

Abstract

Standard JPEG decompression reconstructs quantized DCT coefficients to the center of the quantization bin. This fails to exploit the nonuniform distribution of the AC coefficients. Assuming a Laplacian distribution, we derive a maximum likelihood estimate of the Laplacian parameter, based on the quantized coefficients available at the decoder, and use this estimate to optimally bias the reconstruction levels during decompression. As a decoder enhancement, this technique is fully compatible with the JPEG standard and does not modify the JPEG compressed bit stream. Extensive simulations indicate that, at typical compression ratios, biased reconstruction results in modest PSNR improvements – about 0.25 dB or higher – and slight subjective improvements, for little or no computational cost. Furthermore, simulations show that the PSNR improvements are very close (within 0.07 dB) to the best theoretically possible.

1. Introduction

The JPEG image compression standard [1] is used in a wide variety of consumer and professional digital imaging applications. Since the JPEG standard only defines a decoder syntax, much research in recent years has focused on achieving the highest possible image quality for a given bit rate (or file size) while maintaining compliance with this syntax. This research can be divided into two major categories. One category has focused on improving encoder performance by using sophisticated, image-dependent optimization routines [2]-[5]. This approach generally results in a significantly more complex encoder (compared to the decoder) and can be suitable for applications that need to encode the image only once (and usually off line), but need to decode it multiple times. The second category has focused on enhancing the decoder performance via either decoder modifications or post-processing, such as reducing the blocking artifacts in highly compressed JPEG images. A comprehensive survey and bibliography of such techniques can be found in [6].

One easily implemented decoder modification, which results in a modest PSNR improvement and mildly reduces some artifacts, involves modifying the suboptimal dequantization of the AC discrete cosine transform (DCT) coefficients performed by the standard JPEG decoder. This has been noted previously [7, 8], but in this work we derive a maximum likelihood (ML) estimate of the Laplacian distribution parameter describing the AC coefficients based on the quantized values available at the decoder. We then use this estimate to optimally bias the reconstruction levels. Additionally, we demonstrate that our results are very close (less than 0.07 dB) to the best theoretically possible PSNR improvement resulting from reconstructing to the true centroid of each quantization interval.

In Section 2, a brief overview of the JPEG compression standard as it pertains to the current research is presented. In Section 3, the modeling of the AC coefficient distribution is discussed, and based upon this model, the selection of an optimal reconstruction value is formulated. In Section 4, the estimation of the parameter necessary to characterize the model distribution is discussed. Experimental results are presented in Section 5 and concluding remarks appear in Section 6.

2. Overview of the JPEG compression standard

In JPEG, the fundamental data unit is an 8×8 block of pixels, and each 8×8 block is processed independently (except for the DC value that is differentially coded). A block diagram of the JPEG encoder is depicted in Fig. 1. The original 8×8 block of pixel values, denoted as $f(i, j)$, undergoes a DCT operation resulting in an 8×8 array of transform coefficients denoted by $F(u, v)$. The DCT provides a spatial frequency decomposition of the block and aids compression by packing most of the block energy into only a few coefficients. The top-left DCT coefficient is proportional to the average brightness of the image block and is referred to as the DC coefficient, while the 63 remaining coefficients are referred to as AC coefficients.

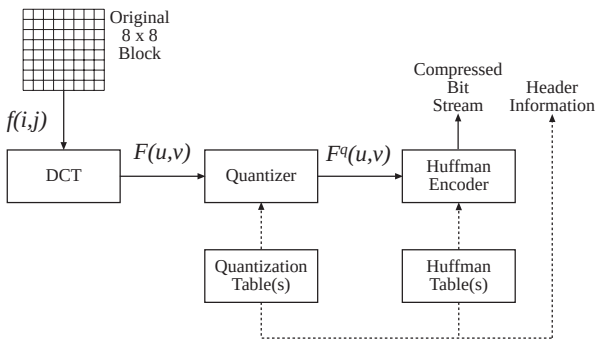

Figure 1. JPEG encoder block diagram.

Each DCT coefficient is then quantized using a simple uniform scalar quantization process with a user-defined step size. An important aspect of JPEG quantization is that the quantizer step size is allowed to vary with spatial frequency. Specifically, the JPEG quantization rule is defined as:

$$F^q(u,v) = \text{nint}\Big[\frac{F(u,v)}{Q(u,v)}\Big] \tag{1}$$

where $Q(u,v)$ are the quantizer step sizes as a function of spatial frequency, and nint$[\cdot]$ denotes rounding to the nearest integer. The 8×8 matrix, $Q(u,v)$, comprises 64 user-defined values, one for each DCT coefficient, and is commonly referred to as the quantization table (or q-table). The baseline JPEG standard restricts the quantization table entries to be integers between 1 and 255 for 8-bit input images where larger values correspond to more quantization. For each coefficient $F(u,v)$, the quantization operation generates a quantizer output index, denoted by $F^q(u,v)$.

The design of the quantization table is the key factor in determining the image quality of a JPEG-compressed image. The quantization table controls how much error is introduced at a given spatial frequency and provides a means of trading image quality for compression ratio. Many systems merely use the JPEG example tables that are provided in the JPEG specification. Table 1 shows the example luminance q-table used in Annex K of the JPEG standard. (There is also an example chrominance q-table provided in the same Annex.)

This example q-table was developed by determining observer threshold response when viewing 720×576 images on a monitor at a viewing distance of six times the screen width [9]. In the ideal scenario, images would be compressed only to threshold levels, i.e., compression errors are just perceptible under specified viewing conditions. However, in practice, it is often impossible to operate at threshold levels as higher compression ratios are required. The common practice is to use a quantization table designed for threshold levels (or often the JPEG example table) and scale the table by a constant multiplicative factor. The results reported in this study use the JPEG example q-tables at various scaling factors to achieve a range of compression ratios. However, it is expected that the reported improvements due to biased reconstruction carry over to any other q-table selection.

Table 1. Example of JPEG luminance quantization table.

	$u=$ 0	1	2	3	4	5	6	7
$v=0$	16	11	10	16	24	40	51	61
1	12	12	14	19	26	58	60	55
2	14	13	16	24	40	57	69	56
3	14	17	22	29	51	87	80	62
4	18	22	37	56	68	109	103	77
5	24	35	55	64	81	104	113	92
6	49	64	78	87	103	121	120	101
7	72	92	95	98	112	100	103	99

As a final stage in JPEG compression, the quantizer indices are Huffman coded to generate the compressed bit stream. The q-table(s) and Huffman table(s) are sent as part of the compressed image header. This information is needed at the decoder to reconstruct the compressed image, and it is sent once for each image (or for a group of images).

At the decoder, the quantization and Huffman tables are parsed from the header prior to decoding. After Huffman decoding of the quantizer indices, it is necessary to map each quantizer output index back to a reconstructed coefficient value, denoted by $F^R(u,v)$, through the process of dequantization. In a standard JPEG decoder, the dequantization process is defined as:

$$F^R(u,v) = F^q(u,v)\,Q(u,v). \tag{2}$$

This dequantization rule, known as midpoint reconstruction, reconstructs a coefficient to the center of the quantization bin. This reconstruction is optimal only if the probability distributions of the DCT coefficients are uniform. However, for a nonuniform distribution, the mean-squared quantization error is minimized if the reconstruction value is chosen as the centroid (mean value) of the portion of the probability distribution enclosed by the bin [10]. It should be noted that the dequantization process does not affect either the compressed file size or the bit stream syntax, so its optimization can improve the reconstructed image quality and PSNR without any impact on the encoder. In the next section, we discuss the modeling of the AC coefficient distribution, and based upon this model, the selection of an optimal reconstruction value.

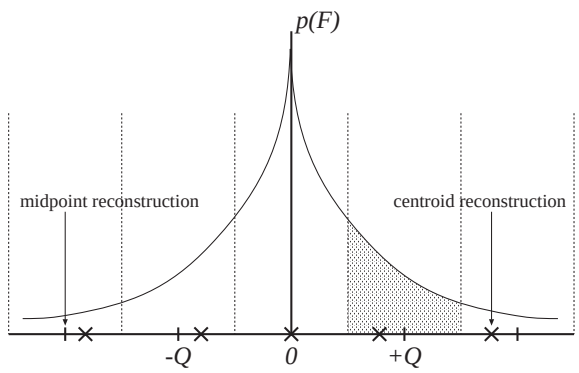

Figure 2. Example of Laplacian distribution with midpoint (hashmark) and centroid (cross) reconstruction.

3. DCT coefficient modeling

Modeling the distributions of the coefficients resulting from the 8×8 DCT of natural images has been studied extensively in the context of JPEG and MPEG compression [11]-[14]. The generalized Gaussian probability density function (GGF) seems to be a good fit for the AC coefficients of most images [13]. The GGF is specified by two parameters: the variance, and the skewness or shape parameter. For a shape parameter of one, the GGF reduces to a Laplacian (double-sided exponential) distribution, while for a shape parameter of two it reduces to a normal (Gaussian) distribution.

Although statistical fitness tests have indicated that the shape parameter depends on the image and the spatial frequency represented by the coefficient [12, 13], the most common assumption for the distribution of the AC DCT coefficients is Laplacian. In this work, we assume the more mathematically tractable Laplacian distribution model for the AC coefficients and derive a maximum likelihood (ML) estimate of the Laplacian parameter based on the quantized coefficients available at the decoder. We then use this estimate to optimally bias the reconstruction levels used during decompression. We justify the Laplacian model by demonstrating that the resulting PSNR improvements are within 0.07 dB of the best theoretically possible (although impractical) strategy of reconstructing to the true bin centroid computed from the unquantized samples available only at the encoder.

The Laplacian distribution, characterized by the single parameter λ, is given by:

$$p(F) = \frac{\lambda}{2}\, e^{-\lambda|F|} \tag{3}$$

An example of Laplacian distribution is shown in Fig. 2. To simplify notation, the derivation presented here pertains to a single 2-D frequency value of u and v, so that the dependence on u and v can be dropped. For a given AC coefficient F, quantized to the bin index F^q, the reconstructed value F^R will be in the bin interval $I_{F^q} = [(F^q - 1/2)Q, (F^q + 1/2)Q]$, where Q indicates the quantization step size. We seek that value of F^R that minimizes the mean-squared quantization error. It is well known that F^R is the centroid of $p(F)$ over I_{F^q} [7, 8, 10, 15] and can be written as:

$$F^R = F^q Q + b \tag{4}$$

where

$$b = -\operatorname{sgn}(F^q)\left[\frac{Q}{2}\left(\frac{1+e^{-\lambda Q}}{1-e^{-\lambda Q}}\right) - \frac{1}{\lambda}\right]. \tag{5}$$

Equation (4) states that F^R is just the bin center, $F^q Q$, plus a bias term b given by (5). The bias term b depends only on the sign of F^q and therefore needs only be computed once for each of the 63 AC coefficients. The $sgn(F^q)$ term in (5) simply ensures that the bias is towards zero.

4. Laplacian parameter estimation

The value of the bias in (4) depends on the Laplacian parameter λ, which needs to be estimated for each DCT coefficient. One approach is to estimate the value of λ on the encoder side prior to coefficient quantization. In general, given a series of N observations of a given coefficient, denoted by F_k, where $k = 1, 2, \ldots, N$ (e.g., $N = 6,144$, for a 768×512 image containing 6,144 8×8 DCT blocks) the maximum likelihood estimate of the Laplacian parameter is given by:

$$\lambda_{ML} = \frac{N}{\sum_{k=1}^{N} |F_k|}, \tag{6}$$

The problem with this approach is that the λ values need to be communicated to the decoder as overhead information, e.g., by being included in the compressed image header.

Since the JPEG syntax does not support this option, a more practical approach is to estimate the λ values at the decoder, based on quantized coefficient values. In what follows, we present a ML estimate of the parameter λ based on quantized observations. Referring to the kth sample of the quantized coefficient (i.e., quantizer index) as F_k^q, we seek λ_{ML}^q. First, we note that quantization effectively transforms the continuous distribution of (3) into the discrete distribution given by:

$$p(F^q) = \int_{(F^q-1/2)Q}^{(F^q+1/2)Q} \frac{\lambda}{2}\, e^{-\lambda|F|}\, dF, \tag{7}$$

Figure 3. The 768×512 monochrome "Boy" image.

where F^q indicates the bin index. Equation (7) leads to:

$$p(F^q) = \begin{cases} \frac{1}{2}\, e^{-\lambda Q(|F^q|-1/2)}\,(1-e^{-\lambda Q}), & \text{for } F^q \neq 0 \\ 1 - e^{-\frac{1}{2}\lambda Q}, & \text{for } F^q = 0. \end{cases} \tag{8}$$

To find λ^q_{ML}, we maximize (over λ) the log-likelihood function of $p(F^q)$ given by:

$$L\big(\lambda; \{F^q_k\}\big) = \ln\Big[\prod_{k=1}^{N} p(F^q_k)\Big] = \sum_{k=1}^{N} \ln[p(F^q_k)] \tag{9}$$

where F^q_k indicates the bin index for the kth observation of a given coefficient. After some tedious manipulation it can be shown that:

$$\lambda^q_{ML} = -\frac{2}{Q}\ln(\gamma) \tag{10}$$

with

$$\gamma = \frac{-N_0 Q}{2NQ+4S} + \frac{\sqrt{N_0^2 Q^2 - (2N_1 Q - 4S)(2NQ+4S)}}{2NQ+4S}, \tag{11}$$

where N_0 is the number of observations that are zero, N_1 is the number of observations that are nonzero, N is the total number of observations ($N = N_0 + N_1$), and

$$S = \sum_{k=1}^{N} Q|F^q_k|. \tag{12}$$

If $S = 0$, which could occur if the JPEG compression ratio is too high, then (12) is not valid. In this case, however, all of the coefficients are in the zero bin where the optimum reconstruction value is zero.

Table 2. Magnitude of the bias b from (5) as a function of 2-D frequency, corresponding to λ_{ML} calculated from (6). Note that u and v range over $0, \ldots 7$ from left-to-right and top-to-bottom, respectively.

0	0.25	0.38	1.39	4.37	12.92	20.43	27.01
0.38	0.64	1.10	2.55	5.52	22.04	25.00	23.97
1.15	1.17	1.99	4.68	12.69	22.30	29.65	24.68
1.74	2.91	4.86	8.32	19.62	38.35	35.94	27.75
4.05	5.94	13.35	23.04	29.35	50.46	48.10	35.78
7.95	13.58	23.58	28.37	36.99	48.82	53.77	43.74
21.56	29.23	36.30	40.77	48.91	58.20	58.05	48.87
34.04	44.03	45.59	47.15	54.20	48.30	50.09	48.29

Table 3. Magnitude of the bias b from (5) as a function of 2-D frequency, corresponding to λ^q_{ML} calculated from (10). Note that u and v range over $0, \ldots 7$ from left-to-right and top-to-bottom, respectively.

0	0.25	0.38	1.36	4.15	11.77	18.45	24.33
0.37	0.63	1.08	2.44	5.12	19.92	22.53	21.46
1.13	1.14	1.91	4.40	11.41	20.07	26.46	22.26
1.68	2.80	4.53	7.57	17.48	34.04	31.91	24.33
3.75	5.44	11.88	20.60	26.48	44.99	42.84	31.96
7.26	12.16	21.09	25.74	33.56	44.03	49.46	39.53
19.48	26.80	34.14	37.79	44.75	59.39	58.89	44.71
31.87	44.89	46.39	47.89	54.89	48.89	50.39	48.39

5. Simulation results

In this section, we first examine the improvements achieved from biased reconstruction on a single example image, and then study the robustness of the results by considering a larger set of 33 test images. As an example, the 768×512 monochrome "Boy" image in Fig. 3 was JPEG compressed using the q-table in Table 1 using a scale factor of 1.0. Tables 2 and 3 denote the bias magnitude tables corresponding to λ_{ML} and λ^q_{ML}, respectively. Recall that the λ_{ML} estimate is computed at the encoder and is not compatible with the JPEG standard as it requires overhead information (which was not accounted for in our simulations) to specify the bias values. These tables were computed by using (6) and (10), where 6,144 samples were available for each AC DCT coefficient. These tables indicate that the bias values resulting from the λ^q_{ML} estimates are slightly more conservative (less bias) than the ones resulting from the λ_{ML} estimates, although in general the differences are fairly small.

Defining the root-mean-squared-error (RMSE) as the standard deviation of the error between the original and the reconstructed (decompressed) image, and the PSNR as $20\log(255/RMSE)$, the performance improvements resulting from biased reconstruction can be quantified. For the "Boy" image, the improvements in PSNR are 0.31 dB for λ_{ML} and 0.35 dB, for λ^q_{ML} . Surprisingly, the ML esti-

Table 4. Average PSNR improvements, in dB, over standard JPEG midpoint reconstruction, for biased, true bin centroid, and fixed percentage reconstructions.

	Quantization Table Scaling			
Dequantization Type	0.50	0.75	1.0	2.0
Midpoint (standard)	0.00	0.00	0.00	0.00
Biased, λ_{ML}, Eq. (6)	0.30	0.27	0.25	0.20
Biased, λ^q_{ML}, Eq. (10)	0.35	0.32	0.30	0.24
Biased, λ from [7]	0.35	0.31	0.29	0.24
True bin centroid	0.42	0.38	0.36	0.30
Fixed percentage	0.20	0.26	0.28	0.26

Table 5. Magnitude of the bias b from (5) as a function of 2-D frequency, corresponding to λ^q_{ML} and expressed as a percentage of the quantizer bin width. A bias of 50% implies reconstruction to the lower end of the bin (towards zero) as opposed to midpoint. Note that u and v range over $0, \ldots 7$ from left-to-right and top-to-bottom, respectively.

0	2.96	6.10	8.78	13.65	20.00	31.79	38.72
2.91	5.87	8.47	13.20	18.76	27.60	37.38	42.75
4.91	9.23	12.31	18.05	26.85	34.00	40.17	43.81
10.14	14.72	19.28	23.67	33.40	37.23	41.84	44.76
17.61	21.36	28.26	32.68	37.16	40.57	43.97	45.76
26.88	33.77	34.73	39.73	42.74	43.71	46.01	46.00
33.24	37.17	39.25	41.23	43.97	45.40	46.95	46.93
38.47	39.64	40.38	41.81	44.09	45.69	47.10	47.44

mate based on the quantized values resulted in a better quantitative performance than the ML estimate based on the unquantized coefficients. This seems to hold true, in general, as will be seen shortly from the results on a larger set of images and a wider range of q-table scales .

Next, a test set of 33 monochrome images (five were 512×512 while the rest were 768×512) were JPEG compressed using the q-table in Table 1 and scaled by four different factors: 0.50, 0.75, 1.0 and 2.0, respectively (for a total of 132 different compressed images). These compressed images were subsequently decompressed using standard JPEG bin center (midpoint) dequantization, biased dequantization using λ_{ML}, and our proposed biased dequantization based on λ^q_{ML}. A less rigorous estimate of λ as suggested in [7] was also tested (see [15] for more details). The ML estimates and the estimate from [7] were derived separately for each image and each compression ratio.

A best case, albeit impractical, scenario was also constructed. Prior to quantization on the encoder side, the true centroid of each bin (the average value of all the coefficients in that bin) for each coefficient was computed and stored. At the decoder, each coefficient in a given bin was reconstructed to the true centroid for that bin. This method is impractical because of both the large overhead (which has been ignored in our simulations) and incompatibility with the JPEG standard. It does, however, provide us with the best possible PSNR improvement for the sake of comparison. The results are summarized in Table 4. All PSNR improvements have been stated with respect to the JPEG bin center dequantization.

The quantities given in Table 4 are the average PSNR improvements over the 33 test images for the indicated quantization table scaling. As is evident from Table 4, biased reconstruction provides modest improvements in PSNR when compared to the bin center reconstruction. There were no individual cases where biased reconstruction caused a relative loss in PSNR. It is well known, however, that PSNR improvements do not necessarily always result in subjective improvements in image quality. Careful analysis of the test images indicated that biased reconstruction produced some subjective improvement; mostly in the form of reducing some mild edge ringing artifacts. Generally, however, the differences between the standard JPEG decoded images and the biased reconstruction images were difficult to detect. Also, the results in Table 4 indicate that estimating the Laplacian parameters from the quantized coefficients actually performs better than estimating them from the unquantized coefficients. Although we do not have a good analytical explanation for this empirical observation, it certainly strengthens the case for using the decoded values for parameter estimation without disturbing the encoder syntax. We additionally note that the less rigorous estimate for λ suggested in [7] performs just as well as our ML estimate, λ^q_{ML}.

Finally, it should be noted that the optimal reconstruction method, using the true bin centroid, is not significantly better than any of the biased reconstructions, and its PSNR improvement is within 0.07 dB of our proposed scheme. This justifies the use of the Laplacian model in our formulation, as it performs almost as well as the best possible and requires little computation when compared to the generalized Gaussian.

The results presented so far are based on using biased values that are specific to each individual image. In most practical situations, it is desirable to add a biased reconstruction functionality to the JPEG decoder at no extra computational cost. This requires the use of a fixed set of representative bias values that are independent of compression ratio or image characteristics. To study the feasibility of such an approach, for each image in the test set and for each scale (132 cases total), the optimum bias values were computed using (10). The bias values were then computed as a percentage of bin width and averaged to create the matrix shown in Table 5. This matrix was used to decompress all the images

in the test set. The last row of Table 4 shows the resulting PSNR improvements. The results are quite encouraging, especially for scale factors of 0.75 or larger, as the improvements are quite close to the ones possible at a higher computational cost. We noted earlier that using image-dependent biased reconstruction improved the results in all cases. In the fixed percentage case, however, biased reconstruction resulted in a PSNR loss in 4 out of the 132 cases. In fact, the well known "Barbara" image (which contains very high frequencies) was responsible for three out of the four cases as it displayed a PSNR loss at quantizer scales of 0.50, 0.75, and 1.0. The average loss for the four cases was -0.13 dB. No images exhibited PSNR loss at scale 2.0.

Note that the elements of the matrix in Table 5 represent the bias as a percentage of the quantization bin width and, as such can be used with any q-table specification. However, there is a slight asymmetry in the matrix that is reminiscent of the asymmetry present in the original JPEG q-table specification. It might be more appropriate to construct a different bias matrix for each q-table specification and scale factor, but that possibility was not explored in our study as the gains are not expected to be large.

6. Conclusions

Assuming a Laplacian distribution for the unquantized AC coefficients, the ML estimate of the Laplacian parameter was derived using only the quantized coefficients available to the decoder. Experiments indicate that biased reconstruction with this estimate gives modest improvements in the PSNR of JPEG decompressed images and that these improvements are close to the best possible by reconstructing to the true bin centroid. It was also shown that by using a fixed (as a percentage of bin width) bias matrix, PSNR improvements on the order of 0.25 dB can be achieved without any increase in decoder computational complexity.

7. References

[1] W. B. Pennebaker and J. L. Mitchell, *JPEG Still Image Data Compression Standard*, Van Nostrand Reinhold, New York, 1993.

[2] M. Crouse and K. Ramchandran, "Joint thresholding and quantizer selection for transform image coding: entropy-constrained analysis and applications to baseline JPEG," *IEEE Trans. on Image Processing* **6**, pp. 285–297, 1997.

[3] V. Ratnakar and M. Livny, "Extending RD-OPT with global thresholding for jpeg optimization," in *Proceedings of the Data Compression Conference*, pp. 379–386, 1996.

[4] V. Ratnakar and M. Livny, "RD-OPT: an efficient algorithm for optimizing DCT quantization tables," in *Proceedings of the Data Compression Conference*, pp. 332–341, 1995.

[5] K. Ramchandran and M. Vetterli, "Rate-distortion optimal fast thresholding with complete JPEG/MPEG decoder compatibility," *IEEE Trans. on Image Processing* **3**, pp. 700–704, 1994.

[6] M.-Y. Shen and C.-C. J. Kuo, "Review of postprocessing techniques for compression artifact removal," *Journal of Visual Communication an Image Representation* **9**(1), pp. 2–14, 1998.

[7] R. L. de Queiroz, "Processing JPEG-compressed images and documents," *IEEE Trans. on Image Processing* **7**(12), pp. 1661–1672, 1998.

[8] G. Lakhani, "Adjustments for JPEG de-quantization coefficients," in *Proceedings of the 1998 Data Compression Conference*, 1998.

[9] H. Lohscheller, "Subjectively adapted image communication system," *IEEE Trans. on Communications* **32**(12), pp. 1316–1322, 1984.

[10] S. P. Lloyd, "Least squares quantization in PCM," *IEEE Trans. on Information Theory* **28**, pp. 129–137, 1982.

[11] S. R. Smoot, "Study of DCT coefficient distributions," in *Human Vision and Electronic Imaging*, vol. 2657 (SPIE), pp. 403–411, 1996.

[12] G. S. Yovanof and S. Liu, "Statistical analysis of the DCT coefficients and their quantization error," in *Proceedings of the 30th Asilomar Conference on Signals, Systems, and Computers*, pp. 601–605, 1996.

[13] F. Müller, "Distribution shape of two-dimensional DCT coefficients of natural images," *Electronics Letters* **29**(22), pp. 1935–1936, 1993.

[14] R. C. Reininger and J. D. Gibson, "Distributions of two-dimensional DCT coefficients for images," *IEEE Trans. on Communications* **31**(6), pp. 835–839, 1983.

[15] J. R. Price and M. Rabbani, "Biased reconstruction for JPEG decoding," *IEEE Signal Processing Letters* **6**(12), pp. 297–299, 1999.