

A REQUANTIZATION-BASED METHOD FOR RECOMPRESSING JPEG IMAGES

H.H. Bauschke,¹ C.H. Hamilton,² M.S. Macklem,³ J.S. McMichael,⁴ and N.R. Swart⁵

¹Department of Mathematics and Statistics, University of Guelph, Guelph, Ontario N1G 2W1, Canada

²Packeteer Inc, #800–1708 Dolphin Avenue, Kelowna, B.C. V1Y 9S4, Canada

³Department of Mathematics, Simon Fraser University, Burnaby, B.C. V5A 1S6, Canada

⁴Packeteer Inc, #800–1708 Dolphin Avenue, Kelowna, B.C. V1Y 9S4, Canada

⁵Analog Devices, 2081 Lillooet Court, Kelowna, B.C. V1V 1Y2, Canada

ABSTRACT

We propose a new method for requantizing JPEG images, based on the well-known Laplacian distribution of the AC discrete cosine transform (DCT) coefficients and on an analysis of the error introduced by requantization. The resulting images have improved perceptual image quality over a “blind” requantization approach, that is, one that does not consider the properties of the quantization matrices.

1. INTRODUCTION

The de facto standard for dealing with photorealistic images is JPEG [8, 3]. There is a tremendous need for on-the-fly recompression of JPEG images to suit, for instance, low bandwidth clients (modem connections, PDAs, and cell phones) on the internet. Ideally, since JPEG is a lossy compression scheme, one would like to work from the original image. In practice, however, the original image may no longer exist or it may be too difficult to retrieve in real time.

Requantizing an already quantized image can lead to apparently unpredictable behavior and unwanted artifacts. The following experiment — carried out using freely available software provided by the Independent JPEG Group (IJG) [6] — is typical. We employ IJG’s *quality scale*, which ranges from Q1 (poor) to Q100 (excellent).

Consider a photorealistic image such as *lena* (available at [4]). Compress *lena* to quality level Q75 (a common quality setting), and denote the resulting image by *lena-75*. Now recompress *lena-75* (*not* the original *lena*) further to Q50 and to Q48. Denote the resulting images by *lena-75-50* (see Fig. 1) and *lena-75-48* (see Fig. 2). The outcome is remarkable: *lena-75-50* is 19% smaller than *lena-75*, but has considerable “grainy” artifacts; however, *lena-75-48* is 37% smaller, while exhibiting “smoother”, less visually annoying artifacts!

It is very striking that one can repeat this experiment for *any* photorealistic image, with the same outcome. Hence the IJG quality rating scale is not perceptually monotone.



Fig. 1. *lena-75-50*



Fig. 2. *lena-75-48*

Chan also observed this in [5], for three other test images.

In this paper, we provide an explanation of this experiment, and present a new heuristic algorithm that requantizes any JPEG image so that the “grainy” artifacts observed in the previous example are avoided — see [2] for a more detailed analysis and discussion of the phenomena and algorithms outlined here. Our approach can also be applied to any DCT-based transform method, such as MPEG [3]. The paper is organized as follows. Section 2 reviews the Laplacian distribution of the AC DCT coefficients. The effect of requantizing an integer in two steps is studied in Section 3. In Section 4, we consider the error introduced when dequantizing a Laplacian distribution twice; this is applied to JPEG images in Section 5. Section 6 describes the new heuristic algorithm and its performance.

2. DISTRIBUTION OF JPEG DCT COEFFICIENTS

2.1. Brief review of the steps in JPEG coding

(See [8, 3] for details.) **(1)** The image is separated into three color components. **(2)** Each component is partitioned into 8-by-8 blocks. **(3)** Each block is transformed using the two-dimensional Discrete Cosine Transform (DCT). **(4)** Each transformed block is quantized with respect to an 8-by-8 quantization matrix. **(5)** The resulting data is compressed, using Huffman or arithmetic coding.

2.2. Continuous Laplacian distribution

We collect the transformed coefficients for each color component by frequency. There are 64 frequencies, and it is known [9, 7, 10, 11] that the histogram of coefficients corresponding to any one of the 63 AC frequencies resembles a *Laplacian distribution*, $f(x) = (\lambda/2) \exp(-\lambda|x|)$, with parameter λ and zero mean. Let X be a random variable with probability density function f . The variance of X is $\text{Var}(X) = 2/\lambda^2$.

2.3. Discrete, dequantized Laplacian distribution

Assume the original image is partitioned into J 8-by-8 blocks, and y_1, y_2, \dots, y_J are the *rounded* DCT coefficients for a fixed AC frequency and a fixed color component. (This corresponds to Q100 on the IJG quality scale.) We view y_1, y_2, \dots, y_J as a random sample of size J taken by the discrete random variable $Y = \text{round}(X)$.

Proposition 1 [2] *Let p be the probability mass function associated with the discrete random variable Y . Then*

$$p(n) = \begin{cases} 1 - \exp(-\lambda/2), & \text{if } n = 0; \\ \exp(-\lambda|n|) \sinh(\lambda/2), & \text{if } n \in \mathbb{Z} \setminus \{0\}. \end{cases}$$

2.4. Estimation of the Laplacian parameter λ

The Laplacian parameter λ for the rounded DCT coefficients y_1, y_2, \dots, y_J can be estimated very well by the *Maximum Likelihood estimator* λ_{ML} :

Proposition 2 (Method of Maximum Likelihood) [2]

Let $\alpha = \#\{j : y_j = 0\}$, $\beta = J - \alpha$, and $\rho = 2 \sum_j |y_j|$. If $\beta = 0$, set $\lambda_{\text{ML}} = +\infty$; otherwise,

$$\lambda_{\text{ML}} = -2 \ln \left(\frac{-\alpha + \sqrt{\alpha^2 - 4\alpha\beta - 4\beta^2 + 4\rho^2 + 4\alpha\rho}}{2(J + \rho)} \right).$$

3. DEQUANTIZATION: DIRECT VS DETOUR

Suppose q_0 and q_1 belong to \mathbb{N}^* with $q_0 < q_1$. Let x be in \mathbb{N}^* . Compare the *direct dequantization* of x with respect to q_1 , i.e., $x \mapsto q_1 \cdot \text{round}(x/q_1)$, to the *two-step dequantization* of x with respect to q_1 , preceding a “*detour*” via q_0 : $x \mapsto q_1 \cdot \text{round}(q_0 \cdot \text{round}(x/q_0)/q_1)$.

We think of q_0 and q_1 as entries of the quantization tables of the old and new requantized images, respectively, both quantizing the same AC entry. Numerical experiments show that the difference between the two dequantizations cannot be described easily except for certain cases, which we summarize now.

Proposition 3 (Direct vs Detour) [2]

Let $x_1 = q_1 \cdot \text{round}(x/q_1)$, $y_0 = q_0 \cdot \text{round}(x/q_0)$, and $y_1 = q_1 \cdot \text{round}(y_0/q_1)$. Let $l = \text{lcm}(q_0, q_1)$. Then:

- (i) $x_1 - y_1 \in \{-q_1, 0, +q_1\}$.
- (ii) If $\tilde{x} = x + l$, then $\tilde{x}_1 = x_1 + l$ and $\tilde{y}_1 = y_1 + l$.
- (iii) If $m \in \mathbb{N}^*$ is odd and $q_1 = mq_0$, then $x_1 = y_1$.
- (iv) Let $m \in \mathbb{N}^*$ be even, $q_1 = mq_0$, and $1 \leq x \leq l$. Then either $y_1 = x_1 + q_1$ if

$$\frac{q_1}{2} - \left\lfloor \frac{q_0}{2} \right\rfloor \leq x \leq \frac{q_1}{2} - 1;$$

or $y_1 = x_1$, otherwise.

- (v) Let $m \in \mathbb{N}^*$ be even, $q_1 = 1 + mq_0$, and $1 \leq x \leq l$. Then $y_1 = x_1$, if $x < (q_1 + 1)/2$; and $y_1 = x_1 - q_1$, if

$$\frac{q_1 + 1}{2} \leq x \leq \frac{q_1 - 3}{2} + \left\lceil \frac{q_0}{2} \right\rceil.$$

4. LAPLACIAN DEQUANTIZATION ERRORS

4.1. Mathematical description

For a fixed Laplacian parameter $\lambda > 0$, and $q_0, q_1 \in \mathbb{N}^*$, the error of dequantizing $n \in \mathbb{Z}$ in two steps is

$$e_n = q_1 \cdot \text{round}(q_0 \cdot \text{round}(n/q_0)/q_1) - n.$$

Because of the dequantization convention in JPEG [6], we restrict our attention to $n \geq 0$. Using Proposition 1, the (weighted absolute) error made by dequantizing the given Laplacian in two steps is

$$e = 2 \sum_{n \in \mathbb{N}^*} p(n) |e_n|.$$

As in [5], we break this error up into two components: $e = e_+ + e_-$, where e_+ and e_- are both nonnegative and measure *increase* and *decrease* in amplitude, respectively:

$$e_+ = 2 \sum_{n \in \mathbb{N}^*: e_n > 0} p(n) e_n, \quad e_- = 2 \sum_{n \in \mathbb{N}^*: e_n < 0} p(n) |e_n|.$$

In the context of JPEG images, the errors can be interpreted as follows: *If the component e_+ (resp. e_-) dominates, then the amplitude increases (resp. decreases) and we thus expect the image dequantized in two steps to be grainier (resp. smoother) than the original.*

Assuming fixed λ and q_0 , one would expect that these errors are monotone with q_1 . Surprisingly, this is not the case; hence we aim for *local minimizers*. As this appears to be intractable analytically, we reason numerically.

4.2. Numerical approach: determining parameters

Using the Method of Maximum Likelihood (Proposition 2) for a collection of standard test images, we estimated the Laplacian parameter λ for each DCT AC coefficient. The resulting statistics guided us in choosing a sensible range for λ , namely $0.01 \leq \lambda \leq 2$. Furthermore, the JPEG image format implicitly constrains the range of values a DCT coefficient can take: in fact, the magnitude of any AC DCT coefficient cannot be larger than 1020 [2].

4.3. Numerical results

We considered various values of λ in $[0.01, 2]$ for the corresponding Laplacian distribution with frequency $p(n)$ (see Proposition 1). We then truncated the distribution at 1020 and renormalized. The errors of Section 4.1 were computed numerically, for all q_0, q_1 in \mathbb{N}^* with $1 \leq q_0 \leq q_1 \leq 255$ (corresponding to the JPEG baseline standard), and collected in upper triangular 255-by-255 matrices, where the (q_0, q_1) entry is the error corresponding to the fixed λ . The resulting matrices of errors e_+, e_-, e prompted us to postulate Observation 4 below. Referring the reader to [2] for further details, we include a visualization of the matrix e_+ in Fig. 3, which illustrates Observation 4.(ii).

Observation 4 *Suppose λ and q_0 are fixed. Consider the errors e, e_+, e_- as functions of q_1 . Then:*

- (i) *The total error e is monotone increasing, with the exception of “dips” around points of the form $q_1 = mq_0 + 1$, where $m \in \mathbb{N}^*$ is even.*

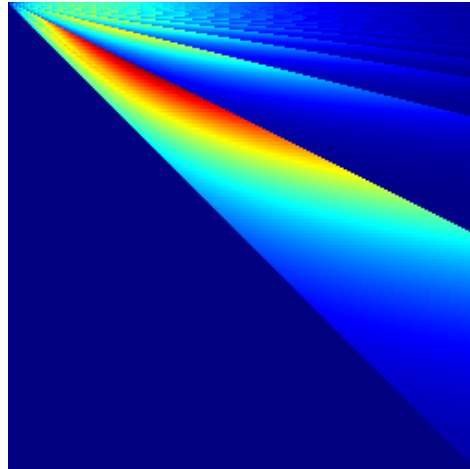


Fig. 3. Error matrix e_+ for $\lambda \approx 0.07$.

- (ii) *The (amplitude-intensifying) error e_+ is overall decreasing to 0. More importantly, it is increasing on intervals $[mq_0 + 1, (m+2)q_0]$, where $m \in \mathbb{N}^*$ is even, and there is a sharp drop between mq_0 and $mq_0 + 1$.*
- (iii) *In contrast, the (amplitude-reducing) error e_- is overall increasing and approaching e . More significantly, it is nearly constant on intervals $[mq_0 + 1, (m+2)q_0]$, where $m \in \mathbb{N}^*$ is even, with a sharp rise between mq_0 and $mq_0 + 1$.*

Note that the sharp changes of e_+ (resp. e_-) in (ii) (resp. (iii)) of Observation 4 parallel the errors in item (iv) (resp. item (v)) of Proposition 3.

5. REQUANTIZATION ERROR FOR JPEG IMAGES

Observation 4 has the following consequences for recompressing JPEG images by requantization.

Observation 5 The key quantity is the quotient q_1/q_0 :

- there is a sharp divide in the type of error for values of q_1 in the vicinity of an even multiple of q_0 , say $q_1 \approx mq_0$, where $m \in \mathbb{N}^*$ is even;
- recompressing with $q_1 = mq_0$ or less is amplitude intensifying and thus results in grainier images;
- recompressing with $q_1 = mq_0 + 1$ or more is amplitude reducing and hence yields smoother images.

Applying Observation 5 to the discussion of *lena* from the Introduction, we see that the crucial quantity is the *element-wise quotient* of the new quantization matrix and Q75. It turns out that all entries of Q50./Q75 and of Q48./Q75 are very close to 2. Specifically:

$$Q_{50}/Q_{75} = \begin{pmatrix} 2 & 2^- & 2 & 2 & 2 & 2 & 2^- & 2^- \\ 2 & 2 & 2 & 2^- & 2 & 2 & 2 & 2^- \\ 2 & 2^- & 2 & 2 & 2 & 2^- & 2^- & 2 \\ 2 & 2^- & 2 & 2^- & 2^- & 2^- & 2 & 2 \\ 2 & 2 & 2^- & 2 & 2 & 2^- & 2^- & 2^- \\ 2 & 2^- & 2^- & 2 & 2^- & 2 & 2^- & 2 \\ 2^- & 2 & 2 & 2^- & 2^- & 2^- & 2 & 2^- \\ 2 & 2 & 2^- & 2 & 2 & 2 & 2^- & 2^- \end{pmatrix},$$

whereas

$$Q_{48}/Q_{75} = \begin{pmatrix} 2^+ & 2^- & 2 & 2^+ & 2^+ & 2^+ & 2^+ & 2^+ \\ 2 & 2 & 2^+ & 2 & 2^+ & 2^+ & 2^+ & 2^+ \\ 2^+ & 2 & 2^+ & 2^+ & 2^+ & 2^+ & 2^+ & 2^+ \\ 2^+ & 2 & 2^+ & 2 & 2^+ & 2^+ & 2^+ & 2^+ \\ 2^+ & 2^+ & 2 & 2^+ & 2^+ & 2^+ & 2^+ & 2^+ \\ 2^+ & 2 & 2^+ & 2^+ & 2^+ & 2^+ & 2^+ & 2^+ \\ 2^+ & 2^+ & 2^+ & 2^+ & 2^+ & 2^+ & 2^+ & 2^+ \\ 2^+ & 2^+ & 2^+ & 2^+ & 2^+ & 2^+ & 2^+ & 2^+ \end{pmatrix};$$

here 2^+ (2^- , resp.) stands for a real number that is strictly greater than (equal to, strictly less than, resp.) 2. In view of Observation 5, this explains why `lena-75-50` appears grainier, and `lena-75-48` appears smoother. *Most importantly, we perceive the latter version as closer to the original than the former* [2].

6. NEW REQUANTIZATION ALGORITHM

We conclude with an algorithm for recompressing JPEG images which avoids the aforementioned “grainy” artifacts and seeks “smooth” artifacts instead.

New heuristic method. Suppose we need to recompress a given JPEG image. Assume Q_0 (resp. Q_{bound}) is the original (resp. target) quantization matrix, respectively. Using these two matrices, we wish to find a new quantization matrix Q_1 , in the vicinity of Q_{bound} , in order to requantize the given image. For a fixed frequency, denote the corresponding quantizers by q_0 , q_{bound} , and q_1 , respectively. Let $k = \lfloor q_{\text{bound}}/q_0 \rfloor$. Then set

$$q_1 = \begin{cases} q_0, & \text{if } k = 0; \\ kq_0, & \text{if } k > 0. \end{cases}$$

A new quantization matrix Q_1 is obtained. Finally, requantize the given image with respect to Q_1 but *observe the following modified rounding convention*: round as always, but if you encounter genuine half-integers ($\pm\frac{1}{2}, \pm\frac{3}{2}, \dots$), then round *towards zero*.

Experimental results. We performed subjective tests with 7 test subjects, comparing the heuristic algorithm to a blind requantization, by using the well-known *Mean Opinion Score (MOS) grading scale* [1]. 63 test images were recompressed using blind requantization to Q50, and with the

heuristic algorithm with $Q_{\text{bound}} = Q_{45}$. The quality level Q50 was selected to give additional compression compared to the original versions of each image, and $Q_{\text{bound}} = Q_{45}$ was chosen to match the corresponding file sizes. The experimental results validated our analysis: we observed a *consistent increase in the overall ratings of the heuristic algorithm, relative to blind requantization* [2].

7. REFERENCES

- [1] J. Allnatt, *Transmitted-Picture Assessment*, John Wiley and Sons, 1983.
- [2] H.H. Bauschke, C.H. Hamilton, M.S. Macklem, J.S. McMichael, and N.R. Swart, “Recompression of JPEG images by Requantization,” 2001, submitted.
- [3] A. Bovik (editor), *Handbook of Image and Video Processing*, Academic Press, 2000.
- [4] links.uwaterloo.ca/bragzone.base.html, *standard test images*.
- [5] S. Chan, “Recompression of still images,” Tech. Rep. 2-92*, University of Kent, Canterbury, UK, March 1992, see also www.cs.ukc.ac.uk/pubs/1992/103.
- [6] Independent JPEG Group, *Various documents*, www.ijg.org.
- [7] E.Y. Lam and J.W. Goodman, “A Mathematical Analysis of the DCT Coefficient Distributions for Images,” *IEEE Transactions on Image Processing*, vol. 9, no. 10, pp. 1661–1666, October 2000.
- [8] W.B. Pennebaker and J.L. Mitchell, *JPEG still image data compression standard*, Van Nostrand Reinhold, 1993.
- [9] R.C. Reininger and J.D. Gibson, “Distributions of the Two-Dimensional DCT Coefficients for Images,” *IEEE Transactions on Communications*, vol. COM-31, no. 6, pp. 835–839, June 1983.
- [10] S.R. Smoot and L.A. Rowe, “Laplacian Model for AC DCT Terms in Image and Video coding,” in *Proceedings of the Ninth IEEE Image and Multidimension Digital Signal Processing Workshop*, 1996.
- [11] S.R. Smoot and L.A. Rowe, “Study of DCT Coefficient Distribution,” in *Human Vision and Electronic Imaging*, B. E. Rogowitz and T. N. Pappas, Eds. 1996, vol. 2657, SPIE — The International Society for Optical Engineering.