Study of DCT coefficient distributions

Stephen R. Smoot

Computer Science Division
University of California at Berkeley
Berkeley, CA 94720
smoot@plateau.cs.berkeley.edu; (510) 643-7106

Lawrence A. Rowe

Computer Science Division
University of California at Berkeley
Berkeley, CA 94720
Rowe@cs.berkeley.edu; (510) 642-5117

ABSTRACT

Many image and video compression schemes perform the discrete cosine transform (DCT) to represent image data in frequency space. An analysis of a broad suite of images confirms previous finding that a Laplacian distribution can be used to model the luminance AC coefficients. This model is expanded and applied to color space (Cr/Cb) coefficients. In MPEG, the DCT is used to code interframe prediction error terms. The distribution of these coefficients is explored. Finally, the distribution model is applied to improve dynamic generation of quantization matrices.

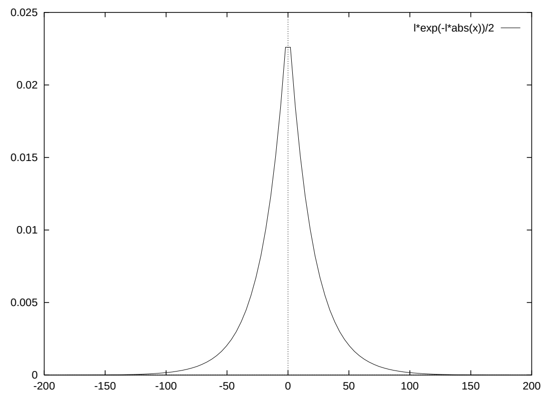
Keywords: DCT, quantization, JPEG, MPEG, video compression

1 INTRODUCTION

Many digital image and video compression schemes use a block-based Discrete Cosine Transform (DCT) as the transform coding step [7]. In particular JPEG [3, 8] and MPEG [4, 5] use the DCT to concentrate image information [6]. This paper presents the results of a study into the properties of images that have been compressed using the DCT.

Image compression systems often divide each image into multiple planes, one for luminance (brightness) and two for color (for example chrominance-red and chrominance-blue). The images are also spatially divided into blocks, usually $8x8$ pixels. The DCT is applied to each block in each plane and the results are quantized and run-length encoded (with additional Huffman or arithmetic coding). This paper considers the properties of images from the middle of this sequence — just after the DCT transform, but before quantization and run-length coding.

It is generally believed that the distribution of the luminance components of a transformed image block



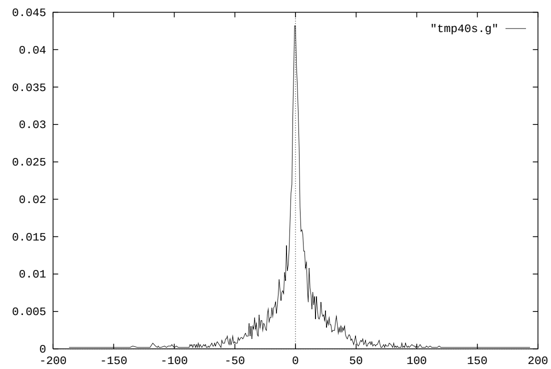


Figure 1: The left plot (a) shows a sample Laplacian distribution, $p(x) = \frac{\lambda}{2}e^{-\lambda|x|}$, with $\lambda = .05$. The histogram on the right (b) is an image distribution with $\lambda = .0500$ (measured) and a bin size of 0.75.

is Laplacian (double sided exponential, see Figure 1a) except for the $[0, 0]$ coefficient [7, 10]. The Laplacian distribution

$$p(x) = \frac{\lambda}{2}e^{-\lambda|x|}$$

is symmetrical about zero, and it can be readily matched to a sample distribution by finding the appropriate parameter (λ).

This paper begins by confirming the Laplacian distribution for the luminance and chrominance AC coefficients (*i.e.*, all coefficients excepting $[0, 0]$). The next Section describes the experimental setup, the video sequences used and the analysis methods. Section 3 describes the results for the luminance coefficients. In Section 4, the results for color coefficients are described. The use of the DCT on the error terms found in P and B MPEG frames is addressed in Section 5. Finally, in Section 7 applications of this research are briefly discussed.

2 EXPERIMENT

The data used in these experiments comes from an evaluation of several MPEG-2 test sequences (fireworks, flowergarden, football, fountain, hockey, marbles, mobile & calendar, popple, tempete, and tennis) provided by AT&T Bell Laboratories. Each sequence is a series of 720x480 color images in progressive format. In several sequences, the active part of the images covers less than the full area, so they were clipped to the active regions for the analysis. Table 1 gives the names, sizes, video lengths, and a brief summary of each sequence.

The first ten images of each clip were put through the UC Berkeley MPEG encoder (mpeg_encode [11, 2]) which has been instrumented to record the exact output of the DCT algorithm. An IEEE compliant floating point implementation of the DCT was employed for these tests. These data files were analyzed to find the appropriate parameter for each distribution.

Fitting the distribution to a particular coefficient's sample requires finding the correct parameter for the Laplacian distribution (λ). The simplest way to find the parameter is to use the standard deviation (1):

$$\sigma^2 = \int_{-\infty}^{\infty} (x - E(x))^2 p(x) dx$$

Name	Size (pixels)	Length (frames)	Description
fireworks	688x480	120	A Japanese woman standing in front of a mostly black background with fireworks going off and Japanese text being colored in (Karoke).
flowergarden	704x480	150	The familiar scene of a windmill with a flower garden seen from a passing car.
football	704x480	150	A famous football play (UC Berkeley v. Stanford) with fumble.
fountain	512x480	150	A computer-graphics fountain.
hockey	624x480	150	Hockey players skating on ice. The camera follows one player, blurring most of the scene.
marbles	688x480	150	Some marbles with Japanese text being filled in (Karoke). The camera pans over the graphics.
mobilcal	720x480	150	A mobile and calendar with panning and zooming.
popple	720x480	150	A saturated-color scene of stuffed animals turning, and zooming.
tempete	704x480	149	Desert scene with leaves blowing, and a zoom out.
tennis	720x480	150	Table tennis (ping-pong) being played. Fast action with cuts.

Table 1: AT&T image sequence summary information.

$$\begin{aligned}
&= \int_{-\infty}^{\infty} x^2 \frac{\lambda}{2} e^{-\lambda|x|} dx \\
&= \frac{1}{2}[e^{\lambda x}(x^2 - \frac{2x}{\lambda} + \frac{2}{\lambda^2})\Big|_{-\infty}^{0} + e^{-\lambda x}(x^2 + \frac{2x}{\lambda} + \frac{2}{\lambda^2})\Big|_{0}^{\infty}] \\
&= \frac{1}{2}(\frac{2}{\lambda^2} + \frac{2}{\lambda^2}) \\
&= \frac{2}{\lambda^2} \\
\lambda &= \frac{\sqrt{2}}{\sigma} \qquad (1)
\end{aligned}$$

3 AC LUMINANCE COEFFICIENTS

The first set of experiments was performed to replicate the original findings [10]. Reininger and Gibson examined individual greyscale images. The parameters for the luminance coefficient of the first frame of the flowergarden sequence are shown in Figure 2 both in tabular form (a) and a more illustrative graph (b), where the DCT coefficients are laid out on the xy plane, and the Laplacian parameters go up the z-axis.

Examining this data, the trends are as expected. The higher order coefficients have larger Laplacian parameters, indicating a stronger peak at zero (more zeros or generally smaller values). The data for the flowergarden sequence is typical for most of the sequences, as seen in Figure 3. However, there are several differences between the different sequences: the hockey sequence has the characteristic shape, but has much higher values, as the featureless ice and blurred motion in the scene gives a large number of zeros for the AC coefficients. Tennis is odd and we have no idea why. The remaining sequences have the same basic pattern with increasing parameters for larger coefficients and more non-zero values in the vertical frequencies than the horizontal (possibly due to smoothing in the camera during pans).

	0	1	2	3	4	5	6	7
0	-	.046	.080	.127	.191	.268	.308	.423
1	.031	.050	.085	.140	.212	.298	.358	.505
2	.036	.054	.086	.144	.228	.334	.430	.632
3	.039	.058	.089	.140	.214	.339	.486	.708
4	.045	.067	.095	.142	.224	.351	.491	.731
5	.054	.063	.082	.120	.191	.290	.406	.631
6	.062	.063	.079	.115	.173	.239	.302	.476
7	.048	.054	.072	.105	.147	.190	.226	.344

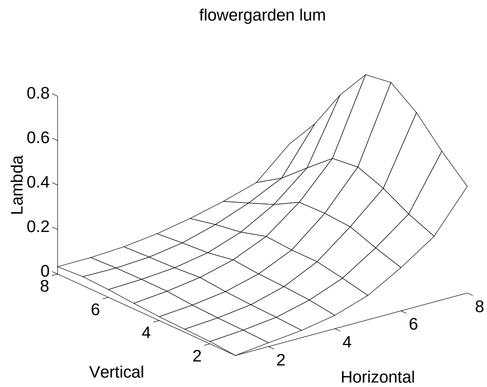


Figure 2: Laplacian distribution parameters for the luminance AC coefficients in the first ten frames of the flowergarden sequence. The graph on the right (b) presents the data from the table on the left (a).

The image distributions are Laplacian, according to the Kolmogorov test [1]. Hypothesis tests were performed using the distribution data and the derived distribution parameters. The Kolmogorov test works by comparing the order statistics of the sample distribution with the theoretical distribution, and testing the largest difference against a theoretically determined difference. In all cases, the largest deviation is near the zero value[1]. The deviations were usually not significant, thus the null hypothesis (that the distributions are the same) could not be rejected at the 5% level. This result indicates that while the Laplacian may not be the absolute best fit for the data, it still fits very well.

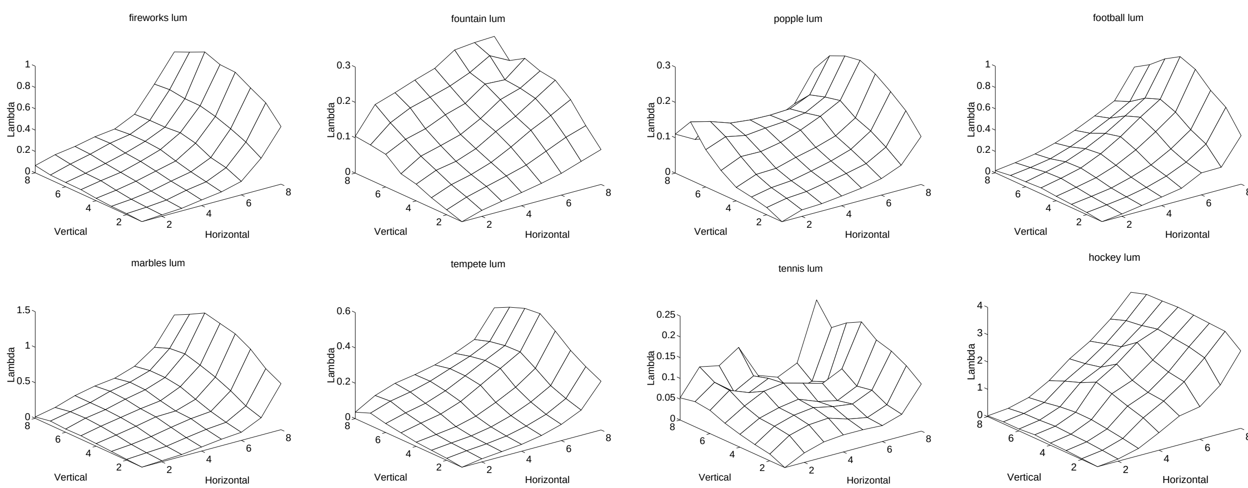


Figure 3: The Laplacian parameters for Luminance channel in the various sequences.

[1] Because the Laplacian distribution over-predicts until very near zero, and then under-predicts until the symmetrical point around zero, the best match of the distribution may be the sum of multiple Laplacians with increasing parameter values. However, this level of detail does not seem to offer any benefits over the simple single Laplacian.

4 AC CHROMINANCE COEFFICIENTS

As mentioned above, the initial work on this problem did not consider the color aspects of images [10]. The chrominance data from the MPEG-2 sequences exhibits the same distribution as for the luminance channel. However the parameters are larger in general, indicating more near-zero values. The parameters for chrominance-red and chrominance-blue are nearly identical, so only those for chrominance-blue are presented in Figures 4 and 5.

	0	1	2	3	4	5	6	7
0	-	.459	.573	.653	.797	.888	1.034	1.085
1	.327	.573	.641	.800	.881	.995	1.022	1.260
2	.360	.590	.755	.845	.940	.971	1.095	1.293
3	.415	.573	.734	.826	.940	.994	1.081	1.271
4	.437	.601	.719	.839	.941	1.011	1.070	1.331
5	.454	.582	.724	.825	.920	.995	1.090	1.321
6	.440	.560	.695	.815	.909	.969	1.028	1.362
7	.413	.519	.646	.789	.890	.945	1.048	1.366

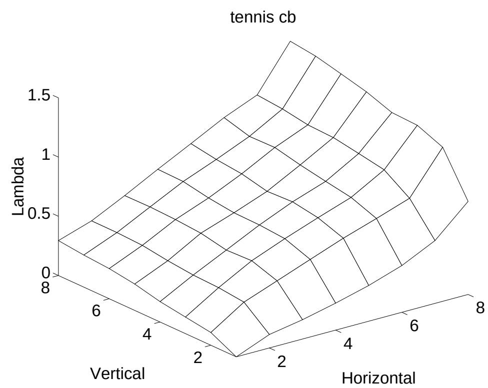


Figure 4: Chrominance-blue for tennis in both tabular and plotted forms.

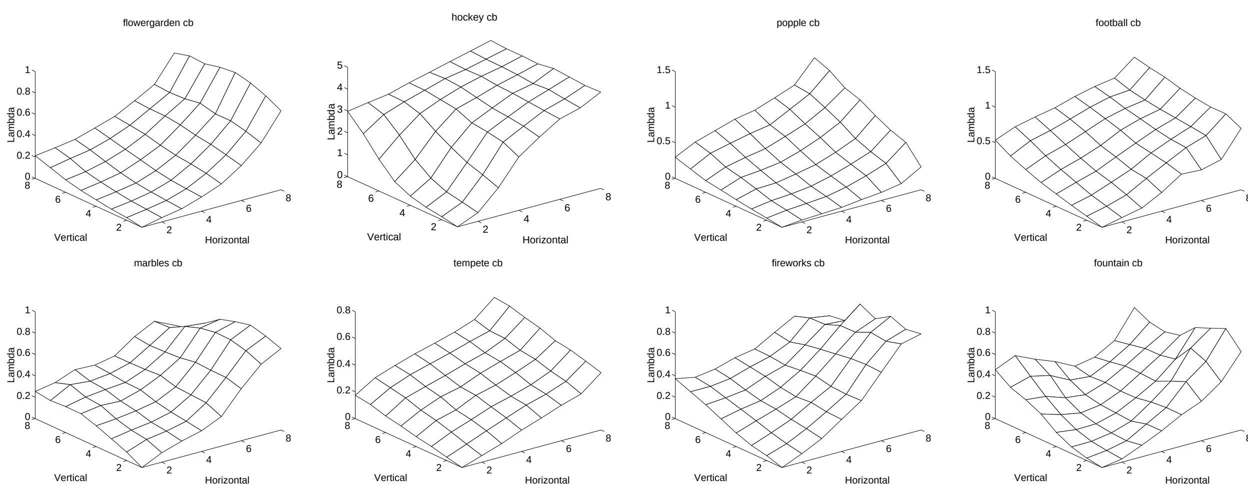


Figure 5: The Laplacian parameters for chrominance-blue channel in the various sequences.

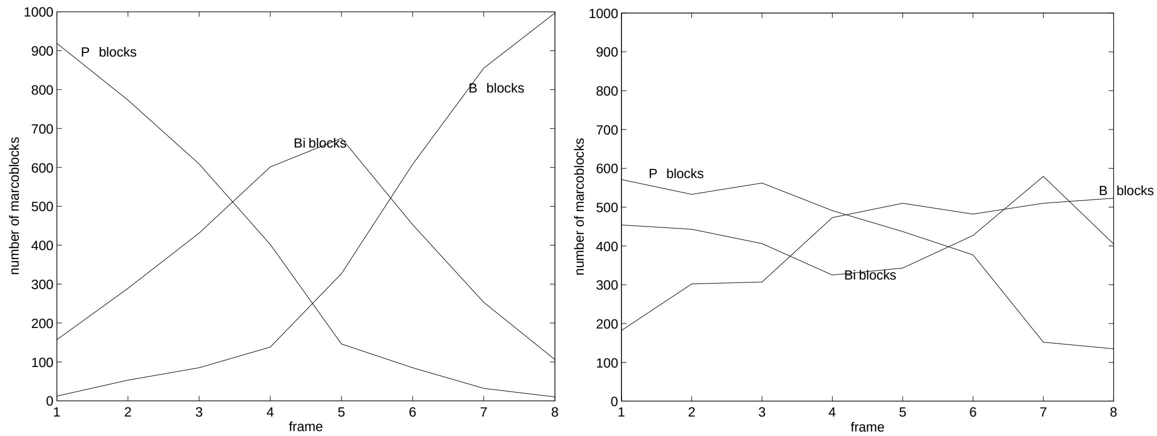


Figure 6: The numbers of forward-predicted (P-blocks), bi-directionally averaged (Bi-blocks), and backward-predicted (B-blocks) blocks in two sequences. The flowergarden sequence (a) has a simple pan that provides a textbook distribution over the course of eight B frames. The fireworks sequence (b) has a more random motion pattern, and the distribution is not as clear.

5 DCT FOR ERROR TERMS

MPEG encoding does not use the DCT solely for picture information; it is also employed to code error terms. Error terms are formed as part of MPEG's motion compensation algorithm. A simple description is that an error term is formed by subtracting the image block from a block on another picture in the sequence, and applying the DCT to the difference. This way the picture can be coded using very few bits if there are only small (or no) changes in the image. The statistical properties of error blocks have not been studied previously, but it seemed likely that they have a Laplacian distribution as well. To investigate this conjecture, the sequences were encoded using motion compensation and the data for the compensated frames was extracted and analyzed.

The first set of data comes from the sequence "IBBBBBBBBI," with the values for the B frames extracted. Thus, it captures the information for a wide range of motion distances. The sequence also varies from mostly forward vectors to mostly averaged vectors to mostly backward vectors, as seen in Figure 6. The results appear in Figure 7. While the Laplacian distribution is still in effect, according to the Kolmogorov tests, the characteristic values are different. In general, the error terms are more balanced across the different coefficients, with the parameters for horizontal and vertical coefficients being more similar than those for the image distributions (*i.e.*, they show more symmetry across the $x = y$ coefficient line). In addition, other than the hockey sequence, the parameters are all in the same general range (a maximum of 0.8–1 excepting hockey), showing more similarity across sequences than the distributions for images did.

The second set of data was encoded with the stream sequence "IPIPIPIPIP" for each sequence over frames 0–9, and then over frames 1–8, thus capturing information on single-frame forward motion for all but the first frame. The DCT coefficients from the P frames were extracted and analyzed. These results, however, do not differ significantly from the B frame case and are not presented here.

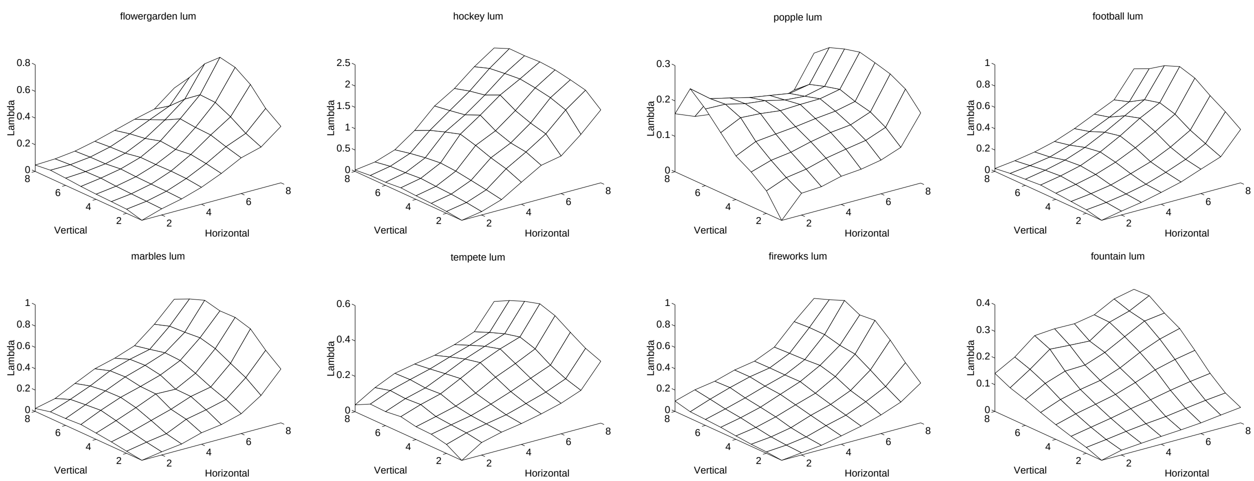


Figure 7: The Laplacian parameters for the luminance channel error blocks in the various sequences.

6 STABILITY

An interesting question is the stability of these parameters. While they clearly can describe a single image, it is unclear that the values will be stable over successive frames. This fact is important for MPEG-1 coding as it is limited to a new quantization matrix for every Group-of-Pictures[2]. To investigate this question, we MPEG-encoded each fourteenth frame of the mobile & calendar sequence, creating a ten frame test sequence. The data for the luminance components are displayed in Figure 8 (next page). Surprisingly, the distribution is stable over the entire sequence.

7 USING THE DISTRIBUTION INFORMATION

Of great interest in image coding is efficient quantization matrices. In MPEG, for example, one can use the distributions measured in previous pictures to generate new quantization matrices for the next Group-of-Pictures to improve both bit-rate control and decoded image quality. One simple method is to gather information about the coefficient distributions (*e.g.*,Laplacian distribution parameters) and use it to estimate the error for the different possible quantizer values. Combining this data with a scale of the importance of the various coefficients produces a base matrix for the new group.

The process begins by determining which quantization factor (Q) results in the smallest mean squared error (MSE). As an example, consider quantizing values with a dead zone around zero:

$$\begin{aligned} MSE &= \min_Q 2\left(\sum_{k=1}^{k=M}\int_{kQ-\frac{Q}{2}}^{kQ+\frac{Q}{2}}(x-kQ)^2p(x)dx + \int_{MQ+\frac{Q}{2}}^{\infty}(x-MQ)^2p(x)dx\right) \\ &= \min_Q 2\left(\sum_{k=1}^{k=M}\int_{kQ-\frac{Q}{2}}^{kQ+\frac{Q}{2}}(x^2-2kQx+k^2Q^2)\frac{\lambda}{2}e^{-\lambda x}dx + \int_{MQ+\frac{Q}{2}}^{\infty}(x-MQ)^2\frac{\lambda}{2}e^{-\lambda x}dx\right) \end{aligned}$$

[2] In MPEG-2 the quantization tables can be resupplied at each frame, giving the encoder more options.

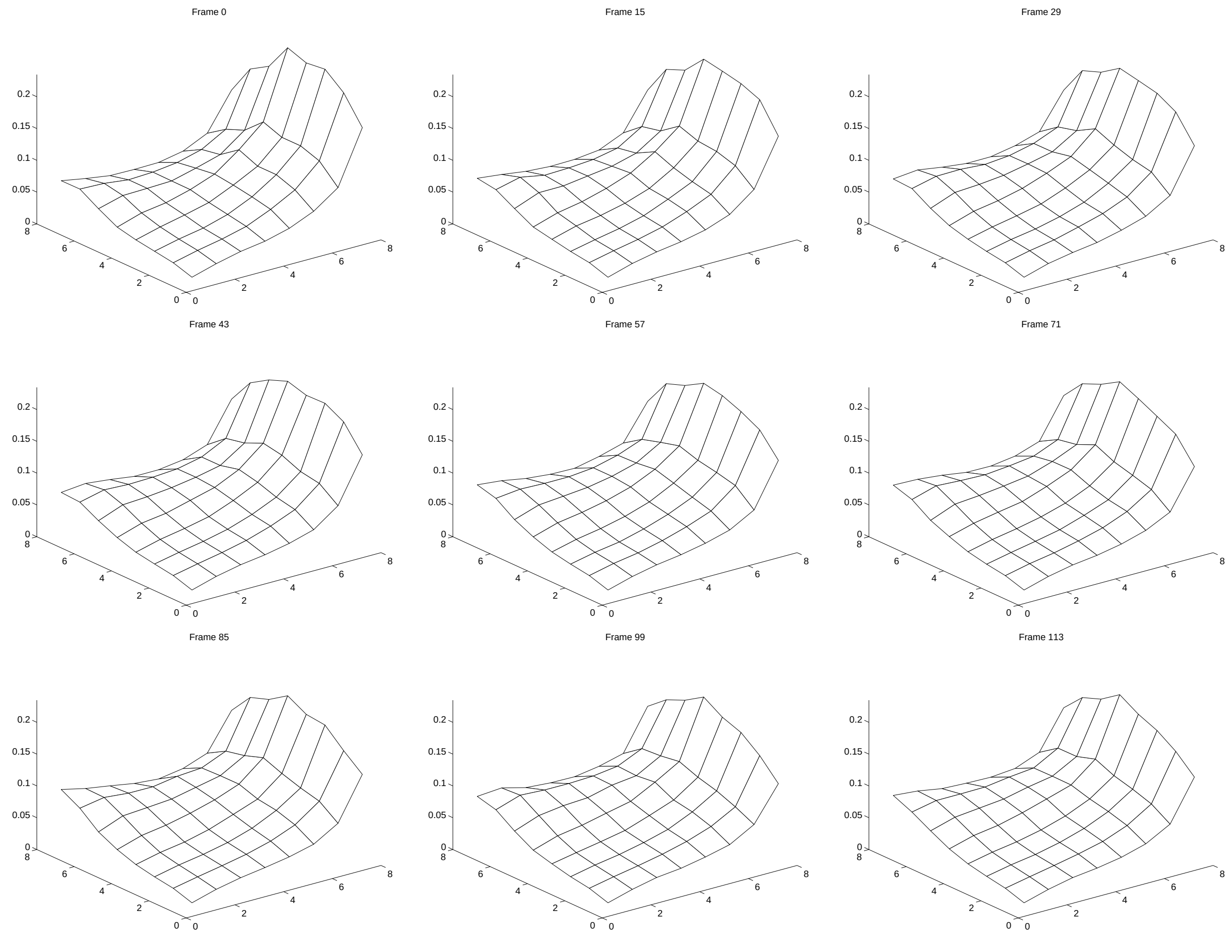


Figure 8: The Laplacian parameters in the mobile & calendar sequence luminance channels distributed through the sequence (frames 0, 15, 28, 43, 57, 71, 85, 99, and 113).

$$= \min_Q \left(\sum_{k=1}^{k=M} e^{-\lambda x} \left[2kQx + \frac{2kQ}{\lambda} - k^2Q^2 - x^2 - \frac{x}{\lambda^2} - \frac{1}{\lambda} \right] \Bigg|_{kQ-\frac{Q}{2}}^{kQ+\frac{Q}{2}} + \right.$$

$$\left. e^{-\lambda x} \left[2kQx + \frac{2kQ}{\lambda} - k^2Q^2 - x^2 - \frac{x}{\lambda^2} - \frac{1}{\lambda} \right] \Bigg|_{MQ+\frac{Q}{2}}^{\infty} \right) \quad (2)$$

While (2) is a relatively complex expression, it can be quickly evaluated to find the best quantization values. A table of perceptually weighted coefficients can be determined from experiments measuring the detectability of the cosine basis functions and normalizing the coefficients (or using techniques like [9]). Then, to find a base matrix, multiply each MSE-optimal Q by the perceptual weighting, and collect them in a matrix. This matrix can then be scaled (and rounded to integers) as appropriate for the bit-rate of the image or video stream.

8 CONCLUSIONS

The primary result of this paper is the confirmation of a Laplacian distribution for both the luminance and chrominance channels of DCT encoded images and error terms. Another surprising result was the stability of distribution values over images in a video sequence. Knowledge of these distribution values may lead to improved on-th-fly quantization matrix generation for video and image coding.

9 REFERENCES

[1] Peter J. Bickel and Kjell A. Doksum. *Mathematical Statistics: Basic Ideas and Selected Topics*. Holden-Day, Inc, 1977. Section 9.6.

[2] Kevin L. Gong. Parallel MPEG-1 video encoding. Master's thesis, University of CA at Berkeley, 1993. Substantially reporduced in PCS-94.

[3] ISO. Draft international standard 10918-1. CCITT Recomendation T.81.

[4] ISO/IEC JTC 1/SC 29. Standard 11172: Coded representation of picture, audio, and multimedia/hypermedia, 1991.

[5] ISO/IEC JTC1/SC29/WG11. Proposed standard 13818: Generic coding of moving pictures and associated audio information. Recommendation ITU-T H.262. Still in draft form.

[6] Anil K. Jain. *Fundamentals of Digital Image Processing*. Prentice Hall, Englewood Cliffs, NJ, 1989.

[7] Arun N. Netravali and Barry G. Haskell. *Digital Pictures: Representation and Compression*. Applications of Communications Theory (Series Editor R. W. Lucky). Plenum Press, NY, NY, 1988.

[8] William B. Pennebaker and Joan L. Mitchel. *JPEG: Still Image Data Compression Standard*. Van Nostrand Reinhold, 1993.

[9] Heidi Peterson, Albert J. Ahumada, and Andrew B. Watson. An improved detection model for DCT coefficient quantization. In *Human Vision, Visual Processing, and Digital Display IV*, pages 191–201. SPIE — The International Society for Optical Engineering, 1993.

[10] Randall C. Reininger and Jerry D. Gibson. Distrubutions of the two-dimensional DCT coefficients for images. *IEEE Transactions on Communications*, COM-31(6):835–839, June 1983.

[11] Stephen R. Smoot, Kevin Gong, and Lawrence A Rowe. UCB MPEG encoder. Source code is available by ftp from mm-ftp.cs.berkeley.edu:/pub/multimedia/mpeg or on the web at http://www-plateau.cs.berkeley.edu/mpeg, 1995.