A Mathematical Analysis of the DCT Coefficient Distributions for Images

Edmund Y. Lam, *Member, IEEE,* and Joseph W. Goodman, *Fellow, IEEE*

Abstract—Over the past two decades, there have been various studies on the distributions of the DCT coefficients for images. However, they have concentrated only on fitting the empirical data from some standard pictures with a variety of well-known statistical distributions, and then comparing their goodness-of-fit. The Laplacian distribution is the dominant choice balancing simplicity of the model and fidelity to the empirical data. Yet, to the best of our knowledge, there has been no mathematical justification as to what gives rise to this distribution. In this paper, we offer a rigorous mathematical analysis using a doubly stochastic model of the images, which not only provides the theoretical explanations necessary, but also leads to insights about various other observations from the literature. This model also allows us to investigate how certain changes in the image statistics could affect the DCT coefficient distributions.

Index Terms—Discrete cosine transforms, Gaussian distributions, image analysis, image coding, probability statistics.

I. INTRODUCTION

WITH the advent of the Internet and multimedia systems, the JPEG standard has gained widespread popularity for lossy compression of still-frame, continuous-tone images. In this algorithm, the image is first divided into nonoverlapping blocks of size 8×8 pixels, where each block is then subjected to the discrete cosine transform (DCT) before quantization and entropy coding. As the centerpiece of the compression algorithm, the DCT has been extensively studied by various researchers.

In fact, there has been an interest in understanding the distributions of the DCT coefficients since more than 20 years ago. The question is as follows: after we have performed the DCT on each of the blocks and collected the corresponding coefficients from them, what is the resulting statistical distribution? Such knowledge would be useful, for instance, in quantizer design and noise mitigation for image enhancement [1], [2]. Fig. 1 shows a typical plot of the histograms of the DCT coefficients. The image used here is the "bridge" picture shown in Fig. 2(a) from the standard image processing library. The upper left coefficient is called the dc coefficient while the rest are ac coefficients. The scaling of the histogram is kept the same for all ac coefficients in this plot. Early on, it was conjectured that the ac coefficients have Gaussian distributions [3]. However, soon
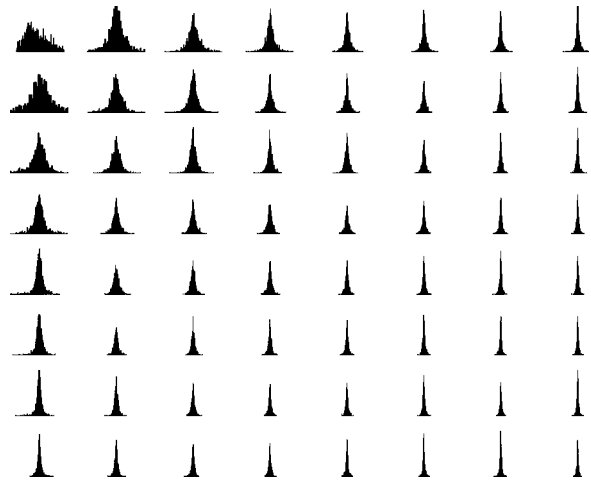
Fig. 1. Histogram of DCT coefficients of "bridge."

experimental results like Fig. 1 indicated that they resemble Laplacian distributions when the Kolmogorov–Smirnov goodness-of-fit test is used [4]. The probability density function of a Laplacian distribution can be written as

$$p(x) = \frac{\mu}{2} \exp\{-\mu|x|\}. \tag{1}$$

This is sometimes also referred to as the double exponential distribution. Since then, different researchers have tried a variety of fitting methods, such as χ^2, Kurtosis, and Watson tests. Many other possible distributions of the coefficients have also been proposed, including Cauchy, generalized Gaussian, and even a sum of Gaussians [5]–[9]. Using different pictures for the experiments, they often differ in opinion as to what distribution model is the most suitable, although the Laplacian distribution remains a popular choice balancing simplicity of the model and fidelity to the empirical data. Yet, none of them provided any analytic justification for their choices of distributions. In this paper, we investigate this problem in two steps: first, we derive the distribution of the DCT coefficients for a fixed block variance, and invoke the central limit theorem to show that this is approximately Gaussian. Then, by allowing the block variance to have a statistical distribution itself, we analyze when the DCT coefficients would follow a Laplacian distribution. We also study the case when little is known of the distribution of the block variance. These are detailed in Section II. In Section III we explain further why the width of the DCT coefficient distributions shrinks as we go to higher frequencies, yet the shape remains about the same. Empowered with these understandings, we proceed in Section IV to offer some insights to various observations reported in the literature.

(a)

(b)

(c)

Fig. 2. Some standard images used for testing. (a) "Bridge" image, (b) "crowd" image, and (c) "baboon" image.

II. DISTRIBUTIONS OF THE COEFFICIENTS: SHAPE

A. Constant Block Variance

Let $i_{p,q}$ denote a pixel, $p = 0, \cdots 7$, $q = 0, \cdots 7$, within an 8×8 block. The $i_{p,q}$'s are assumed to be identically distributed,

but do not have to be Gaussian. Taking the type-II DCT used in JPEG [10]

$$
I_{m,n} = \frac{C(m)}{2} \frac{C(n)}{2} \sum_{p=0}^{7} \sum_{q=0}^{7} i_{p,q}
$$
$$
\cdot \cos\left(\frac{(2p+1)\pi m}{16}\right) \cos\left(\frac{(2q+1)\pi n}{16}\right) \quad (2)
$$

with

$$
C(m) = \begin{cases} \dfrac{1}{\sqrt{2}}, & \text{for } m = 0 \\ 1, & \text{for } m > 0 \end{cases} \quad (3)
$$

and

$$
C(n) = \begin{cases} \dfrac{1}{\sqrt{2}}, & \text{for } n = 0 \\ 1, & \text{for } n > 0. \end{cases} \quad (4)
$$

By the central limit theorem, the (weighted) summation of identically distributed random variables can be well approximated as having a Gaussian distribution. Therefore, $I_{m,n}$ should be approximately distributed as Gaussian. Note that the central limit theorem applies even when the $i_{p,q}$'s are spatially correlated, as long as the magnitude of correlation is less than one. For typical images, the correlation will not be too large, and we generally have enough pixels in the summation to achieve a good approximation to a Gaussian distribution. In fact, the spatial correlation plays an important role in determining the width of the distributions as we shall see in Section III. Finally, because of the unitary nature of DCT, the mean of the Gaussian distribution is zero, while the variance (which we denote as $\sigma_{m,n}^2$ or just σ^2 for short) is proportional to the variance of pixels in the block [3].
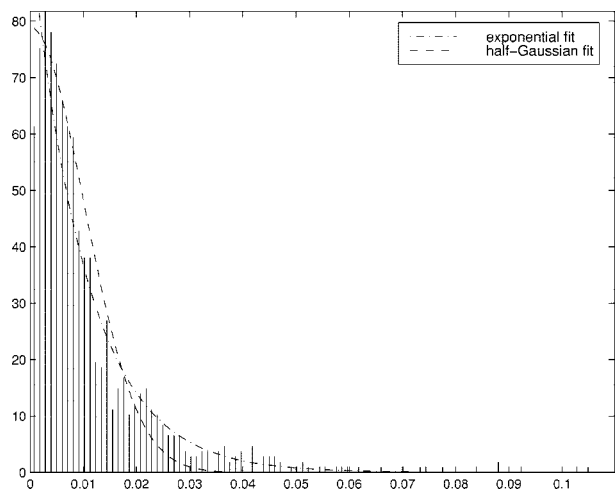
B. Changing Block Variance

One important characteristic of typical images is that the variance of the blocks is varying. We therefore model the image with a doubly stochastic model, with the variance of the Gaussian distributions itself also a random variable. This turns out to be the determining factor for the shape of the coefficient distributions. Let $p(\cdot)$ denote the probability density function. Using conditional probability

$$
p(I_{m,n}) = \int_0^{\infty} p(I_{m,n}|\sigma^2) p(\sigma^2) \, d(\sigma^2). \quad (5)
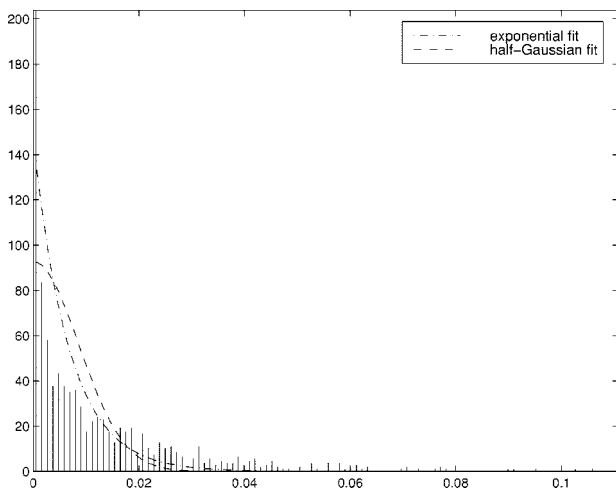$$

From the last section, we know that $p(I_{m,n}|\sigma^2)$ is approximately a zero-mean Gaussian, i.e.,

$$
p(I_{m,n}|\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{I_{m,n}^2}{2\sigma^2}\right\}. \quad (6)
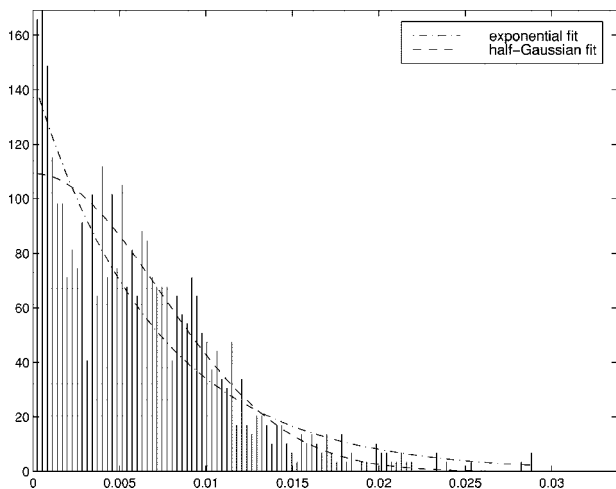$$

It remains to find the probability density function of the variance. Fig. 3 plots the histogram of the variance for the three test images shown in Fig. 2. It appears that their histograms can be reasonably approximated by exponential distributions. In fact, in each plot we also show the exponential and half-Gaussian

Fig. 3.   Histograms and approximations for three images. (a) Histogram and fits for "bridge," (b) histogram and fits for "crowd," and (c) histogram and fits for "baboon."

approximations to the empirical data, where we define the half-Gaussian distribution to be

$$p(x) = \begin{cases} 0 & x < 0 \\ \dfrac{2}{\sqrt{2\pi}s} \exp\left\{-\dfrac{x^2}{2s^2}\right\} & x \geq 0 \end{cases} \tag{7}$$

where s controls the variance of the distribution.

Considering the case that σ^2 is indeed exponentially distributed, we put $p(\sigma^2) = \lambda \exp\{-\lambda\sigma^2\}$ in (5) to get

$$
\begin{aligned}
p(I_{m,n}) &= \int_0^\infty \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{I_{m,n}^2}{2\sigma^2}\right\} \lambda \exp\{-\lambda\sigma^2\}\, d(\sigma^2) \\
&= \sqrt{\frac{2}{\pi}}\,\lambda \int_0^\infty \exp\left\{-\lambda\sigma^2 - \left(\frac{I_{m,n}^2}{2}\right)\frac{1}{\sigma^2}\right\} d\sigma \\
&= \left(\sqrt{\frac{2}{\pi}}\,\lambda\right)\left(\frac{1}{2}\sqrt{\frac{\pi}{\lambda}}\right)\exp\left\{-2\sqrt{\lambda\frac{I_{m,n}^2}{2}}\right\} \\
&= \frac{\sqrt{2\lambda}}{2}\exp\left\{-\sqrt{2\lambda}|I_{m,n}|\right\}
\end{aligned}
\tag{8}
$$

where we have used the relationship

$$\int_0^\infty \exp\left\{-ax^2 - \frac{b}{x^2}\right\} dx = \frac{1}{2}\sqrt{\frac{\pi}{a}}\exp\left\{-2\sqrt{ab}\right\} \tag{9}$$

where $a > 0$, $b > 0$, from an integral table [11]. In this case, the resulting distribution of $I_{m,n}$ is exactly Laplacian with parameter $\mu = \sqrt{2\lambda}$. This explains why the distributions of the ac coefficients are often observed to be approximately Laplacian [1], [4], [5], [8].

If we use the half-Gaussian assumption for the probability density function of σ^2, we substitute (7) in (5) and get

$$
\begin{aligned}
p(I_{m,n}) &= \frac{2}{\pi s}\int_0^\infty \exp\left\{-\frac{I_{m,n}^2}{2\sigma^2} - \frac{\sigma^4}{2s^2}\right\} d\sigma \\
&= \frac{1}{s}\sqrt{\frac{2}{\pi}}\int_0^\infty \left(\sqrt{\frac{2}{\pi}}\exp\left\{-\frac{\sigma^4}{2s^2}\right\}\right) \\
&\quad \cdot \exp\left\{-\frac{I_{m,n}^2}{2\sigma^2}\right\} d\sigma \\
&\approx \frac{1}{s}\sqrt{\frac{2}{\pi}}\int_0^\infty \exp\left\{-\frac{\sigma^2}{s} - \frac{I_{m,n}^2}{2\sigma^2}\right\} d\sigma \\
&= \frac{\sqrt{2/s}}{2}\exp\left\{-\sqrt{2/s}|I_{m,n}|\right\}.
\end{aligned}
\tag{10}
$$

So, this again is a Laplacian distribution, with $\mu = \sqrt{2/s}$. The approximation above is justified by the numerical results shown in Fig. 4, which plots $p(I_{m,n})$ and the approximation for two typical values of s. They appear to be very close. We can therefore argue that the distributions of the ac coefficients are also closely Laplacian in this case. From these results, we can further surmise that if the distribution of the variance is somewhere between exponential and half-Gaussian, the resulting distribution of $I_{m,n}$ is also very close to Laplacian.

Even if the block variance has a distribution that departs significantly from an exponential or half-Gaussian distribution, we can still draw some conclusions about the distribution of $I_{m,n}$ by looking at its high order statistics. A quick review of some definitions and examples is given in Appendix. For further reference, the reader may consult [12].
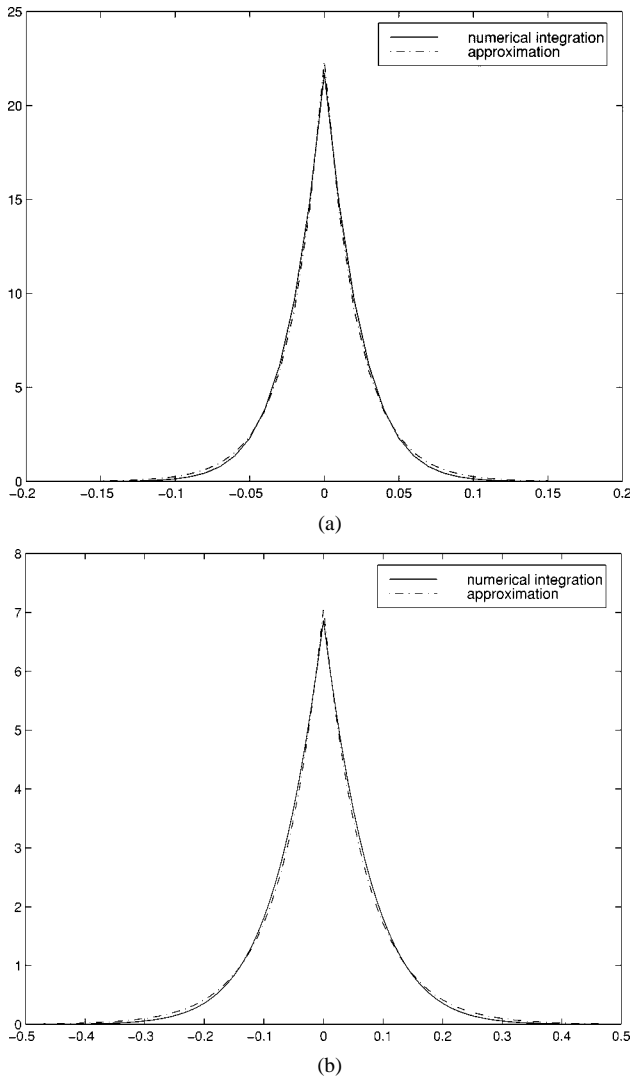
Fig. 4. Approximation with Laplacian distribution. (a) $s = 0.001$, $\mu = \sqrt{2/s}$ and (b) $s = 0.01$, $\mu = \sqrt{2/s}$.

As argued before, $I_{m,n}$ has zero mean. Therefore, we can readily calculate its variance by

$$\mathrm{Var}[I_{m,n}] = E[I_{m,n}^2] = E[E[I_{m,n}^2|\sigma^2]] = E[\sigma^2]. \qquad (11)$$

The skewness coefficient is zero because $E[I_{m,n}^3] = E[E[I_{m,n}^3|\sigma^2]] = 0$. As for the kurtosis coefficient, we know that

$$E[I_{m,n}^4] = E[E[I_{m,n}^4|\sigma^2]] = E[3\sigma^4] = 3E[\sigma^4] \qquad (12)$$

because a Gaussian distribution has a kurtosis of three. Therefore

$$\gamma_2 = \frac{E[I_{m,n}^4]}{\mathrm{Var}[I_{m,n}]^2} = \frac{3E[\sigma^4]}{E[\sigma^2]^2} > 3 \qquad (13)$$

whenever $E[\sigma^4] > E[\sigma^2]^2$. This is always true when the distribution of the block variance is nonsingular, because its variance equals to $E[\sigma^4] - E[\sigma^2]^2$, which must be positive. Furthermore, a large variance also implies a large kurtosis coefficient for $I_{m,n}$.

Therefore, we can conclude from the high order statistics analysis that $I_{m,n}$ must have a fat-tail distribution. For a general kurtosis value, a generalized Gaussian distribution is usually a better fit. It has a probability density function

$$p_{\nu,\beta}(x) = \frac{\nu}{2\beta\Gamma\left(\frac{1}{\nu}\right)} \exp\left\{-\left(\frac{|x|}{\beta}\right)^\nu\right\} \qquad (14)$$

with $\nu > 0$. $\Gamma(\cdot)$ is the standard Gamma function [13]. This distribution is also commonly called the exponential power distribution in statistics [12]. Its kurtosis coefficient is

$$\gamma_2 = \frac{\Gamma\left(\frac{5}{\nu}\right)\Gamma\left(\frac{1}{\nu}\right)}{\left[\Gamma\left(\frac{3}{\nu}\right)\right]^2} \qquad (15)$$

which can take on any value greater than or equal to 1.8. When $\nu = 2$ and $\beta = \sqrt{2}$, it reduces back to a standard Gaussian distribution, with a kurtosis coefficient of three. As $\nu \to \infty$, it approaches the uniform distribution. The Laplacian distribution is also a special case of the generalized Gaussian distribution by setting $\nu = 1$ and $\beta = 1/\lambda$, with a kurtosis coefficient of six. The extra parameter ν therefore allows for greater flexibility in modeling $I_{m,n}$ for different kurtosis values. Therefore, the generalized Gaussian distribution would result in an equal or better fit to the empirical data than Laplacian, yet at the expense of more complex expressions and extra computational cost.

III. DISTRIBUTIONS OF THE COEFFICIENTS: WIDTH

While the discussion above enables us to explain fully the shape of the DCT coefficients, it does not account for the differences in "width," or the variance, of the distributions across various coefficients. Such differences are in fact due to the correlation among pixels. Indeed, if the pixels are not only identically distributed but also independent, we expect all the ac coefficients to have similar widths because energy would be evenly distributed across different frequency bands.

As noted previously, because of the unitary nature of the DCT, the variance of the Gaussian distribution obtained by invoking the central limit theorem is directly proportional to the variance of the block. However, the constant of proportionality is different for different coefficients: it is smaller as we go to higher frequency coefficients. Referring back to (2), we are essentially summing the random variable $i_{p,q}$'s with weights $\cos((2p + 1)\pi m/16)\cos((2q + 1)\pi n/16)$, over all values of p and q. At low frequencies, the weights of adjacent pixels are close, so the overall summation generally leads to values of large magnitude. However, at high frequencies, the weights of adjacent pixels are close in magnitude but opposite in sign. When the pixels are correlated spatially, the overall effect is that they tend to cancel each other in the summation, and therefore the resulting DCT coefficient usually has a small magnitude.

There is an alternative explanation of this phenomenon from a more traditional signal processing perspective. The blockwise

DCT can be interpreted as a subband decomposition after re-arranging the coefficients [14]. Therefore, the coefficient distributions are in fact equivalent to the histograms of the various subbands. Since all natural images possess certain degrees of spatial correlation, their energy is usually concentrated in the lower frequency bands. As a result, it is natural to expect more energy—hence larger variance—in the low frequency subbands of the original image than we would expect from the high-frequency subbands. Therefore, as we look at the histogram, the width of the Laplacian distributions decreases when we move to higher frequency subbands.

IV. INSIGHTS

The previous analyses offer us better insights for some empirical observations put forth in the literature. Here, we collect a few of them.

- Müller used the generalized Gaussian distribution and allowed different values of ν for different DCT coefficients for best fit. He reported that "the value of ν for the different coefficients does not vary much within a single image" [6]. From our discussions above, we could have predicted that to be the case since the distribution shape for different coefficients depends on the same factor, i.e. the distribution of the block variance.
- In [7], Eude *et al.* attempted to fit the distributions with a finite mixture of Gaussian distributions. That is in line with what we are doing, but we go further and show that with particular mixtures of the Gaussians, we can form the Laplacian distribution which is much easier to study analytically.
- Smoot studied the DCT coefficient distributions of color images. Comparing the luminance and chrominance channels, he found that "the chrominance data. . .exhibits the same distribution as for the luminance channel. However the parameters [of the Laplacian distributions] are larger in general, indicating more near-zero values" [8]. Our analyses above would imply that the chrominance channels have larger inter-pixel correlations than the luminance channel, while the distribution of the variance of the blocks is about the same. We verified these to be true for our test images.
- Yovanof and Liu were quite right to notice that "the inability of all those studies to agree upon a single model is partially due to the large variations among the various imagery data sources" [9]. We could now further pinpoint the source of the variation in models to be the fact that the variance of the blocks is not exactly distributed as exponential. When it departs significantly from an exponential or a half-Gaussian distribution, it is possible that the resulting distribution looks less like a Laplacian.

V. CONCLUSIONS

In this paper, we offer a comprehensive mathematical analysis of the DCT coefficient distributions of natural images. We demonstrate how a Laplacian distribution of the coefficients can in fact be derived by using a doubly stochastic model. Furthermore, we also explain the difference in width of the Laplacian
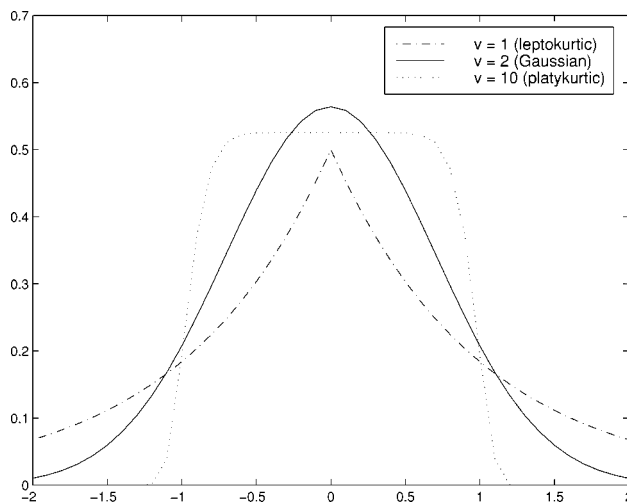

Fig. 5. Generalized Gaussian distribution with different kurtosis coefficients.

distributions across various DCT coefficients. These together bring us better insights into the DCT and subsequently the JPEG compression algorithm.

Further research could look into the distributions of the wavelet coefficients and the DCT coefficients of a motion-compensated block of an interframe-coded video frame, both of which have been observed to also possess Laplacian distributions [15], [16]. We believe the causes of their Laplacian distributions to be similar to the case we have presented here. As wavelets play a larger role for image compression with JPEG 2000 and digital video compression becomes more prevalent, we trust that these analyses would be beneficial for the development of multimedia systems.

APPENDIX

For a random variable Y, let $m = E[Y]$ and $s^2 = E[(Y - m)^2]$. The skewness coefficient γ_1 is defined as

$$\gamma_1 = \frac{E[(Y - m)^3]}{s^3} \tag{16}$$

while the kurtosis coefficient γ_2 is defined as

$$\gamma_2 = \frac{E[(Y - m)^4]}{s^4}. \tag{17}$$

For a Gaussian distribution, $\gamma_2 = 3$. If $\gamma_2 > 3$, the distribution is said to be leptokurtic, or fat-tail. For example, a Laplacian distribution has $\gamma_2 = 6$. If $\gamma_2 < 3$, the distribution is said to be platykurtic. For example, a uniform distribution between -0.5 and 0.5 has $\gamma_2 = 1.8$. Fig. 5 plots the generalized Gaussian distribution with different values of γ_2 by changing ν.

REFERENCES

[1] G. Lakhani, "Adjustments for JPEG de-quantization coefficients," in *1998 Proceedings Data Compression Conference*, J. A. Storer and M. Cohn, Eds. Los Alamitos, CA: IEEE Comput. Soc. Press, Mar. 1998, p. 557.
[2] R. Brown and A. Boden, "*A posteriori* restoration of block transform-compressed data," in *1995 Proceedings Data Compression Conference*, J. A. Storer and M. Cohn, Eds. Los Alamitos, CA: IEEE Comput. Soc. Press, Mar. 1995, p. 426.

[3] W. K. Pratt, *Digital Image Processing*. New York: Wiley, 1978.

[4] R. Reininger and J. Gibson, "Distributions of the two-dimensional DCT coefficients for images," *IEEE Trans. Commun.*, vol. COM-31, pp. 835–839, June 1983.

[5] J. D. Eggerton and M. D. Srinath, "Statistical distributions of image DCT coefficients," *Comput. Elect. Eng.*, vol. 12, no. 3/4, pp. 137–145, 1986.

[6] F. Müller, "Distribution shape of two-dimensional DCT coefficients of natural images," *Electron. Lett.*, vol. 29, no. 22, pp. 1935–1936, Oct. 1993.

[7] T. Eude, R. Grisel, H. Cherifi, and R. Debrie, "On the distribution of the DCT coefficients," in *Proc. 1994 IEEE Int. Conf. Acoustics, Speech, Signal Processing*, vol. 5, Apr. 1994, pp. 365–368.

[8] S. R. Smoot and L. A. Rowe, "Study of DCT coefficient distributions," *Proc. SPIE*, pp. 403–411, Jan. 1996.

[9] G. S. Yovanof and S. Liu, "Statistical analysis of the DCT coefficients and their quantization error," in *Conf. Rec. 30th Asilomar Conf. Signals, Systems, Computers*, vol. 1, 1997, pp. 601–605.

[10] W. Pennebaker and J. Mitchell, *JPEG Still Image Data Compression Standard*. New York: Van Nostrand Reinhold, 1992.

[11] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 5th ed. New York: Academic, 1994.

[12] G. E. P. Box and G. C. Tiao, *Bayesian Inference in Statistical Analysis*. New York: Wiley, 1992.

[13] W. H. Pun and B. D. Jeffs, "Adaptive image restoration using a generalized Gaussian model for unknown noise," *IEEE Trans. Image Processing*, vol. 4, pp. 1451–1456, Oct. 1995.

[14] A. Akansu and M. Smith, *Subband and Wavelet Transforms: Design and Applications*. Norwell, MA: Kluwer, 1996.

[15] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, pp. 674–693, July 1989.

[16] F. Bellifemine, A. Capellino, A. Chimienti, R. Picco, and R. Ponti, "Statistical analysis of the 2D-DCT coefficients of the differential signal for images," *Signal Process. Image Commun.*, vol. 4, pp. 477–488, Nov. 1992.

Joseph W. Goodman (S'58–M'63–SM'72–F'74) received the A.B. degree in engineering and applied physics from Harvard University, Cambridge, MA, and the M.S. and Ph.D. degrees in electrical engineering from Stanford University, Stanford, CA.

He is currently William E. Ayer Professor Emeritus of electrical engineering at Stanford University. He served as Chair of the Department of Electrical Engineering (1988–1996) and as Senior Associate Dean of Engineering (1996–1998). He is the author of the textbooks *Introduction to Fourier Optics* (New York: McGraw-Hill, 1st ed., 1968, 2nd ed., 1996) and *Statistical Optics* (New York: Wiley, 1985), and co-author (with R. Gray) of *Fourier Transforms: An Introduction for Engineers* (Norwell, MA: Kluwer, 1995), and editor of *International Trends in Optics* (New York: Academic, 1991), as well as more than 200 technical publications. His fields of research have covered Fourier optics, statistical optics, image processing, and, in recent years, the use of photonics in communications and computing.

Dr. Goodman received the F. E. Terman Award of the ASEE, the Dennis Gabor Award of the SPIE, the Max Bom Award of the OSA, the Frederic Ives Medal of the OSA (the society's highest award), the Esther Hoffman Beller award of the OSA for excellence in optics education, and the Education Medal of the IEEE. He is a member of the National Academy of Engineering as well as a Fellow of SPIE, OSA, and the American Academy of Arts and Sciences. He has served as President of both the International Commission for Optics and the Optical Society of America.

Edmund Y. Lam (S'97–M'00) was born in Hong Kong in 1974. He received the B.S. degree in 1995, the M.S. degree in 1996, and the Ph.D. degree in 2000, all in electrical engineering from Stanford University, Stanford, CA.

From 1995 to 1997, he was a Teaching Assistant with the Department of Electrical Engineering, Stanford University. Since 1996, he has been conducting research in the Information Systems Laboratory for the Stanford Programmable Digital Camera Project, with particular focus on developing image restoration algorithms for digital photography. Outside Stanford, he has also performed consulting work for industry in the areas of digital camera systems design and algorithms development. His current research interests are in image enhancement, restoration and compression, optical imaging systems, color science, and statistical signal processing.

Dr. Lam is a member of the Optical Society of America (OSA), the International Society for Optical Engineering (SPIE), and Tau Beta Pi.