CHAPTER

Image Noise Models

7

Charles Boncelet

University of Delaware

7.1 SUMMARY

This chapter reviews some of the more commonly used *image noise models*. Some of these are naturally occurring, e.g., Gaussian noise, some sensor induced, e.g., photon counting noise and speckle, and some result from various processing, e.g., quantization and transmission.

7.2 PRELIMINARIES

7.2.1 What is Noise?

Just what is noise, anyway? Somewhat imprecisely we will define *noise* as an unwanted component of the image. Noise occurs in images for many reasons. Gaussian noise is a part of almost any signal. For example, the familiar white noise on a weak television station is well modeled as Gaussian. Since image sensors must count photons—especially in low-light situations—and the number of photons counted is a random quantity, images often have photon counting noise. The grain noise in photographic films is sometimes modeled as Gaussian and sometimes as Poisson. Many images are corrupted by salt and pepper noise, as if someone had sprinkled black and white dots on the image. Other noises include quantization noise and speckle in coherent light situations.

Let $\mathbf{f}(\cdot)$ denote an image. We will decompose the image into a desired component, $g(\cdot)$, and a noise component, $\mathbf{q}(\cdot)$. The most common decomposition is *additive*:

$$\mathbf{f}(\cdot) = g(\cdot) + \mathbf{q}(\cdot). \tag{7.1}$$

For instance, Gaussian noise is usually considered to be an additive component.

The second most common decomposition is *multiplicative*:

$$\mathbf{f}(\cdot) = g(\cdot)\mathbf{q}(\cdot). \tag{7.2}$$

An example of a noise often modeled as multiplicative is speckle.

Note, the multiplicative model can be transformed into the additive model by taking logarithms and the additive model into the multiplicative one by exponentiation. For instance, (7.1) becomes

$$e^{\mathbf{f}} = e^{g+\mathbf{q}} = e^{g}e^{\mathbf{q}}. \tag{7.3}$$

Similarly, (7.2) becomes

$$\log \mathbf{f} = \log(g\mathbf{q}) = \log g + \log \mathbf{q}. \tag{7.4}$$

If the two models can be transformed into one another, what is the point? Why do we bother? The answer is that we are looking for *simple* models that properly describe the behavior of the system. The additive model, (7.1), is most appropriate when the noise *in that model* is independent of f. There are many applications of the additive model. Thermal noise, photographic noise, and quantization noise, for instance, obey the additive model well.

The multiplicative model is most appropriate when the noise in that model is independent of f. One common situation where the multiplicative model is used is for speckle in coherent imagery.

Finally, there are important situations when neither the additive nor the multiplicative model fits the noise well. Poisson counting noise and salt and pepper noise fit neither model well.

The questions about noise models one might ask include: What are the properties of $\mathbf{q}(\cdot)$? Is $\mathbf{q}$ related to g or are they independent? Can $\mathbf{q}(\cdot)$ be eliminated or at least, mitigated? As we will see in this chapter and in others, it is only occasionally true that $\mathbf{q}(\cdot)$ will be independent of $g(\cdot)$. Furthermore, it is usually impossible to remove all the effects of the noise.

Figure 7.1 is a picture of the San Francisco, CA, skyline. It will be used throughout this chapter to illustrate the effects of various noises. The image is 432×512, 8 bits per pixel, grayscale. The largest value (the whitest pixel) is 220 and the minimum value is 32. This image is relatively noise free with sharp edges and clear details.

7.2.2 Notions of Probability

The various noises considered in this chapter are random in nature. Their exact values are random variables whose values are best described using probabilistic notions. In this section, we will review some of the basic ideas of probability. A fuller treatment can be found in many texts on probability and randomness, including Feller [1], Billingsley [2], and Woodroofe [3].

Let $\mathbf{a} \in R^n$ be a n-dimensional *random vector* and $a \in R^n$ be a point. Then the *distribution function* of $\mathbf{a}$ (also known as the cumulative distribution function) will be denoted as $P_{\mathbf{a}}(a) = \Pr[\mathbf{a} \leq a]$ and the corresponding *density function,* $p_{\mathbf{a}}(a) = dP_{\mathbf{a}}(a)/da$. Probabilities of events will be denoted as $\Pr[A]$.

The *expected value* of a function, $\psi(\mathbf{a})$ is

$$E[\psi(\mathbf{a})] = \int_{-\infty}^{\infty} \psi(a) p_{\mathbf{a}}(a)\, da. \tag{7.5}$$

FIGURE 7.1
Original picture of San Francisco skyline.

Note that for discrete distributions the integral is replaced by the corresponding sum:

$$E[\psi(\mathbf{a})] = \sum_k \psi(a_k)\Pr[\mathbf{a} = a_k]. \tag{7.6}$$

The *mean* is $\mu_\mathbf{a} = E[\mathbf{a}]$ (i.e., $\psi(\mathbf{a}) = \mathbf{a}$), the *variance* of a single random variable is $\sigma_\mathbf{a}^2 = E\big[(\mathbf{a} - \mu_\mathbf{a})^2\big]$, and the *covariance matrix* of a random vector is $\Sigma_\mathbf{a} = E\big[(\mathbf{a} - \mu_\mathbf{a})(\mathbf{a} - \mu_\mathbf{a})^T\big]$.

Related to the covariance matrix is the *correlation matrix*,

$$\mathbf{R}_\mathbf{a} = E\big[\mathbf{a}\mathbf{a}^T\big]. \tag{7.7}$$

The various moments are related by the well-known relation, $\Sigma = \mathbf{R} - \mu\mu^T$.

The *characteristic function*, $\Phi_\mathbf{a}(u) = E\big[\exp(ju\mathbf{a})\big]$, has two main uses in analyzing probabilistic systems: calculating moments and calculating the properties of sums of independent random variables. For calculating moments, consider the power series of $\exp(ju\mathbf{a})$:

$$e^{ju\mathbf{a}} = 1 + ju\mathbf{a} + \frac{(ju\mathbf{a})^2}{2!} + \frac{(ju\mathbf{a})^3}{3!} + \cdots \tag{7.8}$$

After taking expected values,

$$E\big[e^{ju\mathbf{a}}\big] = 1 + juE[\mathbf{a}] + \frac{(ju)^2 E\big[\mathbf{a}^2\big]}{2!} + \frac{(ju)^3 E\big[\mathbf{a}^3\big]}{3!} + \cdots, \tag{7.9}$$

One can isolate the kth moment by taking k derivatives with respect to u and then setting $u = 0$:

$$E\left[\mathbf{a}^k\right] = \frac{1}{j^k} \frac{d^k E\left[e^{ju\mathbf{a}}\right]}{d^k u}\bigg|_{u=0}. \tag{7.10}$$

Consider two independent random variables, $\mathbf{a}$ and $\mathbf{b}$, and their sum $\mathbf{c}$. Then,

$$\Phi_{\mathbf{c}}(u) = E\left[e^{ju(\mathbf{c})}\right] \tag{7.11}$$

$$= E\left[e^{ju(\mathbf{a}+\mathbf{b})}\right] \tag{7.12}$$

$$= E\left[e^{ju\mathbf{a}} e^{ju\mathbf{b}}\right] \tag{7.13}$$

$$= E\left[e^{ju\mathbf{a}}\right] E\left[e^{ju\mathbf{b}}\right] \tag{7.14}$$

$$= \Phi_{\mathbf{a}}(u)\Phi_{\mathbf{b}}(u), \tag{7.15}$$

where (7.14) used the independence of $\mathbf{a}$ and $\mathbf{b}$. Since the characteristic function is the (complex conjugate of the) Fourier transform of the density, the density of $\mathbf{c}$ is easily calculated by taking an inverse Fourier transform of $\Phi_{\mathbf{c}}(u)$.

7.3 ELEMENTS OF ESTIMATION THEORY

As we said in the introduction, noise is generally an *unwanted* component in an image. In this section, we review some of the techniques to eliminate—or at least minimize—the noise.

The basic estimation problem is to find a good estimate of the noise-free image, g, given the noisy image, $\mathbf{f}$. Some authors refer to this as an *estimation problem*, while others say it is a *filtering* problem. Let the estimate be denoted $\hat{\mathbf{g}} = \hat{\mathbf{g}}(\mathbf{f})$. The most common performance criterion is the mean squared error (MSE):

$$\text{MSE}(g, \hat{\mathbf{g}}) = E\left[(g - \hat{\mathbf{g}})^2\right]. \tag{7.16}$$

The estimator that minimizes the MSE is called the *minimum mean squared error estimator* (MMSE). Many authors prefer to measure the performance in a positive way using the *peak signal-to-noise ratio* (PSNR) measured in dB:

$$\text{PSNR} = 10\log_{10}\left(\frac{\text{MAX}^2}{\text{MSE}}\right), \tag{7.17}$$

where MAX is the maximum pixel value, e.g., 255 for 8 bit images.

While the MSE is the most common error criterion, it is by no means the only one. Many researchers argue that MSE results are not well correlated with the human visual system. For instance, the mean absolute error (MAE) is often used in motion compensation in video compression. Nevertheless, MSE has the advantages of easy tractability and intuitive appeal since MSE can be interpreted as "noise power." Estimators can be classified in many different ways. The primary division we will consider here is linear versus nonlinear estimators.

The *linear estimators* form estimates by taking linear combinations of the sample values. For example, consider a small region of an image modeled as a constant value plus additive noise:

$$\mathbf{f}(x,y) = \mu + \mathbf{q}(x,y). \tag{7.18}$$

A linear estimate of μ is

$$\hat{\mu} = \sum_{x,y} \alpha(x,y)\mathbf{f}(x,y) \tag{7.19}$$

$$= \mu \sum_{x,y} \alpha(x,y) + \sum_{x,y} \alpha(x,y)\mathbf{q}(x,y). \tag{7.20}$$

An estimator is called *unbiased* if $E\left[\mu - \hat{\mu}\right] = 0$. In this case, assuming $E[\mathbf{q}] = 0$, unbiasedness requires $\sum_{x,y} \alpha(x,y) = 1$. If the $\mathbf{q}(x,y)$ are independent and identically distributed (i.i.d.), meaning that the random variables are independent and each has the same distribution function, then the MMSE for this example is the sample mean:

$$\hat{\mu} = \frac{1}{M} \sum_{(x,y)} \mathbf{f}(x,y), \tag{7.21}$$

where M is the number of samples averaged over.

Linear estimators in image filtering get more complicated primarily for two reasons: Firstly, the noise may not be i.i.d., and secondly and more commonly, the noise-free image is not well modeled as a constant. If the noise-free image is Gaussian and the noise is Gaussian, then the optimal estimator is the well-known Weiner filter [4].

In many image filtering applications, linear filters do not perform well. Images are not well modeled as Gaussian, and linear filters are not optimal. In particular, images have small details and sharp edges. These are blurred by linear filters. It is often true that the filtered image is more objectionable than the original. The blurriness is worse than the noise.

Largely because of the blurring problems of linear filters, nonlinear filters have been widely studied in image filtering. While there are many classes of nonlinear filters, we will concentrate on the class based on order statistics. Many of these filters were invented to solve image processing problems.

Order statistics are the result of sorting the observations from smallest to largest. Consider an image *window* (a small piece of an image) centered on the pixel to be

estimated. Some windows are square, some are "x" shaped, some are "+" shaped, and some more oddly shaped. The choice of a window size and shape is usually up to the practitioner. Let the samples in the window be denoted simply as $\mathbf{f}_i$ for $i = 1,\ldots,N$. The order statistics are denoted $\mathbf{f}_{(i)}$ for $i = 1,\ldots,N$ and obey the ordering $\mathbf{f}_{(1)} \leq \mathbf{f}_{(2)} \leq \cdots \leq \mathbf{f}_{(N)}$.

The simplest order statistic-based estimator is the sample median, $\mathbf{f}_{((N+1)/2)}$. For example, if $N = 9$, the median is $\mathbf{f}_{(5)}$. The median has some interesting properties. Its value is one of the samples. The median tends to blur images much less than the mean. The median can pass an edge without any blurring at all.

Some other order statistic estimators are the following:

Linear Combinations of Order Statistics $\hat{\mu} = \sum_{i=1}^{N} \alpha_i \mathbf{f}_{(i)}$. The α_i determine the behavior of the filter. In some cases, the coefficients can be determined optimally, see Lloyd [5] and Bovik *et al.* [6].

Weighted Medians and the LUM Filter Another way to weight the samples is to repeat certain samples more than once before the data is sorted. The most common situation is to repeat the center sample more than once. The center weighted median does "less filtering" than the ordinary median and is suitable when the noise is not too severe. (See Salt and Pepper Noise below.)

The LUM filter [7] is a rearrangement of the center weighted median. It has the advantages of being easy to understand and extensible to image sharpening applications.

Iterated and Recursive Forms The various filtering operations can be combined or iterated upon. One might first filter horizontally, then vertically. One might compute the outputs of three or more filters and then use "majority rule" techniques to choose between them.

To analyze or optimally design order statistics filters, we need descriptions of the probability distributions of the order statistics. Initially, we will assume the $\mathbf{f}_i$ are i.i.d. Then the $\Pr[\mathbf{f}_{(i)} \leq x]$ equals the probability that at least i of the $\mathbf{f}_i$ are less than or equal to x. Thus,

$$\Pr[\mathbf{f}_{(i)} \leq x] = \sum_{k=i}^{N} \binom{N}{k} (P_{\mathbf{f}}(x))^k (1 - P_{\mathbf{f}}(x))^{N-k}. \tag{7.22}$$

We see immediately that the order statistic probabilities are related to the binomial distribution.

Unfortunately (7.22) does not hold when the observations are not i.i.d. In the special case when the observations are independent (or Markov), but not identically distributed, there are simple recursive formulas to calculate the probabilities [8, 9]. For example, even if the additive noise in (7.1) is i.i.d, the image may not be constant throughout the window. One may be interested in how much blurring of an edge is done by a particular order statistics filter.

7.4 TYPES OF NOISE AND WHERE THEY MIGHT OCCUR

In this section, we present some of the more common image noise models and show sample images illustrating the various degradations.

7.4.1 Gaussian Noise

Probably the most frequently occurring noise is additive Gaussian noise. It is widely used to model thermal noise and, under some often reasonable conditions, is the limiting behavior of other noises, e.g., photon counting noise and film grain noise. Gaussian noise is used in many places in this *Guide*.

The density function of univariate Gaussian noise, $\mathbf{q}$, with mean μ and variance σ^2 is

$$p_{\mathbf{q}}(x) = (2\pi\sigma^2)^{-1/2} e^{-(x-\mu)^2/2\sigma^2} \tag{7.23}$$

for $-\infty < x < \infty$. Notice that the support, which is the range of values of x where the probability density is nonzero, is infinite in both the positive and negative directions. But, if we regard an image as an intensity map, then the values must be nonnegative. In other words, the noise cannot be strictly Gaussian. If it were, there would be some nonzero probability of having negative values. In practice, however, the range of values of the Gaussian noise is limited to about $\pm 3\sigma$, and the Gaussian density is a useful and accurate model for many processes. If necessary, the noise values can be truncated to keep $f > 0$.

In situations where $\mathbf{a}$ is a random vector, the multivariate Gaussian density becomes

$$p_{\mathbf{a}}(a) = (2\pi)^{-n/2} |\Sigma|^{-1/2} e^{-(a-\mu)^T \Sigma^{-1} (a-\mu)/2}, \tag{7.24}$$

where $\mu = E[\mathbf{a}]$ is the mean vector and $\Sigma = E\left[(\mathbf{a} - \mu)(\mathbf{a} - \mu)^T\right]$ is the covariance matrix. We will use the notation $\mathbf{a} \sim N(\mu, \Sigma)$ to denote that $\mathbf{a}$ is Gaussian (also known as *Normal*) with mean μ and covariance Σ.

The Gaussian characteristic function is also Gaussian in shape:

$$\Phi_{\mathbf{a}}(u) = e^{u^T \mu - u^T \Sigma u/2}. \tag{7.25}$$

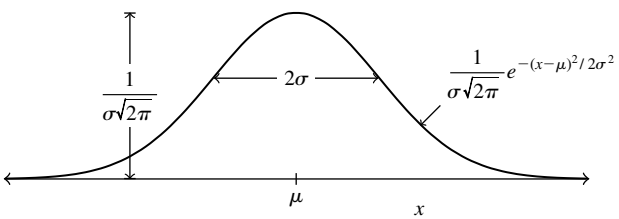

FIGURE 7.2

The Gaussian density.

The Gaussian distribution has many convenient mathematical properties—and some not so convenient ones. Certainly the least convenient property of the Gaussian distribution is that the cumulative distribution function cannot be expressed in closed form using elementary functions. However, it is tabulated numerically. See almost any text on probability, e.g., [10].

Linear operations on Gaussian random variables yield Gaussian random variables. Let $\mathbf{a}$ be $N(\mu, \Sigma)$ and $\mathbf{b} = G\mathbf{a} + h$. Then a straightforward calculation of $\Phi_{\mathbf{b}}(u)$ yields

$$\Phi_{\mathbf{b}}(u) = e^{ju^T(G\mu+h)-u^TG\Sigma G^Tu/2}, \tag{7.26}$$

which is the characteristic function of a Gaussian random variable with mean, $G\mu + h$, and covariance, $G\Sigma G^T$.

Perhaps the most significant property of the Gaussian distribution is called the Central Limit Theorem, which states that the distribution of a sum of a large number of independent, small random variables has a Gaussian distribution. Note the individual random variables do not need to have a Gaussian distribution themselves, nor do they even need to have the same distribution. For a detailed development, see, e.g., Feller [1] or Billingsley [2]. A few comments are in order:

- There must be a large number of random variables that contribute to the sum. For instance, thermal noise is the result of the thermal vibrations of an astronomically large number of tiny electrons.
- The individual random variables in the sum must be independent, or nearly so.
- Each term in the sum must be small compared to the sum.

As one example, thermal noise results from the vibrations of a very large number of electrons, the vibration of any one electron is independent of that of another, and no one electron contributes significantly more than the others. Thus, all three conditions are satisfied and the noise is well modeled as Gaussian. Similarly, binomial probabilities approach the Gaussian. A binomial random variable is the sum of N independent Bernoulli (0 or 1) random variables. As N gets large, the distribution of the sum approaches a Gaussian distribution.

In Fig. 7.3 we see the effect of a small amount of Gaussian noise ($\sigma = 10$). Notice the "fuzziness" overall. It is often counterproductive to try to use signal processing techniques to remove this level of noise—the filtered image is usually visually less pleasing than the original noisy one (although sometimes the image is filtered to reduce the noise, then sharpened to eliminate the blurriness introduced by the noise reducing filter).

In Fig. 7.4, the noise has been increased by a factor of 3 ($\sigma = 30$). The degradation is much more objectionable. Various filtering techniques can improve the quality, though usually at the expense of some loss of sharpness.

7.4.2 Heavy Tailed Noise

In many situations, the conditions of the Central Limit Theorem are almost, but not quite, true. There may not be a large enough number of terms in the sum, or the terms

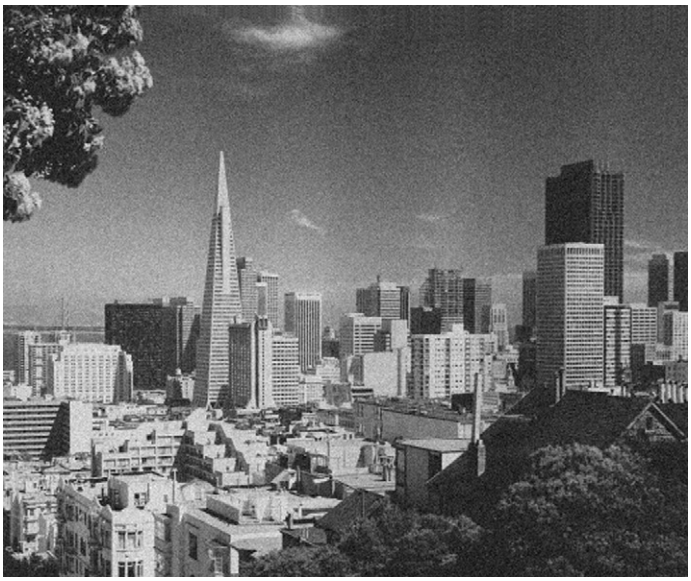

FIGURE 7.3

San Francisco corrupted by additive Gaussian noise with standard deviation equal to 10.

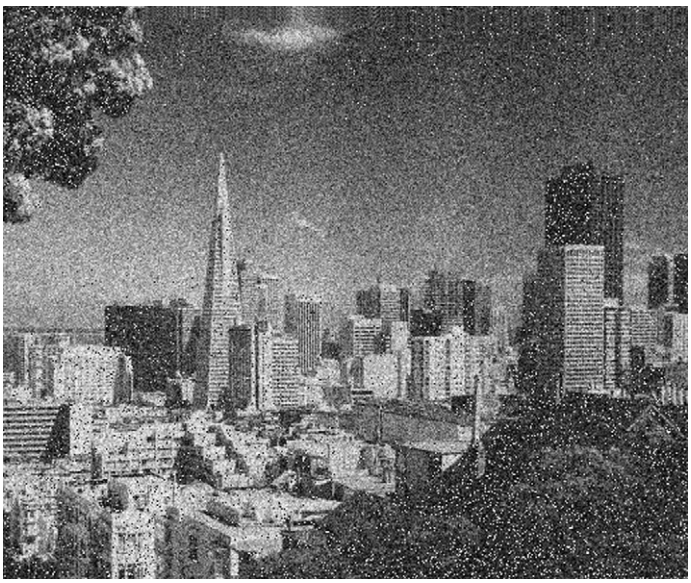

FIGURE 7.4

San Francisco corrupted by additive Gaussian noise with standard deviation equal to 30.

may not be sufficiently independent, or a small number of the terms may contribute a disproportionate amount to the sum. In these cases, the noise may only be approximately Gaussian. One should be careful. Even when the center of the density is approximately Gaussian, the tails may not be.

The *tails* of a distribution are the areas of the density corresponding to large x, i.e., as $|x| \to \infty$. A particularly interesting case is when the noise has *heavy tails*. "Heavy tails" means that for large values of x, the density, $p_{\mathbf{a}}(x)$, approaches 0 more slowly than the Gaussian. For example, for large values of x, the Gaussian density goes to 0 as $\exp(-x^2/2\sigma^2)$; the Laplacian density (also known as the double exponential density) goes to 0 as $\exp(-\lambda|x|)$. The Laplacian density is said to have heavy tails.

In Table 7.1, we present the tail probabilities, $\Pr[|\mathbf{x}| > x_0]$, for the "standard" Gaussian and Laplacian ($\mu = 0, \sigma = 1$, and $\lambda = 1$). Note the probability of exceeding 1 is approximately the same for both distributions, while the probability of exceeding 3 is about 20 times greater for the double exponential than for the Gaussian.

An interesting example of heavy tailed noise that should be familiar is static on a weak, broadcast AM radio station during a lightning storm. Most of the time, the

TABLE 7.1 Comparison of tail probabilities for the Gaussian and Laplacian distributions. Specifically, the values of $\Pr[|\mathbf{x}| > x_0]$ are listed for both distributions (with $\sigma = 1$ and $\lambda = 1$)

x_0	Gaussian	Laplacian
1	0.32	0.37
2	0.046	0.14
3	0.0027	0.05

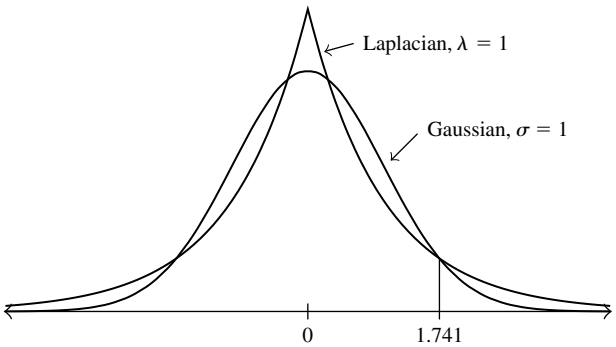

FIGURE 7.5

Comparison of the Laplacian ($\lambda = 1$) and Gaussian ($\sigma = 1$) densities, both with $\mu = 0$. Note, for deviations larger than 1.741, the Laplacian density is larger than the Gaussian.

conditions of the central limit theorem are well satisfied and the noise is Gaussian. Occasionally, however, there may be a lightning bolt. The lightning bolt overwhelms the tiny electrons and dominates the sum. During the time period of the lightning bolt, the noise is non-Gaussian and has much heavier tails than the Gaussian.

Some of the heavy tailed models that arise in image processing include the following:

7.4.2.1 Laplacian or Double Exponential

$$p_{\mathbf{a}}(x) = \frac{\lambda}{2} e^{-\lambda|x-\mu|} \tag{7.27}$$

The mean is μ and the variance is $2/\lambda^2$. The Laplacian is interesting in that the best estimate of μ is the median, not the mean, of the observations. Not truly "noise," the prediction error in many image compression algorithms is modeled as Laplacian. More simply, the difference between successive pixels is modeled as Laplacian.

7.4.2.2 Negative Exponential

$$p_{\mathbf{a}}(x) = \lambda e^{-\lambda x} \tag{7.28}$$

for $x > 0$. The mean is $1/\lambda > 0$ and variance, $1/\lambda^2$. The negative exponential is used to model speckle, for example, in SAR systems.

7.4.2.3 Alpha-Stable

In this class, appropriately normalized sums of independent and identically distributed random variables have the same distribution as the individual random variables. We have already seen that sums of Gaussian random variables are Gaussian, so the Gaussian is in the class of alpha-stable distributions. In general, these distributions have characteristic functions that look like $\exp(-|u|^\alpha)$ for $0 < \alpha \leq 2$. Unfortunately, except for the Gaussian ($\alpha = 2$) and the Cauchy ($\alpha = 1$), it is not possible to write the density functions of these distributions in closed form.

As $\alpha \to 0$, these distributions have very heavy tails.

7.4.2.4 Gaussian Mixture Models

$$p_{\mathbf{a}}(x) = (1-\alpha)p_{\mathbf{0}}(x) + \alpha p_{\mathbf{1}}(x), \tag{7.29}$$

where $p_{\mathbf{0}}(x)$ and $p_{\mathbf{1}}(x)$ are Gaussian densities with differing means, μ_0 and μ_1, or variances, σ_0^2 and σ_1^2. In modeling heavy tailed distributions, it is often true that α is small, say $\alpha = 0.05$, $\mu_0 = \mu_1$, and $\sigma_1^2 \gg \sigma_0^2$.

In the "static in the AM radio" example above, at any given time, α would be the probability of a lightning strike, σ_0^2 the average variance of the thermal noise, and σ_1^2 the variance of the lightning induced signal.

Sometimes this model is generalized further and $p_{\mathbf{1}}(x)$ is allowed to be non-Gaussian (and sometimes completely arbitrary). See Huber [11].

7.4.2.5 Generalized Gaussian

$$p_{\mathbf{a}}(x) = Ae^{-\beta|x-\mu|^{\alpha}}, \tag{7.30}$$

where μ is the mean and A, β, and α are constants. α determines the shape of the density: $\alpha = 2$ corresponds to the Gaussian and $\alpha = 1$ to the double exponential. Intermediate values of α correspond to densities that have tails in between the Gaussian and double exponential. Values of $\alpha < 1$ give even heavier tailed distributions.

The constants, A and β, can be related to α and the standard deviation, σ, as follows:

$$\beta = \frac{1}{\sigma}\left(\frac{\Gamma(3/\alpha)}{\Gamma(1/\alpha)}\right)^{0.5} \tag{7.31}$$

$$A = \frac{\beta\alpha}{2\Gamma(1/\alpha)}. \tag{7.32}$$

The generalized Gaussian has the advantage of being able to fit a large variety of (symmetric) noises by appropriate choice of the three parameters, μ, σ, and α [12].

One should be careful to use estimators that behave well in heavy tailed noise. The sample mean, optimal for a constant signal in additive Gaussian noise, can perform quite poorly in heavy tailed noise. Better choices are those estimators designed to be robust against the occasional outlier [11]. For instance, the median is only slightly worse than the mean in Gaussian noise, but can be much better in heavy tailed noise.

7.4.3 Salt and Pepper Noise

Salt and pepper noise refers to a wide variety of processes that result in the same basic image degradation: only a few pixels are noisy, but they are *very* noisy. The effect is similar to sprinkling white and black dots—salt and pepper—on the image.

One example where salt and pepper noise arises is in transmitting images over noisy digital links. Let each pixel be quantized to B bits in the usual fashion. The value of the pixel can be written as $X = \sum_{i=0}^{B-1} b_i 2^i$. Assume the channel is a binary symmetric one with a crossover probability of ϵ. Then each bit is flipped with probability ϵ. Call the received value, Y. Then, assuming the bit flips are independent,

$$\Pr\left[|X - Y| = 2^i\right] = \epsilon(1-\epsilon)^{B-1} \tag{7.33}$$

for $i = 0, 1, \ldots, B-1$. The MSE due to the most significant bit is $\epsilon 4^{B-1}$ compared to $\epsilon(4^{B-1} - 1)/3$ for all the other bits combined. In other words, the contribution to the MSE from the most significant bit is approximately three times that of all the other bits. The pixels whose most significant bits are changed will likely appear as black or white dots.

Salt and pepper noise is an example of (very) heavy tailed noise. A simple model is the following: Let $f(x,y)$ be the original image and $\mathbf{q}(x,y)$ be the image after it has been

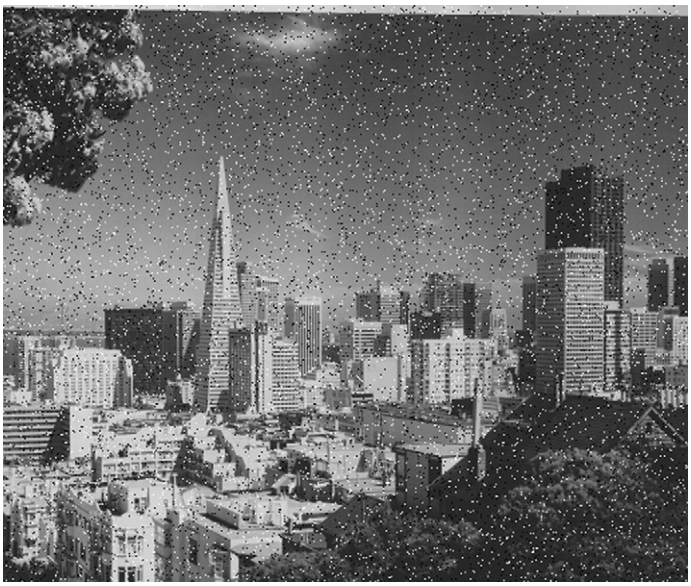

FIGURE 7.6

San Francisco corrupted by salt and pepper noise with a probability of occurrence of 0.05.

altered by salt and pepper noise.

$$\Pr[\mathbf{q} = f] = 1 - \alpha \tag{7.34}$$

$$\Pr[\mathbf{q} = \text{MAX}] = \alpha/2 \tag{7.35}$$

$$\Pr[\mathbf{q} = \text{MIN}] = \alpha/2, \tag{7.36}$$

where MAX and MIN are the maximum and minimum image values, respectively. For 8 bit images, MIN = 0 and MAX = 255. The idea is that with probability $1 - \alpha$ the pixels are unaltered; with probability α the pixels are changed to the largest or smallest values. The altered pixels look like black and white dots sprinkled over the image.

Figure 7.6 shows the effect of salt and pepper noise. Approximately 5% of the pixels have been set to black or white (95% are unchanged). Notice the sprinkling of the black and white dots. Salt and pepper noise is easily removed with various order statistic filters, especially the center weighted median and the LUM filter [13].

7.4.4 Quantization and Uniform Noise

Quantization noise results when a continuous random variable is converted to a discrete one or when a discrete random variable is converted to one with fewer levels. In images, quantization noise often occurs in the acquisition process. The image may be continuous initially, but to be processed it must be converted to a digital representation.

As we shall see, quantization noise is usually modeled as uniform. Various researchers use uniform noise to model other impairments, e.g., dither signals. Uniform noise is the opposite of the heavy tailed noise discussed above. Its tails are very light (zero!).

Let $\mathbf{b} = Q(\mathbf{a}) = \mathbf{a} + \mathbf{q}$, where $-\Delta/2 \leq \mathbf{q} \leq \Delta/2$ is the quantization noise and $\mathbf{b}$ is a discrete random variable usually represented with β bits. In the case where the number of quantization levels is large (so Δ is small), $\mathbf{q}$ is usually modeled as being uniform between $-\Delta/2$ and $\Delta/2$ and independent of $\mathbf{a}$. The mean and variance of $\mathbf{q}$ are

$$E[\mathbf{q}] = \frac{1}{\Delta}\int_{-\Delta/2}^{\Delta/2} s\,ds = 0 \tag{7.37}$$

and

$$E\Big[(\mathbf{q} - E[\mathbf{q}])^2\Big] = \frac{1}{\Delta}\int_{-\Delta/2}^{\Delta/2} s^2\,ds = \Delta^2/12. \tag{7.38}$$

Since $\Delta \sim 2^{-\beta}$, $\sigma_\nu^2 \sim 2^{2\beta}$, the signal-to-noise ratio increases by 6 dB for each additional bit in the quantizer.

When the number of quantization levels is small, the quantization noise becomes signal dependent. In an image of the noise, signal features can be discerned. Also, the noise is correlated on a pixel by pixel basis and not uniformly distributed.

The general appearance of an image with too few quantization levels may be described as "scalloped." Fine graduations in intensities are lost. There are large areas of constant color separated by clear boundaries. The effect is similar to transforming a smooth ramp into a set of discrete steps.

In Fig. 7.7, the San Francisco image has been quantized to only 4 bits. Note the clear "stair-stepping" in the sky. The previously smooth gradations have been replaced by large constant regions separated by noticeable discontinuities.

7.4.5 Photon Counting Noise

Fundamentally, most image acquisition devices are photon counters. Let $\mathbf{a}$ denote the number of photons counted at some location (a pixel) in an image. Then, the distribution of $\mathbf{a}$ is usually modeled as Poisson with parameter λ. This noise is also called *Poisson noise* or Poisson counting noise.

$$P(\mathbf{a} = k) = \frac{e^{-\lambda}\lambda^k}{k!} \tag{7.39}$$

for $k = 0, 1, 2, \ldots$

The Poisson distribution is one for which calculating moments by using the characteristic function is much easier than by the usual sum.

FIGURE 7.7
San Francisco quantized to 4 bits.

$$\Phi(u) = \sum_{k=0}^{\infty} \frac{e^{juk} e^{-\lambda} \lambda^k}{k!} \tag{7.40}$$

$$= e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda e^{ju})^k}{k!} \tag{7.41}$$

$$= e^{-\lambda} e^{\lambda e^{ju}} \tag{7.42}$$

$$= e^{\lambda(e^{ju}-1)}. \tag{7.43}$$

While this characteristic function does not *look* simple, it does yield the moments:

$$E[\mathbf{a}] = \frac{1}{j} \frac{d}{du} e^{\lambda(e^{ju}-1)} \Big|_{u=0} \tag{7.44}$$

$$= \frac{1}{j} \lambda j e^{ju} e^{\lambda(e^{ju}-1)} \Big|_{u=0} \tag{7.45}$$

$$= \lambda. \tag{7.46}$$

Similarly, $E\left[\mathbf{a}^2\right] = \lambda + \lambda^2$ and $\sigma^2 = (\lambda + \lambda^2) - \lambda^2 = \lambda$. We see one of the most interesting properties of the Poisson distribution, that the variance is equal to the expected value.

When λ is large, the central limit theorem can be invoked and the Poisson distribution is well approximated by the Gaussian with mean and variance both equal to λ.

Consider two different regions of an image, one brighter than the other. The brighter one has a higher λ and therefore a higher noise variance.

As another example of Poisson counting noise, consider the following:

Example: Effect of Shutter Speed on Image Quality Consider two pictures of the same scene, one taken with a shutter speed of 1 unit time and the other with $\Delta > 1$ unit of time. Assume that an area of an image emits photons at the rate λ per unit time. The first camera measures a random number of photons, whose expected value is λ and whose variance is also λ. The second, however, has an expected value and variance equal to $\lambda\Delta$. When time averaged (divided by Δ), the second now has an expected value of λ and a variance of $\lambda/\Delta < \lambda$. Thus, we are led to the intuitive conclusion: all other things being equal, slower shutter speeds yield better pictures.

For example, astro-photographers traditionally used long exposures to average over a long enough time to get good photographs of faint celestial objects. Today's astronomers use CCD arrays and average many short exposure photographs, but the principal is the same.

Figure 7.8 shows the image with Poisson noise. It was constructed by taking each pixel value in the original image and generating a Poisson random variable with λ equal to that value. Careful examination reveals that the white areas are noisier than the dark areas. Also, compare this image with Fig. 7.3 which shows Gaussian noise of almost the same power.

FIGURE 7.8

San Francisco corrupted by Poisson noise.

7.4.6 Photographic Grain Noise

Photographic grain noise is a characteristic of photographic films. It limits the effective magnification one can obtain from a photograph. A simple model of the photography process is as follows:

A photographic film is made up from millions of tiny *grains*. When light strikes the film, some of the grains absorb the photons and some do not. The ones that do change their appearance by becoming metallic silver. In the developing process, the unchanged grains are washed away.

We will make two simplifying assumptions: (1) the grains are uniform in size and character and (2) the probability that a grain changes is proportional to the number of photons incident upon it. Both assumptions can be relaxed, but the basic answer is the same. In addition, we will assume the grains are independent of each other.

Slow film has a large number of small fine grains, while fast film has a smaller number of larger grains. The small grains give slow film a better, less grainy picture; the large grains in fast film cause a grainier picture.

In a given area, A, assume there are L grains, with the probability of each grain changing, p, proportionate to the number of incident photons. Then the number of grains that change, $\mathbf{N}$, is binomial

$$\Pr[\mathbf{N} = k] = \binom{L}{k} p^k (1-p)^{L-k}. \tag{7.47}$$

Since L is large, when p small but $\lambda = Lp = E[\mathbf{N}]$ moderate, this probability is well approximated by a Poisson distribution

$$\Pr[\mathbf{N} = k] = \frac{e^{-\lambda}\lambda^k}{k!} \tag{7.48}$$

and by a Gaussian when p is larger:

$$\begin{aligned} &\Pr[k \le \mathbf{N} < k + \Delta_k] \\ &\quad = \Pr\left[\frac{k - Lp}{\sqrt{Lp(1-p)}} \le \frac{\mathbf{N} - Lp}{\sqrt{Lp(1-p)}} \le \frac{k + \Delta_k - Lp}{\sqrt{Lp(1-p)}}\right] \qquad (7.49) \\ &\quad \approx e^{-0.5\left(\frac{k-Lp}{Lp(1-p)}\right)^2} \Delta_k \qquad (7.50) \end{aligned}$$

The probability interval on the right-hand side of (7.49) is exactly the same as that on the left except that it has been normalized by subtracting the mean and dividing by the standard deviation. (7.50) results from (7.49) by applying the central limit theorem. In other words, the distribution of grains that change is approximately Gaussian with mean Lp and variance $Lp(1-p)$. This variance is maximized when $p = 0.5$. Sometimes, however, it is sufficiently accurate to ignore this variation and model grain noise as additive Gaussian with a constant noise power.

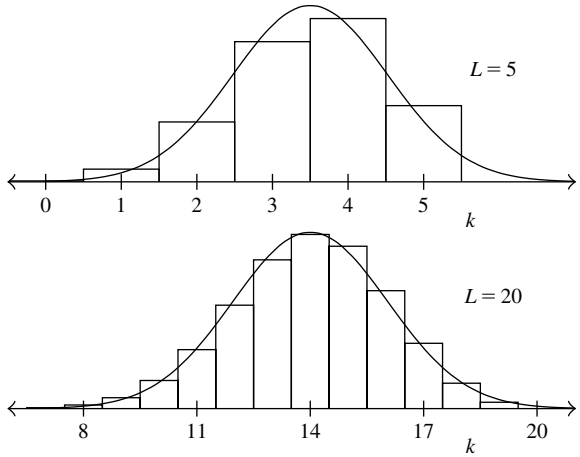

FIGURE 7.9

Illustration of the Gaussian approximation to the binomial. In both figures, $p = 0.7$ and the Gaussians have the same means and variances as the binomials. Even for L as small as 5, the Gaussian reasonably approximates the binomial PMF. For $L = 20$, the approximation is very good.

7.5 CCD IMAGING

In the past 20 years or so, CCD (charge-coupled devices) imaging has replaced photographic film as the dominant imaging form. First CCDs appeared in scientific applications, such as astronomical imaging and microscopy. Recently, CCD digital cameras and videos have become widely used consumer items. In this section, we analyze the various noise sources affecting CCD imagery.

CCD arrays work on the photoelectric principle (first discovered by Hertz and explained by Einstein, for which he was awarded the Nobel prize). Incident photons are absorbed, causing electrons to be elevated into a high energy state. These electrons are captured in a well. After some time, the electrons are counted by a "read out" device.

The number of electrons counted, N, can be written as

$$N = N_I + N_{th} + N_{ro}, \tag{7.51}$$

where N_I is the number of electrons due to the image, N_{th} the number due to thermal noise, and N_{ro} the number due to read out effects.

N_I is Poisson, with the expected value $E[N_I] = \lambda$ proportional to the incident image intensity. The variance of N_I is also λ, thus the standard deviation is $\sqrt{\lambda}$. The signal-to-noise ratio (neglecting the other noises) is $\lambda/\sqrt{\lambda} = \sqrt{\lambda}$. The only way to increase the signal-to-noise ratio is to increase the number of electrons recorded. Sometimes the image intensity can be increased (e.g., a photographer's flash), the aperature increased

(e.g., a large telescope), or the exposure time increased. However, CCD arrays saturate: only a finite number of electrons can be captured. The effect of long exposures is achieved by averaging many short exposure images.

Even without incident photons, some electrons obtain enough energy to get captured. This is due to thermal effects and is called thermal noise or *dark current*. The amount of thermal noise is proportional to the temperature, T, and the exposure time. N_{th} is modeled as Gaussian.

The read out process introduces its own uncertainties and can inject electrons into the count. Read out noise is a function of the read out process and is independent of the image and the exposure time. Like image noise, N_{ro} is modeled as Poisson noise.

There are two different regimes in which CCD imaging is used: low light and high light levels. In low light, the number of image electrons is small. In this regime, thermal noise and read out noise are both significant and can dominate the process. For instance, much scientific and astronomical imaging is in low light. Two important steps are taken to reduce the effects of thermal and read out noise. The first is obvious: since thermal noise increases with temperature, the CCD is cooled as much as practicable. Often liquid nitrogen is used to lower the temperature.

The second is to estimate the means of the two noises and subtract them from measured image. Since the two noises arise from different effects, the means are measured separately. The mean of the thermal noise is measured by averaging several images taken with the shutter closed, but with the same shutter speed and temperature. The mean of the read out noise is estimated by taking the median of several (e.g., 9) images taken with the shutter closed and a zero exposure time (so that any signal measured is due to read out effects).

In high light levels, the image noise dominates and thermal and read out noises can be ignored. This is the regime in which consumer imaging devices are normally used. For large values of N_I, the Poisson distribution is well modeled as Gaussian. Thus the overall noise looks Gaussian, but the signal-to-noise ratio is higher in bright regions than in dark regions.

7.6 SPECKLE

In this section, we discuss two kinds of *speckle*, a curious distortion in images created by coherent light or by atmospheric effects. Technically not noise in the same sense as other noise sources considered so far, speckle is noise-like in many of its characteristics.

7.6.1 Speckle in Coherent Light Imaging

Speckle is one of the more complex image noise models. It is signal dependent, non-Gaussian, and spatially dependent. Much of this discussion is taken from [14, 15]. We will first discuss the origins of speckle, then derive the first-order density of speckle, and conclude this section with a discussion of the second-order properties of speckle.

In coherent light imaging, an object is illuminated by a coherent source, usually a laser or a radar transmitter. For the remainder of this discussion, we will consider the illuminant to be a light source, e.g., a laser, but the principles apply to radar imaging as well.

When coherent light strikes a surface, it is reflected back. Due to the microscopic variations in the surface roughness within one pixel, the received signal is subjected to random variations in phase and amplitude. Some of these variations in phase add constructively, resulting in strong intensities, and others add deconstructively, resulting in low intensities. This variation is called *speckle*.

Of crucial importance in the understanding of speckle is the point spread function of the optical system. There are three regimes:

- The point spread function is so narrow that the individual variations in surface roughness can be resolved. The reflections off the surface are random (if, indeed, we can model the surface roughness as random in this regime), but we cannot appeal to the central limit theorem to argue that the reflected signal amplitudes are Gaussian. Since this case is uncommon in most applications, we will ignore it.
- The point spread function is broad compared to the feature size of the surface roughness, but small compared to the features of interest in the image. This is a common case and leads to the conclusion, presented below, that the noise is exponentially distributed and uncorrelated on the scale of the features in the image. Also, in this situation, the noise is often modeled as multiplicative.
- The point spread function is broad compared to both the feature size of the object and the feature size of the surface roughness. Here, the speckle is correlated and its size distribution is interesting and is determined by the point spread function.

The development will proceed in two parts. Firstly, we will derive the first-order probability density of speckle and, secondly, we will discuss the correlation properties of speckle.

In any given macroscopic area, there are many microscopic variations in the surface roughness. Rather than trying to characterize the surface, we will content ourselves with finding a statistical description of the speckle.

We will make the (standard) assumptions that the surface is very rough on the scale of the optical wavelengths. This roughness means that each microscopic reflector in the surface is at a random height (distance from the observer) and a random orientation with respect to the incoming polarization field. These random reflectors introduce random changes in the reflected signal's amplitude, phase, and polarization. Further, we assume these variations at any given point are independent from each other and independent from the changes at any other point.

These assumptions amount to assuming that the system cannot resolve the variations in roughness. This is generally true in optical systems, but may not be so in some radar applications.

The above assumptions on the physics of the situation can be translated to statistical equivalents: the amplitude of the reflected signal at any point, (x,y), is multiplied by a random amplitude, denoted $\mathbf{a}(x,y)$, and the polarization, $\phi(x,y)$, is uniformly distributed between 0 and 2π.

Let $\mathbf{u}(x,y)$ be the complex phasor of the incident wave at a point (x,y), $\mathbf{v}(x,y)$ be the reflected signal, and $\mathbf{w}(x,y)$ be the received phasor. From the above assumptions,

$$\mathbf{v}(x,y) = \mathbf{u}(x,y)\mathbf{a}(x,y)e^{j\phi(x,y)} \tag{7.52}$$

and, letting $h(\cdot,\cdot)$ denote the 2D point spread function of the optical system,

$$\mathbf{w}(x,y) = h(x,y) * \mathbf{v}(x,y). \tag{7.53}$$

One can convert the phasors to rectangular coordinates:

$$\mathbf{v}(x,y) = \mathbf{v}_R(x,y) + j\mathbf{v}_I(x,y) \tag{7.54}$$

and

$$\mathbf{w}(x,y) = \mathbf{w}_R(x,y) + j\mathbf{w}_I(x,y). \tag{7.55}$$

Since the change in polarization is uniform between 0 and 2π, $\mathbf{v}_R(x,y)$ and $\mathbf{v}_I(x,y)$ are statistically independent. Similarly, $\mathbf{w}_R(x,y)$ and $\mathbf{w}_I(x,y)$ are statistically independent. Thus,

$$\mathbf{w}_R(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(\alpha,\beta)v_R(x-\alpha,y-\beta)\,d\alpha\,d\beta \tag{7.56}$$

and similarly for $\mathbf{w}_I(x,y)$.

The integral in (7.56) is basically a sum over many tiny increments in x and y. By assumption, the increments are independent of one another. Thus, we can appeal to the central limit theorem and conclude that the distributions of $\mathbf{w}_R(x,y)$ and $\mathbf{w}_I(x,y)$ are each Gaussian with mean 0 and variance σ^2. Note, this conclusion does not depend on the details of the roughness, as long as the surface is rough on the scale of the wavelength of the incident light and the optical system cannot resolve the individual components of the surface.

The measured intensity, $\mathbf{f}(x,y)$, is the squared magnitude of the received phasors:

$$\mathbf{f}(x,y) = \mathbf{w}_R(x,y)^2 + \mathbf{w}_I(x,y)^2. \tag{7.57}$$

The distribution of $\mathbf{f}$ can be found by integrating the joint density of $\mathbf{w}_R$ and $\mathbf{w}_I$ over a circle of radius $f^{0.5}$:

$$\Pr[\mathbf{f}(x,y) \le f] = \int_0^{2\pi}\int_0^{f^{0.5}} \frac{1}{2\pi\sigma^2} e^{-\rho/2\sigma^2}\rho\,d\rho\,d\phi \tag{7.58}$$

$$= 1 - e^{-f/2\sigma^2}. \tag{7.59}$$

The corresponding density is $p_{\mathbf{f}}(f)$:

$$p_{\mathbf{f}}(f) = \begin{cases} \frac{1}{g} e^{-f/g} & f \geq 0 \\ 0 & f < 0, \end{cases} \tag{7.60}$$

where we have taken the liberty to introduce the mean intensity, $g = g(x,y) = 2\sigma^2(x,y)$. A little rearrangement can put this into a multiplicative noise model:

$$\mathbf{f}(x,y) = g(x,y)\mathbf{q}, \tag{7.61}$$

where $\mathbf{q}$ has a exponential density

$$p_{\mathbf{q}}(x) = \begin{cases} e^{-x} & x \geq 0 \\ 0 & x < 0. \end{cases} \tag{7.62}$$

The mean of $\mathbf{q}$ is 1 and the variance is 1.

The exponential density is much heavier tailed than the Gaussian density, meaning that much greater excursions from the mean occur. In particular, the standard deviation of $\mathbf{f}$ equals $E[\mathbf{f}]$, i.e., the typical deviation in the reflected intensity is equal to the typical intensity. It is this large variation that causes speckle to be so objectionable to human observers.

It is sometimes possible to obtain multiple images of the same scene with independent realizations of the speckle pattern, i.e., the speckle in any one image is independent of the speckle in the others. For instance, there may be multiple lasers illuminating the same object from different angles or with different optical frequencies. One means of speckle reduction is to average these images:

$$\hat{\mathbf{f}}(x,y) = \frac{1}{M} \sum_{i=1}^{M} \mathbf{f}_i(x,y) \tag{7.63}$$

$$= g(x,y) \frac{\sum_{i=1}^{M} \mathbf{q}_i(x,y)}{M}. \tag{7.64}$$

Now, the average of the negative exponentials has mean 1 (the same as each individual negative exponential) and variance $1/M$. Thus, the average of the speckle images has a mean equal to $g(x,y)$ and variance $g^2(x,y)/M$.

Figure 7.10 shows an uncorrelated speckle image of San Francisco. Notice how severely degraded this image is. Careful examination will show that the light areas are noisier than the dark areas. This image was created by generating an "image" of exponential variates and multiplying each by the corresponding pixel value. Intensity values beyond 255 were truncated to 255.

The correlation structure of speckle is largely determined by the width of the point spread function. As above the real and imaginary components (or, equivalently, the X and Y components) of the reflected wave are independent Gaussian. These components ($\mathbf{w_R}$ and $\mathbf{w_I}$ above) are individually filtered by the point spread function of the imaging

FIGURE 7.10
San Francisco with uncorrelated speckle.

system. The intensity image is formed by taking the complex magnitude of the resulting filtered components.

Figure 7.11 shows a correlated speckle image of San Francisco. The image was created by filtering $\mathbf{w_R}$ and $\mathbf{w_I}$ with a 2D square filter of size 5×5. This size filter is too big for the fine details in the original image, but is convenient to illustrate the correlated speckle. As above, intensity values beyond 255 were truncated to 255. Notice the correlated structure to the "speckles." The image has a pebbly appearance.

We will conclude this discussion with a quote from Goodman [16]:

> *The general conclusions to be drawn from these arguments are that, in any speckle pattern, large-scale-size fluctuations are the most populous, and no scale sizes are present beyond a certain small-size cutoff. The distribution of scale sizes in between these limits depends on the autocorrelation function of the object geometry, or on the autocorrelation function of the pupil function of the imaging system in the imaging geometry.*

7.6.2 Atmospheric Speckle

The twinkling of stars is similar in cause to speckle in coherent light, but has important differences. Averaging multiple frames of independent coherent imaging speckle results in an image estimate whose mean equals the underlying image and whose variance is reduced by the number of frames averaged over. However, averaging multiple images of twinkling stars results in a blurry image of the star.

FIGURE 7.11

San Francisco with correlated speckle.

From the earth, stars (except the Sun!) are point sources. Their light is spatially coherent and planar when it reaches the atmosphere. Due to thermal and other variations, the diffusive properties of the atmosphere changes in an irregular way. This causes the index of refraction to change randomly. The star appears to twinkle. If one averages multiple images of the star, one obtains a blurry image.

Until recently, the preferred way to eliminate atmospheric-induced speckle (the "twinkling") was to move the observer to a location outside the atmosphere, i.e., in space. In recent years, new techniques to estimate and track the fluctuations in atmospheric conditions have allowed astronomers to take excellent pictures from the earth. One class is called "speckle interferometry" [17]. It uses multiple short duration (typically less than 1 second each) images and a nearby star to estimate the random speckle pattern. Once estimated, the speckle pattern can be removed, leaving the unblurred image.

7.7 CONCLUSIONS

In this chapter, we have tried to summarize the various image noise models and give some recommendations for minimizing the noise effects. Any such summary is, by necessity, limited. We do, of course, apologize to any authors whose work we may have omitted.

For further information, the interested reader is urged to consult the references for this and other chapters.

REFERENCES

[1] W. Feller. *An Introduction to Probability Theory and its Applications*. J. Wiley & Sons, New York, 1968.

[2] P. Billingsley. *Probability and Measure*. J. Wiley & Sons, New York, 1979.

[3] M. Woodroofe. *Probability with Applications*. McGraw-Hill, New York, 1975.

[4] C. Helstrom. *Probability and Stochastic Processes for Engineers*. Macmillan, New York, 1991.

[5] E. H. Lloyd. Least-squares estimations of location and scale parameters using order statistics. *Biometrika*, 39:88–95, 1952.

[6] A. C. Bovik, T. S. Huang, and D. C. Munson, Jr. A generalization of median filtering using linear combinations of order statistics. *IEEE Trans. Acoust.*, ASSP-31(6):1342–1350, 1983.

[7] R. C. Hardie and C. G. Boncelet, Jr. LUM filters: a class of order statistic based filters for smoothing and sharpening. *IEEE Trans. Signal Process.*, 41(3):1061–1076, 1993.

[8] C. G. Boncelet, Jr. Algorithms to compute order statistic distributions. *SIAM J. Sci. Stat. Comput.*, 8(5):868–876, 1987.

[9] C. G. Boncelet, Jr. Order statistic distributions with multiple windows. *IEEE Trans. Inf. Theory*, IT-37(2):436–442, 1991.

[10] P. Peebles. *Probability, Random Variables, and Random Signal Principles*. McGraw Hill, New York, 1993.

[11] P. J. Huber. *Robust Statistics*. J. Wiley & Sons, New York, 1981.

[12] J. H. Miller and J. B. Thomas. Detectors for discrete-time signals in non-Gaussian noise. *IEEE Trans. Inf. Theory*, IT-18(2):241–250, 1972.

[13] J. Astola and P. Kuosmanen. *Fundamentals of Nonlinear Digital Filtering*. CRC Press, Boca Raton, FL, 1997.

[14] D. Kuan, A. Sawchuk, T. Strand, and P. Chavel. Adaptive restoration of images with speckle. *IEEE Trans. Acoust.*, ASSP-35(3):373–383, 1987.

[15] J. Goodman. *Statistical Optics*. Wiley-Interscience, New York, 1985.

[16] J. Goodman. Some fundamental properties of speckle. *J. Opt. Soc. Am.*, 66:1145–1150, 1976.

[17] A. Labeyrie. Attainment of diffraction limited resolution in large telescopes by fourier analysis speckle patterns in star images. *Astron. Astrophys.*, VI:85–87, 1970.