CHAPTER 5

IMAGE RESTORATION

> Things which we see are not by themselves what we see. . . . It remains completely unknown to us what the objects may be by themselves and apart from the receptivity of our senses. We know nothing but our manner of perceiving them. . . .
>
> *Immanuel Kant*

As in image enhancement, the ultimate goal of restoration techniques is to improve an image in some sense. For the purpose of differentiation, we consider restoration to be a process that attempts to reconstruct or recover an image that has been degraded by using some a priori knowledge of the degradation phenomenon. Thus restoration techniques are oriented toward modeling the degradation and applying the inverse process in order to recover the original image. This approach usually involves formulating a criterion of goodness that will yield some optimal estimate of the desired result. By contrast, enhancement techniques basically are heuristic procedures designed to manipulate an image in order to take advantage of the psychophysical aspects of the human visual system. For example, contrast stretching is considered an enhancement technique because it is based primarily on the pleasing aspects it might present to the viewer, whereas removal of image blur by applying a deblurring function is considered a restoration technique.

Early techniques for digital image restoration were derived mostly from frequency domain concepts. However, this chapter focuses on a more modern, algebraic approach, which has the advantage of allowing the derivation of numerous restoration techniques from the same basic principles. Although a direct solution by algebraic methods generally involves the manipulation of large systems of simultaneous equations, we show that, under certain conditions, computational complexity can be reduced to the same level as that required by traditional frequency domain restoration techniques.

The material developed in this chapter is strictly introductory. We consider the restoration problem only from the point where a degraded, *digital* image is given; thus we do not consider topics dealing with sensor, digitizer, and display degradations. These subjects, although of importance in the overall treatment of image restoration applications, are beyond the present discussion. The references cited at the end of the chapter provide a guide to the voluminous literature on these and related topics.

5.1 DEGRADATION MODEL

As Fig. 5.1 shows, the degradation process is modeled in this chapter as an operator (or system) H, which together with an additive noise term $\eta(x, y)$ operates on an input image $f(x, y)$ to produce a degraded image $g(x, y)$. Digital image restoration may be viewed as the process of obtaining an approximation to $f(x, y)$, given $g(x, y)$ and a knowledge of the degradation in the form of the operator H. We assume that knowledge of $\eta(x, y)$ is limited to information of a statistical nature.

5.1.1 Some Definitions

The input–output relationship in Fig. 5.1 is expressed as

$$g(x, y) = H[f(x, y)] + \eta(x, y). \tag{5.1-1}$$

For the moment, let us assume that $\eta(x, y) = 0$ so that $g(x, y) = H[f(x, y)]$. Then H is *linear* if

$$H[k_1 f_1(x, y) + k_2 f_2(x, y)] = k_1 H[f_1(x, y)] + k_2 H[f_2(x, y)] \tag{5.1-2}$$

where k_1 and k_2 are constants and $f_1(x, y)$ and $f_2(x, y)$ are any two input images.

If $k_1 = k_2 = 1$, Eq. (5.1-2) becomes

$$H[f_1(x, y) + f_2(x, y)] = H[f_1(x, y)] + H[f_2(x, y)] \tag{5.1-3}$$

which is called the property of *additivity*; this property simply says that, if H is a linear operator, the response to a sum of two inputs is equal to the sum of the two responses.

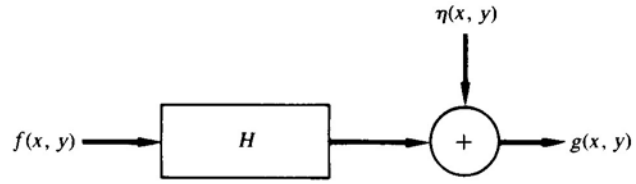


Figure 5.1 *A model of the image degradation process.*

With $f_2(x, y) = 0$, Eq. (5.1-2) becomes

$$H[k_1 f_1(x, y)] = k_1 H[f_1(x, y)] \tag{5.1-4}$$

which is called the property of *homogeneity*. It says that the response to a constant multiple of any input is equal to the response to that input multiplied by the same constant. Thus a linear operator possesses both the property of additivity and the property of homogeneity.

An operator having the input–output relationship $g(x, y) = H[f(x, y)]$ is said to be *position* (or *space*) *invariant* if

$$H[f(x - \alpha, y - \beta)] = g(x - \alpha, y - \beta) \tag{5.1-5}$$

for any $f(x, y)$ and any α and β. This definition indicates that the response at any point in the image depends only on the value of the input at that point and not on the position of the point.

5.1.2 Degradation Model for Continuous Functions

With a slight (but equivalent) change in notation in the definition of the impulse function, Eq. (3.3-46), $f(x, y)$ can be expressed in the form

$$f(x, y) = \int_{-\infty}^{\infty}\int f(\alpha, \beta)\delta(x - \alpha, y - \beta)\, d\alpha\, d\beta. \tag{5.1-6}$$

Then, if $\eta(x, y) = 0$ in Eq. (5.1-1),

$$g(x, y) = H[f(x, y)] = H\left[\int_{-\infty}^{\infty}\int f(\alpha, \beta)\delta(x - \alpha, y - \beta)\, d\alpha\, d\beta\right]. \tag{5.1-7}$$

If H is a linear operator and we extend the additivity property to integrals, then

$$g(x, y) = \int_{-\infty}^{\infty}\int H[f(\alpha, \beta)\delta(x - \alpha, y - \beta)]\, d\alpha\, d\beta. \tag{5.1-8}$$

Since $f(\alpha, \beta)$ is independent of x and y, and from the homogeneity property,

$$g(x, y) = \int_{-\infty}^{\infty}\int f(\alpha, \beta)H[\delta(x - \alpha, y - \beta)]\, d\alpha\, d\beta. \tag{5.1-9}$$

The term

$$h(x, \alpha, y, \beta) = H[\delta(x - \alpha, y - \beta)] \tag{5.1-10}$$

is called the *impulse response* of H. In other words, if $\eta(x, y) = 0$ in Eq. (5.1-1), then $h(x, \alpha, y, \beta)$ is the response of H to an impulse of strength 1 at coordinates (α, β). In optics, the impulse becomes a point of light and $h(x, \alpha, y, \beta)$ is commonly referred to in this case as the *point spread function* (PSF), as discussed in Section 4.1.2.

Substituting Eq. (5.1-10) into Eq. (5.1-9) yields the expression

$$g(x, y) = \int_{-\infty}^{\infty}\int f(\alpha, \beta)h(x, \alpha, y, \beta)\, d\alpha\, d\beta \tag{5.1-11}$$

which is called the *superposition* (or *Fredholm*) *integral of the first kind*. This expression is of fundamental importance in linear system theory. It states that if the response of H to an impulse is known, the response to any input $f(\alpha, \beta)$ can be calculated by means of Eq. (5.1-11). In other words, a linear system H is completely characterized by its impulse response.

If H is position invariant, from Eq. (5.1-5),

$$H[\delta(x - \alpha, y - \beta)] = h(x - \alpha, y - \beta). \tag{5.1-12}$$

Equation (5.1-11) reduces in this case to

$$g(x, y) = \int_{-\infty}^{\infty}\int f(\alpha, \beta)h(x - \alpha, y - \beta)\, d\alpha\, d\beta \tag{5.1-13}$$

which is the convolution integral defined in Eq. (3.3-30).

In the presence of additive noise the expression describing a linear degradation model becomes

$$g(x, y) = \int_{-\infty}^{\infty}\int f(\alpha, \beta)h(x, \alpha, y, \beta)\, d\alpha\, d\beta + \eta(x, y). \tag{5.1-14}$$

If H is position invariant, Eq. (5.1-14) becomes

$$g(x, y) = \int_{-\infty}^{\infty}\int f(\alpha, \beta)h(x - \alpha, y - \beta)\, d\alpha\, d\beta + \eta(x, y). \tag{5.1-15}$$

The noise, of course, is assumed in both cases to be independent of position in the image.

Many types of degradations can be approximated by linear, position invariant processes. The advantage of this approach is that the extensive tools of linear system theory then become available for the solution of image restoration problems. Nonlinear and space variant techniques, although more general (and usually more accurate), introduce difficulties that often have no known solution or are very difficult to solve computationally. This chapter focuses on linear, space invariant restoration techniques. However, even this simplification can result in computational problems that, if attacked directly, are beyond the practical capabilities of most present-day computers.

5.1.3 Discrete Formulation

The development of a discrete, space invariant degradation model is simplified by starting with the 1-D case and temporarily neglecting the noise term. Suppose that two functions $f(x)$ and $h(x)$ are sampled uniformly to form arrays of dimensions A and B, respectively. In this case, x is a discrete variable in the ranges, $0, 1, 2, \ldots, A - 1$ for $f(x)$ and $0, 1, 2, \ldots, B - 1$ for $h(x)$.

The discrete convolution formulation given in Section 3.3.8 is based on the assumption that the sampled functions are periodic, with a period M. Overlap in the individual periods of the resulting convolution is avoided by choosing $M \geq A + B - 1$ and extending the functions with zeros so that their length is equal to M. Letting $f_e(x)$ and $h_e(x)$ represent the extended functions yields, from Eq. (3.3-29), their convolution:

$$g_e(x) = \sum_{m=0}^{M-1} f_e(m) h_e(x - m) \tag{5.1-16}$$

for $x = 0, 1, 2, \ldots, M - 1$. As both $f_e(x)$ and $h_e(x)$ are assumed to have a period equal to M, $g_e(x)$ also has this period.

Using matrix notation, Eq. (5.1-16) can be expressed in the form

$$\mathbf{g} = \mathbf{H}\mathbf{f} \tag{5.1-17}$$

where $\mathbf{f}$ and $\mathbf{g}$ are M-dimensional column vectors:

$$\mathbf{f} = \begin{bmatrix} f_e(0) \\ f_e(1) \\ \vdots \\ f_e(M - 1) \end{bmatrix} \tag{5.1-18}$$

and

$$\mathbf{g} = \begin{bmatrix} g_e(0) \\ g_e(1) \\ \vdots \\ g_e(M-1) \end{bmatrix} \tag{5.1-19}$$

and $\mathbf{H}$ is the $M \times M$ matrix

$$\mathbf{H} = \begin{bmatrix} h_e(0) & h_e(-1) & h_e(-2) & \cdots & h_e(-M+1) \\ h_e(1) & h_e(0) & h_e(-1) & \cdots & h_e(-M+2) \\ h_e(2) & h_e(1) & h_e(0) & \cdots & h_e(-M+3) \\ \vdots & & & & \\ h_e(M-1) & h_e(M-2) & h_e(M-3) & \cdots & h_e(0) \end{bmatrix} \tag{5.1-20}$$

Because of the periodicity assumption on $h_e(x)$, it follows that $h_e(x) = h_e(M + x)$. This property allows Eq. (5.1-20) to be written in the form

$$\mathbf{H} = \begin{bmatrix} h_e(0) & h_e(M-1) & h_e(M-2) & \cdots & h_e(1) \\ h_e(1) & h_e(0) & h_e(M-1) & \cdots & h_e(2) \\ h_e(2) & h_e(1) & h_e(0) & \cdots & h_e(3) \\ \vdots & & & & \\ h_e(M-1) & h_e(M-2) & h_e(M-3) & \cdots & h_e(0) \end{bmatrix} \tag{5.1-21}$$

The structure of this matrix plays a fundamental role throughout the remainder of this chapter. In Eq. (5.1-21) the rows are related by a *circular shift* to the right; that is, the right-most element in one row is equal to the left-most element in the row immediately below. The shift is called circular because an element shifted off the right end of a row reappears at the left end of the next row. Moreover, in Eq. (5.1-21) the circularity of $\mathbf{H}$ is complete in the sense that it extends from the last row back to the first row. A square matrix in which each row is a circular shift of the preceding row, and the first row is a circular shift of the last row, is called a *circulant matrix*. Keep in mind that the circular behavior of $\mathbf{H}$ is a direct consequence of the assumed periodicity of $h_e(x)$.

Example: Suppose that $A = 4$ and $B = 3$. We may choose $M = 6$ and then append two zeros to the samples of $f(x)$ and three zeros to samples of $h(x)$.

In this case **f** and **g** are 6-D vectors and **H** is the 6 × 6 matrix

$$\mathbf{H} = \begin{bmatrix} h_e(0) & h_e(5) & h_e(4) & \cdots & h_e(1) \\ h_e(1) & h_e(0) & h_e(5) & \cdots & h_e(2) \\ h_e(2) & h_e(1) & h_e(0) & \cdots & h_e(3) \\ \vdots & & & & \\ h_e(5) & h_e(4) & h_e(3) & \cdots & h_e(0) \end{bmatrix}$$

However, as $h_e(x) = 0$ for $x = 3, 4, 5$, and $h_e(x) = h(x)$ for $x = 0, 1, 2$,

$$\mathbf{H} = \begin{bmatrix} h(0) & & & & h(2) & h(1) \\ h(1) & h(0) & & & & h(2) \\ h(2) & h(1) & h(0) & & & \\ & h(2) & h(1) & h(0) & & \\ & & h(2) & h(1) & h(0) & \\ & & & h(2) & h(1)) & h(0) \end{bmatrix}$$

where all elements not indicated in the matrix are zero. ❑

Extension of the discussion to a 2-D, discrete degradation model is straightforward. For two digitized images $f(x, y)$ and $h(x, y)$ of sizes $A \times B$ and $C \times D$, respectively, extended images of size $M \times N$ may be formed by padding the above functions with zeros. As indicated in Section 3.3.8, one procedure for doing this is to let

$$f_e(x, y) = \begin{cases} f(x, y) & 0 \leq x \leq A - 1 \quad \text{and} \quad 0 \leq y \leq B - 1 \\ 0 & A \leq x \leq M - 1 \quad \text{or} \quad B \leq y \leq N - 1 \end{cases}$$

and

$$h_e(x, y) = \begin{cases} h(x, y) & 0 \leq x \leq C - 1 \quad \text{and} \quad 0 \leq y \leq D - 1 \\ 0 & C \leq x \leq M - 1 \quad \text{or} \quad D \leq y \leq N - 1. \end{cases}$$

Treating the extended functions $f_e(x, y)$ and $h_e(x, y)$ as periodic in two dimensions, with periods M and N in the x and y directions, respectively, yields, from Eq. (3.3-35), the convolution of these two functions:

$$g_e(x, y) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} f_e(m, n) h_e(x - m, y - n) \tag{5.1-22}$$

for $x = 0, 1, 2, \ldots, M - 1$ and $y = 0, 1, 2, \ldots, N - 1$. The convolution function $g_e(x, y)$ is periodic with the same period of $f_e(x, y)$ and $h_e(x, y)$. Overlap of the individual convolution periods is avoided by choosing $M \geq A + C - 1$ and $N \geq B + D - 1$. To complete the discrete degradation model requires adding an $M \times N$ extended discrete noise term $\eta_e(x, y)$ to Eq. (5.1-22) so that

$$g_e(x, y) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} f_e(m, n) h_e(x - m, y - n) + \eta_e(x, y) \tag{5.1-23}$$

for $x = 0, 1, 2, \ldots, M - 1$ and $y = 0, 1, 2, \ldots, N - 1$.

Let **f**, **g**, and **n** represent MN-dimensional column vectors formed by stacking the rows of the $M \times N$ functions $f_e(x, y)$, $g_e(x, y)$, and $\eta_e(x, y)$. The first N elements of **f**, for example, are the elements in the first row of $f_e(x, y)$, the next N elements are from the second row, and so on for all M rows of $f_e(x, y)$. This convention allows Eq. (5.1-23) to be expressed in vector-matrix form:

$$\mathbf{g} = \mathbf{H}\mathbf{f} + \mathbf{n} \tag{5.1-24}$$

where **f**, **g**, and **n** are of dimension $(MN) \times 1$ and **H** is of dimension $MN \times MN$. This matrix consists of M^2 partitions, each partition being of size $N \times N$ and ordered according to

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_0 & \mathbf{H}_{M-1} & \mathbf{H}_{M-2} & \cdots & \mathbf{H}_1 \\ \mathbf{H}_1 & \mathbf{H}_0 & \mathbf{H}_{M-1} & \cdots & \mathbf{H}_2 \\ \mathbf{H}_2 & \mathbf{H}_1 & \mathbf{H}_0 & \cdots & \mathbf{H}_3 \\ \vdots & & & & \\ \mathbf{H}_{M-1} & \mathbf{H}_{M-2} & \mathbf{H}_{M-3} & \cdots & \mathbf{H}_0 \end{bmatrix} \tag{5.1-25}$$

Each partition $\mathbf{H}_j$ is constructed from the jth row of the extended function $h_e(x, y)$, as follows:

$$\mathbf{H}_j = \begin{bmatrix} h_e(j,0) & h_e(j,N-1) & h_e(j,N-2) & \cdots & h_e(j,1) \\ h_e(j,1) & h_e(j,0) & h_e(j,N-1) & \cdots & h_e(j,2) \\ h_e(j,2) & h_e(j,1) & h_e(j,0) & \cdots & h_e(j,3) \\ \vdots & & & & \\ h_e(j,N-1) & h_e(j,N-2) & h_e(j,N-3) & \cdots & h_e(j,0) \end{bmatrix} \tag{5.1-26}$$

where, as in Eq. (5.1-21), use was made of the periodicity of $h_e(x, y)$. Here, $\mathbf{H}_j$ is a circulant matrix, and the blocks of **H** are subscripted in a circular manner.

For these reasons, the matrix **H** in Eq. (5.1-25) is often called a *block-circulant* matrix.

Most of the discussion in the following sections centers on the discrete degradation model given in Eq. (5.1-24). Keep in mind that derivation of this expression was based on the assumption of a linear, space invariant degradation process. As indicated earlier, the objective is to estimate the image $f(x, y)$ given $g(x, y)$ and a knowledge of $h(x, y)$ and $\eta(x, y)$. In terms of Eq. (5.1-24), this objective requires estimating **f**, given **g** and some knowledge about **H** and **n**.

Although Eq. (5.1-24) seems deceptively simple, a direct solution of this expression to obtain the elements of **f** is a monumental processing task for images of practical size. If, for example, $M = N = 512$, **H** is of size 262,144 $\times$ 262,144. Thus to obtain **f** directly would require the solution of a system of 262,144 simultaneous linear equations. Fortunately, the complexity of this problem can be reduced considerably by taking advantage of the circulant properties of **H**.

5.2 DIAGONALIZATION OF CIRCULANT AND BLOCK-CIRCULANT MATRICES

We show in this section that solutions that are computationally feasible may be obtained from the model in Eq. (5.1-24) by diagonalizing the **H** matrix. In order to simplify the explanation we begin the discussion by considering circulant matrices and then extend the procedure to block-circulant matrices.

5.2.1 Circulant Matrices

Consider an $M \times M$ circulant matrix **H** of the form

$$\mathbf{H} = \begin{bmatrix} h_e(0) & h_e(M-1) & h_e(M-2) & \cdots & h_e(1) \\ h_e(1) & h_e(0) & h_e(M-1) & \cdots & h_e(2) \\ h_e(2) & h_e(1) & h_e(0) & \cdots & h_e(3) \\ \vdots & & & & \\ h_e(M-1) & h_e(M-2) & h_e(M-3) & \cdots & h_e(0) \end{bmatrix} \tag{5.2-1}$$

Let us define a scalar function $\lambda(k)$ and a vector $\mathbf{w}(k)$ as

$$\lambda(k) = h_e(0) + h_e(M-1)\exp\left[j\frac{2\pi}{M}k\right] + h_e(M-2)\exp\left[j\frac{2\pi}{M}2k\right] + \cdots + h_e(1)\exp\left[j\frac{2\pi}{M}(M-1)k\right] \tag{5.2-2}$$

where $j = \sqrt{-1}$, and

$$\mathbf{w}(k) = \begin{bmatrix} 1 \\ \exp\left[j\frac{2\pi}{M}k\right] \\ \exp\left[j\frac{2\pi}{M}2k\right] \\ \vdots \\ \exp\left[j\frac{2\pi}{M}(M-1)k\right] \end{bmatrix} \tag{5.2-3}$$

for $k = 0, 1, 2, \ldots, M-1$. It can be shown by matrix multiplication that

$$\mathbf{H}\mathbf{w}(k) = \lambda(k)\mathbf{w}(k). \tag{5.2-4}$$

This expression indicates that $\mathbf{w}(k)$ is an eigenvector of the circulant matrix $\mathbf{H}$ and that $\lambda(k)$ is its corresponding eigenvalue (see Section 3.6).

Next, let us form an $M \times M$ matrix $\mathbf{W}$ by using the M eigenvectors of $\mathbf{H}$ as columns:

$$\mathbf{W} = [\mathbf{w}(0) \quad \mathbf{w}(1) \quad \mathbf{w}(2) \quad \cdots \quad \mathbf{w}(M-1)]. \tag{5.2-5}$$

The kith element of $\mathbf{W}$, denoted by $W(k, i)$, is given by

$$W(k, i) = \exp\left[j\frac{2\pi}{M}ki\right] \tag{5.2-6}$$

for $k, i = 0, 1, 2, \ldots, M-1$. The orthogonality properties of the complex exponential allows writing the inverse matrix, $\mathbf{W}^{-1}$, by inspection; its kith element, symbolized as $W^{-1}(k, i)$, is

$$W^{-1}(k, i) = \frac{1}{M}\exp\left[-j\frac{2\pi}{M}ki\right]. \tag{5.2-7}$$

From Eqs. (5.2-6) and (5.2-7),

$$\mathbf{W}\mathbf{W}^{-1} = \mathbf{W}^{-1}\mathbf{W} = \mathbf{I} \tag{5.2-8}$$

where $\mathbf{I}$ is the $M \times M$ identity matrix.

The importance of the existence of the inverse matrix $\mathbf{W}^{-1}$ is that it guarantees that the columns of $\mathbf{W}$ (the eigenvectors of $\mathbf{H}$) are *linearly independent*.

From elementary matrix theory (Noble [1969]) $\mathbf{H}$ then may be expressed in the form

$$\mathbf{H} = \mathbf{W}\mathbf{D}\mathbf{W}^{-1} \tag{5.2-9}$$

or

$$\mathbf{D} = \mathbf{W}^{-1}\mathbf{H}\mathbf{W} \tag{5.2-10}$$

where $\mathbf{D}$ is a diagonal matrix whose elements $D(k, k)$ are the eigenvalues of $\mathbf{H}$; that is,

$$D(k, k) = \lambda(k). \tag{5.2-11}$$

Equation (5.2-10) indicates that $\mathbf{H}$ is diagonalized by using $\mathbf{W}^{-1}$ and $\mathbf{W}$ in the order indicated.

5.2.2 Block-Circulant Matrices

The transformation matrix for diagonalizing block circulants is constructed as follows. Let

$$w_M(i, m) = \exp\left[j\frac{2\pi}{M}im\right] \tag{5.2-12}$$

and

$$w_N(k, n) = \exp\left[j\frac{2\pi}{N}kn\right]. \tag{5.2-13}$$

Based on this notation, we define a matrix $\mathbf{W}$ of size $MN \times MN$ and containing M^2 partitions of size $N \times N$. The imth partition of $\mathbf{W}$ is

$$\mathbf{W}(i, m) = w_M(i, m)\mathbf{W}_N \tag{5.2-14}$$

for $i, m = 0, 1, 2, \ldots, M - 1$. Then $\mathbf{W}_N$ is an $N \times N$ matrix with elements

$$W_N(k, n) = w_N(k, n) \tag{5.2-15}$$

for $k, n = 0, 1, 2, \ldots, N - 1$.

The inverse matrix $\mathbf{W}^{-1}$ is also of size $MN \times MN$ with M^2 partitions of size $N \times N$. The imth partition of $\mathbf{W}^{-1}$, symbolized as $\mathbf{W}^{-1}(i, m)$, is

$$\mathbf{W}^{-1}(i, m) = \frac{1}{M}w_M^{-1}(i, m)\mathbf{W}_N^{-1} \tag{5.2-16}$$

where $w_M^{-1}(i, m)$ is

$$w_M^{-1}(i, m) = \exp\left[-j\frac{2\pi}{M}im\right] \tag{5.2-17}$$

for $i, m = 0, 1, 2, \ldots, M - 1$. The matrix $\mathbf{W}_N^{-1}$ has elements

$$W_N^{-1}(k, n) = \frac{1}{N} w_N^{-1}(k, n) \tag{5.2-18}$$

where

$$w_N^{-1}(k, n) = \exp\left[-j\frac{2\pi}{N}kn\right] \tag{5.2-19}$$

for $k, n = 0, 1, 2, \ldots, N - 1$. It can be verified by direct substitution of the elements of $\mathbf{W}$ and $\mathbf{W}^{-1}$ that

$$\mathbf{W}\mathbf{W}^{-1} = \mathbf{W}^{-1}\mathbf{W} = \mathbf{I} \tag{5.2-20}$$

where $\mathbf{I}$ is the $MN \times MN$ identity matrix.

From the results in Section 5.2.1, and if $\mathbf{H}$ is a block-circulant matrix, it can be shown (Hunt [1973]) that

$$\mathbf{H} = \mathbf{W}\mathbf{D}\mathbf{W}^{-1} \tag{5.2-21}$$

or

$$\mathbf{D} = \mathbf{W}^{-1}\mathbf{H}\mathbf{W} \tag{5.2-22}$$

where $\mathbf{D}$ is a diagonal matrix whose elements $D(k, k)$ are related to the discrete Fourier transform of the extended function $h_e(x, y)$ discussed in Section 5.1.3. Moreover, the transpose of $\mathbf{H}$, denoted $\mathbf{H}^T$, is

$$\mathbf{H}^T = \mathbf{W}\mathbf{D}^*\mathbf{W}^{-1} \tag{5.2-23}$$

where $\mathbf{D}^*$ is the complex conjugate of $\mathbf{D}$.

5.2.3 Effects of Diagonalization on the Degradation Model

The matrix $\mathbf{H}$ in the discrete, 1-D model of Eq. (5.1-17) is circulant, so it may be expressed in the form of Eq. (5.2-9). Equation (5.1-17) then becomes

$$\mathbf{g} = \mathbf{W}\mathbf{D}\mathbf{W}^{-1}\mathbf{f}. \tag{5.2-24}$$

Rearranging this equation yields

$$\mathbf{W}^{-1}\mathbf{g} = \mathbf{D}\mathbf{W}^{-1}\mathbf{f}. \tag{5.2-25}$$

The product $\mathbf{W}^{-1}\mathbf{f}$ is an M-dimensional column vector. From Eq. (5.2-7) and the definition of $\mathbf{f}$ in Section 5.1.3, the kth element of the product $\mathbf{W}^{-1}\mathbf{f}$, denoted $F(k)$, is

$$F(k) = \frac{1}{M}\sum_{i=0}^{M-1} f_e(i)\exp\left[-j\frac{2\pi}{M}ki\right] \tag{5.2-26}$$

for $k = 0, 1, 2, \ldots, M-1$. This expression is recognized as the discrete Fourier transform of the extended sequence $f_e(x)$. In other words, multiplication of $\mathbf{f}$ by $\mathbf{W}^{-1}$ yields a vector whose elements are the Fourier transforms of the elements of $\mathbf{f}$. Similarly, $\mathbf{W}^{-1}\mathbf{g}$ yields the Fourier transform of the elements of $\mathbf{g}$, denoted $G(k)$, $k = 0, 1, 2, \ldots, M-1$.

Next, we examine the matrix $\mathbf{D}$ in Eq. (5.2-25). The discussion in Section 5.2.1 showed that the main diagonal elements of $\mathbf{D}$ are the eigenvalues of the circulant matrix $\mathbf{H}$. The eigenvalues are given in Eq. (5.2-2) which, using the fact that

$$\exp\left[j\frac{2\pi}{M}(M-i)k\right] = \exp\left[-j\frac{2\pi}{M}ik\right] \tag{5.2-27}$$

may be written in the form

$$\begin{aligned}\lambda(k) = {} & h_e(0) + h_e(1)\exp\left[-j\frac{2\pi}{M}k\right] + h_e(2)\exp\left[-j\frac{2\pi}{M}2k\right] \\ & + \cdots + h_e(M-1)\exp\left[-j\frac{2\pi}{M}(M-1)k\right].\end{aligned} \tag{5.2-28}$$

From Eqs. (5.2-11) and (5.2-28),

$$D(k, k) = \lambda(k) = \sum_{i=0}^{M-1} h_e(i)\exp\left[-j\frac{2\pi}{M}ki\right] \tag{5.2-29}$$

for $k = 0, 1, 2, \ldots, M-1$. The right-hand side of this equation is $MH(k)$, where $H(k)$ is the discrete Fourier transform of the extended sequence $h_e(x)$. Thus

$$D(k, k) = MH(k). \tag{5.2-30}$$

These transforms can be combined into one result. Since $\mathbf{D}$ is a diagonal matrix, the product of $\mathbf{D}$ with any vector multiplies each element of that vector by a single diagonal element of $\mathbf{D}$. Consequently, the matrix formulation given in Eq. (5.2-25) can be reduced to a term-by-term product of 1-D Fourier transform sequences. In other words,

$$G(k) = MH(k)F(k) \tag{5.2-31}$$

for $k = 0, 1, 2, \ldots, M - 1$, where $G(k)$ are the elements of the vector $\mathbf{W}^{-1}\mathbf{g}$ and $MH(k)F(k)$ the elements of vector $\mathbf{D}\mathbf{W}^{-1}\mathbf{f}$. The right-hand side of Eq. (5.2-31) is the convolution of $f_e(x)$ and $h_e(x)$ in the frequency domain (see Section 3.3.8). Computationally, this result implies considerable simplification, because $G(k)$, $H(k)$, and $F(k)$ are M-sample discrete transforms, which can be obtained by using a fast Fourier transform algorithm.

A procedure similar to the preceding development yields equivalent results for the 2-D degradation model. Multiplying both sides of Eq. (5.1-24) by $\mathbf{W}^{-1}$ and using Eqs. (5.2-20) and (5.2-21) yields

$$\mathbf{W}^{-1}\mathbf{g} = \mathbf{D}\mathbf{W}^{-1}\mathbf{f} + \mathbf{W}^{-1}\mathbf{n} \tag{5.2-32}$$

where $\mathbf{W}^{-1}$ is an $MN \times MN$ matrix whose elements are given in Eq. (5.2-16), $\mathbf{D}$ is an $MN \times MN$ diagonal matrix, $\mathbf{H}$ is the $MN \times MN$ block-circulant matrix defined in Eq. (5.1-25), and $\mathbf{f}$ and $\mathbf{g}$ are vectors of dimension MN formed by stacking the rows of the extended images $f_e(x, y)$ and $g_e(x, y)$, respectively.

The left-hand side of Eq. (5.2-32) is a vector of dimension $MN \times 1$. Let us denote its elements $G(0, 0)$, $G(0, 1)$, $\ldots$, $G(0, N - 1)$; $G(1, 0)$, $G(1, 1)$, $\ldots$, $G(1, N - 1)$; $\ldots$; $G(M - 1, 0)$, $G(M - 1, 1)$, $\ldots$, $G(M - 1, N - 1)$. It can be shown (Hunt [1973]) that

$$G(u, v) = \frac{1}{MN}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1} g_e(x, y)\exp\left[-j2\pi\left(\frac{ux}{M} + \frac{vy}{N}\right)\right] \tag{5.2-33}$$

for $u = 0, 1, 2, \ldots, M - 1$, and $v = 0, 1, 2, \ldots, N - 1$. Equation (5.2-33) is the 2-D Fourier transform of $g_e(x, y)$. In other words, the elements of $\mathbf{W}^{-1}\mathbf{g}$ correspond to the stacked rows of the Fourier transform matrix with elements $G(u, v)$ for $u = 0, 1, 2, \ldots, M - 1$ and $v = 0, 1, 2, \ldots, N - 1$. Similarly, the vectors $\mathbf{W}^{-1}\mathbf{f}$ and $\mathbf{W}^{-1}\mathbf{n}$ are MN-dimensional and contain elements $F(u, v)$ and $N(u, v)$, where

$$F(u, v) = \frac{1}{MN}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1} f_e(x, y)\exp\left[-j2\pi\left(\frac{ux}{M} + \frac{vy}{N}\right)\right] \tag{5.2-34}$$

and

$$N(u, v) = \frac{1}{MN}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1} \eta_e(x, y)\exp\left[-j2\pi\left(\frac{ux}{M} + \frac{vy}{N}\right)\right] \tag{5.2-35}$$

for $u = 0, 1, 2, \ldots, M - 1$ and $v = 0, 1, 2, \ldots, N - 1$.

Finally, the elements of the diagonal matrix $\mathbf{D}$ are related to the Fourier transform of the extended impulse response function $h_e(x, y)$; that is,

$$H(u, v) = \frac{1}{MN}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1} h_e(x, y)\exp\left[-j2\pi\left(\frac{ux}{M} + \frac{vy}{N}\right)\right] \tag{5.2-36}$$

for $u = 0, 1, 2, \ldots, M - 1$, and $v = 0, 1, 2, \ldots, N - 1$. The MN diagonal elements of $\mathbf{D}$ are formed as follows. The first N elements are $H(0, 0)$, $H(0, 1), \ldots, H(0, N - 1)$; the next, $H(1, 0), H(1, 1), \ldots, H(1, N - 1)$; and so on, with the last N diagonal elements being $H(M - 1, 0), H(M - 1, 1), \ldots, H(M - 1, N - 1)$. The off-diagonal elements, of course, are zero. The entire matrix formed from the preceding elements is then multiplied by MN to obtain $\mathbf{D}$. A more concise way of expressing this construction is as

$$D(k, i) = \begin{cases} MNH\left(\left[\frac{k}{N}\right], k \bmod N\right) & \text{if } i = k \\ 0 & \text{if } i \neq k \end{cases} \tag{5.2-37}$$

where $[c]$ is used to denote the greatest integer not exceeding c, and $k \bmod N$ is the remainder obtained by dividing k by N.

Equations (5.2-33)–(5.2-36) can be used to show that the individual elements of Eq. (5.2-32) are related by the expression

$$G(u, v) = MNH(u, v)F(u, v) + N(u, v) \tag{5.2-38}$$

for $u = 0, 1, 2, \ldots, M - 1$, and $v = 0, 1, 2, \ldots, N - 1$.

The term MN is simply a scale factor, which for notational purposes can be absorbed conveniently in $H(u, v)$. With this notation, Eqs. (5.2-37) and (5.2-38) may be expressed as

$$D(k, i) = \begin{cases} H\left(\left[\frac{k}{N}\right], k \bmod N\right) & \text{if } i = k \\ 0 & \text{if } i \neq k \end{cases} \tag{5.2-39}$$

for $k, i = 0, 1, 2, \ldots, MN - 1$, and

$$G(u, v) = H(u, v)F(u, v) + N(u, v) \tag{5.2-40}$$

for $u = 0, 1, 2, \ldots, M - 1$, and $v = 0, 1, 2, \ldots, N - 1$, with $H(u, v)$ now scaled by the factor MN.

The significance of Eq. (5.2-38) or (5.2-40) is that the large system of equations implicit in the model in Eq. (5.1-24) can be reduced to computation of a few discrete Fourier transforms of size $M \times N$. If M and N are integer powers of 2, for example, this is a simple problem if we use an FFT algorithm. As mentioned earlier, however, the problem becomes an almost impossible computational task if approached directly from the model in Eq. (5.1-24).

We use the model in Eq. (5.1-24) in the following sections as the basis for deriving several image restoration approaches. We then simplify the results,

which are in matrix form, by using the concepts introduced in this section. Keep in mind that the simplifications achieved are the result of assuming that (1) the degradation is a linear, space invariant process, and (2) all images are treated as extended, periodic functions.

Equation (5.2-40) could have been written directly from Eq. (5.1-15) via the convolution theorem. However, our objective was to show that the same result could be achieved by a matrix formulation. In so doing, we established a number of important matrix properties for use in Section 5.3 to develop a unified approach to restoration.

5.3 ALGEBRAIC APPROACH TO RESTORATION

As indicated in Section 5.1.3, the objective of image restoration is to estimate an original image $\mathbf{f}$ from a degraded image $\mathbf{g}$ and some knowledge or assumptions about $\mathbf{H}$ and $\mathbf{n}$. Assuming that these quantities are related according to the model in Eq. (5.1-24) allows formulation of a class of image restoration problems in a unified linear algebraic framework.

Central to the algebraic approach is the concept of seeking an estimate of $\mathbf{f}$, denoted $\hat{\mathbf{f}}$, that minimizes a predefined criterion of performance. Because of their simplicity, this chapter focuses on least squares criterion functions. This choice has the added advantage of yielding a central approach for the derivation of several well-known restoration methods. These methods are the result of considering either an unconstrained or a constrained approach to the least squares restoration problem.

5.3.1 Unconstrained Restoration

From Eq. (5.1-24), the noise term in the degradation model is

$$\mathbf{n} = \mathbf{g} - \mathbf{H}\mathbf{f}. \tag{5.3-1}$$

In the absence of any knowledge about $\mathbf{n}$, a meaningful criterion function is to seek an $\hat{\mathbf{f}}$ such that $\mathbf{H}\hat{\mathbf{f}}$ approximates $\mathbf{g}$ in a least squares sense by assuming that the norm of the noise term is as small as possible. In other words, we want to find an $\hat{\mathbf{f}}$ such that

$$\|\mathbf{n}\|^2 = \|\mathbf{g} - \mathbf{H}\hat{\mathbf{f}}\|^2 \tag{5.3-2}$$

is minimum, where, by definition,

$$\|\mathbf{n}\|^2 = \mathbf{n}^T\mathbf{n} \quad \text{and} \quad \|\mathbf{g} - \mathbf{H}\hat{\mathbf{f}}\|^2 = (\mathbf{g} - \mathbf{H}\hat{\mathbf{f}})^T(\mathbf{g} - \mathbf{H}\hat{\mathbf{f}})$$

are the squared norms of $\mathbf{n}$ and $(\mathbf{g} - \mathbf{H}\hat{\mathbf{f}})$, respectively. Equation (5.3-2) allows the equivalent view of this problem as one of minimizing the criterion function

$$J(\hat{\mathbf{f}}) = \|\mathbf{g} - \mathbf{H}\hat{\mathbf{f}}\|^2 \tag{5.3-3}$$

with respect to $\hat{\mathbf{f}}$. Aside from the requirement that it minimize Eq. (5.3-3), $\hat{\mathbf{f}}$ is not constrained in any other way.

Minimization of Eq. (5.3-3) is straightforward. We simply differentiate J with respect to $\hat{\mathbf{f}}$ and set the result equal to the zero vector; that is,

$$\frac{\partial J(\hat{\mathbf{f}})}{\partial \hat{\mathbf{f}}} = \mathbf{0} = -2\mathbf{H}^T(\mathbf{g} - \mathbf{H}\hat{\mathbf{f}}). \tag{5.3-4}$$

Solving Eq. (5.3-4) for $\hat{\mathbf{f}}$ yields

$$\hat{\mathbf{f}} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{g}. \tag{5.3-5}$$

Letting $M = N$ so that $\mathbf{H}$ is a square matrix and assuming that $\mathbf{H}^{-1}$ exists reduces Eq. (5.3-5) to

$$\begin{aligned}\hat{\mathbf{f}} &= \mathbf{H}^{-1}(\mathbf{H}^T)^{-1}\mathbf{H}^T\mathbf{g} \\ &= \mathbf{H}^{-1}\mathbf{g}.\end{aligned} \tag{5.3-6}$$

5.3.2 Constrained Restoration

In this section, we consider the least squares restoration problem as one of minimizing functions of the form $\|\mathbf{Q}\hat{\mathbf{f}}\|^2$, where $\mathbf{Q}$ is a linear operator on $\mathbf{f}$, subject to the constraint $\|\mathbf{g} - \mathbf{H}\hat{\mathbf{f}}\|^2 = \|\mathbf{n}\|^2$. This approach introduces considerable flexibility in the restoration process because it yields different solutions for different choices of $\mathbf{Q}$. The constraint imposed on a solution is consistent with the model in Eq. (5.1-24).

The addition of an equality constraint in the minimization problem can be handled without difficulty by using the method of *Lagrange multipliers* (Elsgolc [1961]). The procedure calls for expressing the constraint in the form $\alpha(\|\mathbf{g} - \mathbf{H}\hat{\mathbf{f}}\|^2 - \|\mathbf{n}\|^2)$ and then appending it to the function $\|\mathbf{Q}\hat{\mathbf{f}}\|^2$. In other words, we seek an $\hat{\mathbf{f}}$ that minimizes the criterion function

$$J(\hat{\mathbf{f}}) = \|\mathbf{Q}\hat{\mathbf{f}}\|^2 + \alpha(\|\mathbf{g} - \mathbf{H}\hat{\mathbf{f}}\|^2 - \|\mathbf{n}\|^2) \tag{5.3-7}$$

where α is a constant called the *Lagrange multiplier*. After the constraint has been appended, minimization is carried out in the usual way.

Differentiating Eq. (5.3-7) with respect to $\hat{\mathbf{f}}$ and setting the result equal to the zero vector yields

$$\frac{\partial J(\hat{\mathbf{f}})}{\partial \hat{\mathbf{f}}} = \mathbf{0} = 2\mathbf{Q}^T\mathbf{Q}\hat{\mathbf{f}} - 2\alpha\mathbf{H}^T(\mathbf{g} - \mathbf{H}\hat{\mathbf{f}}). \tag{5.3-8}$$

The solution is obtained by solving Eq. (5.3-8) for $\hat{\mathbf{f}}$; that is,

$$\hat{\mathbf{f}} = (\mathbf{H}^T\mathbf{H} + \gamma\mathbf{Q}^T\mathbf{Q})^{-1}\mathbf{H}^T\mathbf{g} \tag{5.3-9}$$

where $\gamma = 1/\alpha$. This quantity must be adjusted so that the constraint is satisfied, a problem considered later in this chapter. Equations (5.3-6) and (5.3-9) are the bases for all the restoration procedures discussed in the following sections. In Section 5.4, for example, we show that Eq. (5.3-6) leads to the traditional inverse-filter restoration method. Similarly, the general formulation in Eq. (5.3-9) can be used to derive results such as the classical Wiener filter, as well as other restoration techniques. To do so simply requires selecting an appropriate transformation matrix $\mathbf{Q}$ and using the simplifications derived in Section 5.2.

5.4 INVERSE FILTERING

5.4.1 Formulation

We begin the derivation of image restoration techniques by considering the unconstrained result in Eq. (5.3-6). If we assume that $M = N$ and use Eq. (5.2-21), Eq. (5.3-6) becomes

$$\begin{aligned}\hat{\mathbf{f}} &= \mathbf{H}^{-1}\mathbf{g} \\ &= (\mathbf{W}\mathbf{D}\mathbf{W}^{-1})^{-1}\mathbf{g} \\ &= \mathbf{W}\mathbf{D}^{-1}\mathbf{W}^{-1}\mathbf{g}.\end{aligned} \tag{5.4-1}$$

Premultiplying both sides of Eq. (5.4-1) by $\mathbf{W}^{-1}$ yields

$$\mathbf{W}^{-1}\hat{\mathbf{f}} = \mathbf{D}^{-1}\mathbf{W}^{-1}\mathbf{g}. \tag{5.4-2}$$

From the discussion in Section 5.2.3, the elements comprising Eq. (5.4-2) may be written in the form

$$\hat{F}(u, v) = \frac{G(u, v)}{H(u, v)} \tag{5.4-3}$$

for $u, v = 0, 1, 2, \ldots, N-1$. According to Eq. (5.2-39), $H(u, v)$ is assumed to be scaled by N^2 and, because $\mathbf{D}$ is a diagonal matrix, its inverse is easily obtained by inspection.

The image restoration approach given by Eq. (5.4-3) is commonly referred to as the *inverse filter* method. This terminology arises from considering $H(u, v)$ as a "filter" function that multiplies $F(u, v)$ to produce the transform of the degraded image $g(x, y)$. The division of $G(u, v)$ by $H(u, v)$ indicated in Eq. (5.4-3) then constitutes an inverse filtering operation in this context. The restored image, of course, is obtained by using the relation

$$\begin{aligned} \hat{f}(x, y) &= \mathfrak{F}^{-1}[\hat{F}(u, v)] \\ &= \mathfrak{F}^{-1}[G(u, v)/H(u, v)] \end{aligned} \tag{5.4-4}$$

for $x, y = 0, 1, 2, \ldots, N-1$. This procedure is normally implemented by means of an FFT algorithm.

Note in Eq. (5.4-4) that computational difficulties will be encountered in the restoration process if $H(u, v)$ vanishes or becomes very small in any region of interest in the uv plane. If the zeros of $H(u, v)$ are located at a few known points in the uv plane, they generally can be neglected in the computation of $\hat{F}(u, v)$ without noticeably affecting the restored result.

A more serious difficulty arises in the presence of noise. Substituting Eq. (5.2-40) into Eq. (5.4-3) yields

$$\hat{F}(u, v) = F(u, v) + \frac{N(u, v)}{H(u, v)}. \tag{5.4-5}$$

This expression clearly indicates that if $H(u, v)$ is zero or becomes very small, the term $N(u, v)/H(u, v)$ could dominate the restoration result $\mathfrak{F}^{-1}[\hat{F}(u, v)]$. In practice $H(u, v)$ often drops off rapidly as a function of distance from the origin of the uv plane. The noise term, however, usually falls off at a much slower rate. In such situations, reasonable results often can be obtained by carrying out the restoration in a limited neighborhood about the origin in order to avoid small values of $H(u, v)$.

Example: Figure 5.2(a) shows a point image $f(x, y)$, and Fig. 5.2(b) shows a degraded image $g(x, y)$ obtained by blurring $f(x, y)$. Considering the point source to be an approximation to a unit impulse function gives

$$\begin{aligned} G(u, v) &= H(u, v)F(u, v) \\ &\approx H(u, v) \end{aligned}$$

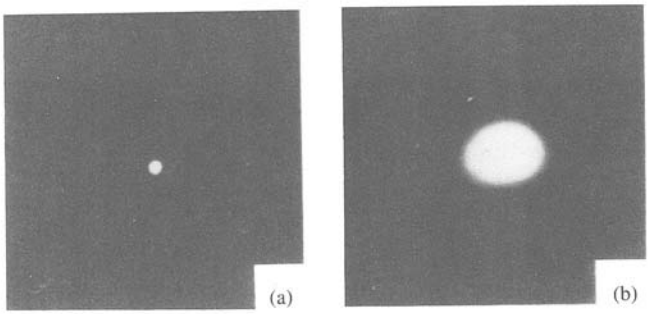


Figure 5.2 *Blurring of a point source to obtain $H(u, v)$.*

because $\mathfrak{F}[\delta(x, y)] = 1$. This expression indicates that the transfer function $H(u, v)$ can be approximated by the Fourier transform of the degraded image. The procedure of blurring a known function to obtain an approximation to $H(u, v)$ is a useful one in practice because it can often be used in a trial-and-error approach to restore images for which the blurring function $H(u, v)$ is not known a priori.

The result of applying the same blurring function as above to the ideal image shown in Fig. 5.3(a) is shown in Fig. 5.3(b). The restored image shown in Fig. 5.3(c) was obtained by using Eq. (5.4-4) for values of u and v near enough to the origin of the uv plane to avoid excessively small values of $H(u, v)$. The result of carrying out the restoration for a larger neighborhood is shown in Fig. 5.3(d). These results clearly point out the difficulties introduced by a vanishing function $H(u, v)$. ❑

If $H(u, v)$, $G(u, v)$, and $N(u, v)$ all are known, an exact inverse filtering expression can be obtained directly from Eq. (5.2-40); that is,

$$F(u, v) = \frac{G(u, v)}{H(u, v)} - \frac{N(u, v)}{H(u, v)}. \tag{5.4-6}$$

In addition to the potential difficulties with $H(u, v)$ outlined in the preceding example, a problem with this formulation is that the noise is seldom known well enough to allow computation of $N(u, v)$.

5.4.2 Removal of Blur Caused by Uniform Linear Motion

There are practical applications in which $H(u, v)$ can be obtained analytically, but the solution has zero values in the frequency range of interest. In Section 5.4.1, we gave an example of the difficulties caused by a vanishing $H(u, v)$. In

the following discussion we consider the problem of restoring an image that has been blurred by uniform linear motion. We singled out this problem because of its practical implications and also because it lends itself well to an analytical formulation. Solution of the uniform blurring case also demonstrates how zeros of $H(u, v)$ can be handled computationally. These considerations are important, because they often arise in practice in other contexts of image restoration by inverse filtering.

Suppose that an image $f(x, y)$ undergoes planar motion and that $x_0(t)$ and $y_0(t)$ are the time varying components of motion in the x and y directions, respectively. The total exposure at any point of the recording medium (say, film) is obtained in this case by integrating the instantaneous exposure over the time interval during which the shutter is open. Assuming that shutter opening and closing takes place instantaneously and that the optical imaging process is perfect isolates the effect of image motion. Then, if T is the duration of the

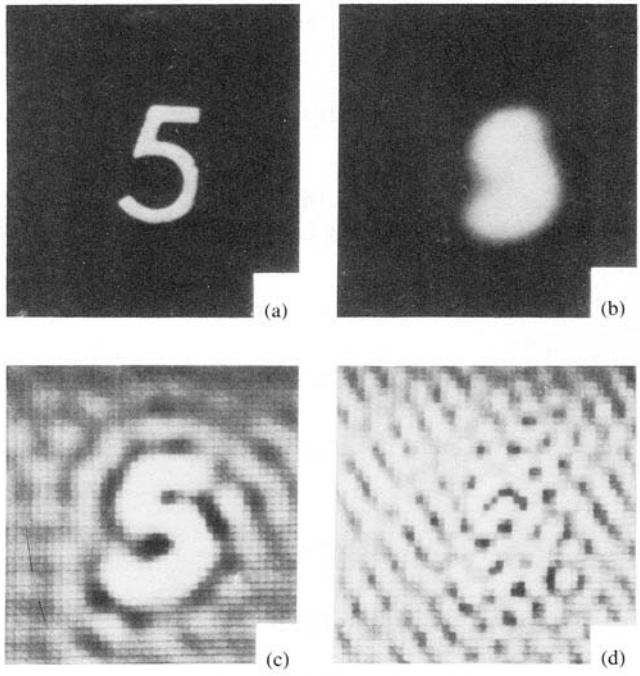


Figure 5.3 *Example of image restoration by inverse filtering: (a) original image $f(x, y)$; (b) degraded (blurred) image $g(x, y)$; (c) result of restoration by considering a neighborhood about the origin of the uv plane that does not include excessively small values of $H(u, v)$; (d) result of using a larger neighborhood in which this condition does not hold. (From McGlamery [1967].)*

exposure, it follows that

$$g(x, y) = \int_0^T f[x - x_0(t), y - y_0(t)]\, dt \tag{5.4-7}$$

where $g(x, y)$ is the blurred image.

From Eq. (3.1-9), the Fourier transform of Eq. (5.4-7) is

$$\begin{aligned} G(u, v) &= \int_{-\infty}^{\infty}\int g(x, y)\exp[-j2\pi(ux + vy)]\, dx\, dy \\ &= \int_{-\infty}^{\infty}\int \left[\int_0^T f[x - x_0(t), y - y_0(t)]\, dt\right]\exp[-j2\pi(ux + vy)]\, dx\, dy. \end{aligned} \tag{5.4-8}$$

Reversing the order of integration allows Eq. (5.4-8) to be expressed in the form

$$G(u, v) = \int_0^T\left[\int_{-\infty}^{\infty}\int f[x - x_0(t), y - y_0(t)]\exp[-j2\pi(ux + vy)]\, dx\, dy\right] dt. \tag{5.4-9}$$

The term inside the outer brackets is the Fourier transform of the displaced function $f[x - x_0(t), y - y_0(t)]$. Using Eq. (3.3-7b) then yields the relation

$$\begin{aligned} G(u, v) &= \int_0^T F(u, v)\exp\{-j2\pi[ux_0(t) + vy_0(t)]\}\, dt \\ &= F(u, v)\int_0^T \exp\{-j2\pi[ux_0(t) + vy_0(t)]\}\, dt \end{aligned} \tag{5.4-10}$$

where the last step follows from the fact that $F(u, v)$ is independent of t.

By defining

$$H(u, v) = \int_0^T \exp\{-j2\pi[ux_0(t) + vy_0(t)]\}\, dt \tag{5.4-11}$$

Eq. (5.4-10) may be expressed in the familiar form

$$G(u, v) = H(u, v)F(u, v). \tag{5.4-12}$$

If the nature of the motion variables $x_0(t)$ and $y_0(t)$ is known, the transfer function $H(u, v)$ can be obtained directly from Eq. (5.4-11). As an illustration, suppose that the image in question undergoes uniform linear motion in the x direction only, at a rate given by $x_0(t) = at/T$. When $t = T$, the image has

been displaced by a total distance a. With $y_0(t) = 0$, Eq. (5.4-11) yields

$$\begin{aligned} H(u, v) &= \int_0^T \exp[-j2\pi u x_0(t)]\, dt \\ &= \int_0^T \exp[-j2\pi uat/T]\, dt \\ &= \frac{T}{\pi ua} \sin(\pi ua) e^{-j\pi ua}. \end{aligned} \tag{5.4-13}$$

Obviously, H vanishes at values of u given by $u = n/a$, where n is an integer.

When $f(x, y)$ is zero (or known) outside an interval $0 \leq x \leq L$, the problem presented by Eq. (5.4-13) can be avoided and the image completely reconstructed from a knowledge of $g(x, y)$ in this interval. Because y is time invariant, suppressing this variable temporarily allows Eq. (5.4-7) to be written as

$$\begin{aligned} g(x) &= \int_0^T f[x - x_0(t)]\, dt \\ &= \int_0^T f\left(x - \frac{at}{T}\right) dt \qquad 0 \leq x \leq L. \end{aligned} \tag{5.4-14}$$

Substituting $\tau = x - at/T$ in this expression and ignoring a scale factor yields

$$g(x) = \int_{x-a}^{x} f(\tau)\, d\tau \qquad 0 \leq x \leq L. \tag{5.4-15}$$

Then, by differentiation with respect to x (using Liebnitz's rule),

$$g'(x) = f(x) - f(x - a) \qquad 0 \leq x \leq L \tag{5.4-16}$$

or

$$f(x) = g'(x) + f(x - a) \qquad 0 \leq x \leq L. \tag{5.4-17}$$

In the following development a convenient assumption is that $L = Ka$, where K is an integer. Then the variable x may be expressed in the form

$$x = z + ma \tag{5.4-18}$$

where z takes on values in the interval $[0, a]$ and m is the integral part of (x/a). For example, if $a = 2$ and $x = 3.5$, then $m = 1$ (the integral part of 3.5/2), and $z = 1.5$. Clearly, $z + ma = 3.5$, as required. Note also that, for $L = Ka$, the index m can assume any of the integer values $0, 1, \ldots, K - 1$. For instance, when $x = L$, then $z = a$ and $m = K - 1$.

Substitution of Eq. (5.4-18) into Eq. (5.4-17) yields

$$f(z + ma) = g'(z + ma) + f[z + (m - 1)a]. \tag{5.4-19}$$

Next, denoting $\phi(z)$ as the portion of the scene that moves into the range $0 \leq z < a$ during exposure gives

$$\phi(z) = f(z - a) \qquad 0 \leq z < a. \tag{5.4-20}$$

Equation (5.4-19) can be solved recursively in terms of $\phi(z)$. Thus for $m = 0$,

$$\begin{aligned} f(z) &= g'(z) + f(z - a) \\ &= g'(z) + \phi(z). \end{aligned} \tag{5.4-21}$$

For $m = 1$, Eq. (5.4-19) becomes

$$f(z + a) = g'(z + a) + f(z). \tag{5.4-22}$$

Substituting Eq. (5.4-21) into Eq. (5.4-22) yields

$$f(z + a) = g'(z + a) + g'(z) + \phi(z). \tag{5.4-23}$$

In the next step, letting $m = 2$ results in the expression

$$f(z + 2a) = g'(z + 2a) + f(z + a) \tag{5.4-24}$$

or, substituting Eq. (5.4-23) for $f(z + a)$,

$$f(z + 2a) = g'(z + 2a) + g'(z + a) + g'(z) + \phi(z). \tag{5.4-25}$$

Continuing with this procedure finally yields

$$f(z + ma) = \sum_{k=0}^{m} g'(z + ka) + \phi(z). \tag{5.4-26}$$

However, as $x = z + ma$, Eq. (5.4-26) may be expressed in the form

$$f(x) = \sum_{k=0}^{m} g'(x - ka) + \phi(x - ma) \qquad 0 \leq x \leq L. \tag{5.4-27}$$

Because $g(x)$ is known, the problem is reduced to that of estimating $\phi(x)$.

One way to estimate this function directly from the blurred image is as follows. First note that, as x varies from 0 to L, m ranges from 0 to $K - 1$. The argument of ϕ is $(x - ma)$, which is always in the range $0 \leq x - ma <$

a, so ϕ is repeated K times during the evaluation of $f(x)$ for $0 \leq x \leq L$. Next, defining

$$\tilde{f}(x) = \sum_{j=0}^{m} g'(x - ja) \tag{5.4-28}$$

allows rewriting Eq. (5.4-27) as

$$\phi(x - ma) = f(x) - \tilde{f}(x). \tag{5.4-29}$$

Evaluating the left-hand and right-hand sides of Eq. (5.4-29) for $ka \leq x < (k + 1)a$, and adding the results for $k = 0, 1, \ldots, K - 1$ gives

$$K\phi(x) = \sum_{k=0}^{K-1} f(x + ka) - \sum_{k=0}^{K-1} \tilde{f}(x + ka) \qquad 0 \leq x < a \tag{5.4-30}$$

where $m = 0$ because $0 \leq x < a$. Dividing through by K yields

$$\phi(x) = \frac{1}{K} \sum_{k=0}^{K-1} f(x + ka) - \frac{1}{K} \sum_{k=0}^{K-1} \tilde{f}(x + ka). \tag{5.4-31}$$

The first sum on the right-hand side of this expression is, of course, unknown. However, for large values of K it approaches the average value of f. Thus this sum may be taken as a constant A, giving the approximation

$$\phi(x) \approx A - \frac{1}{K} \sum_{k=0}^{K-1} \tilde{f}(x + ka) \qquad 0 \leq x < a \tag{5.4-32}$$

or

$$\phi(x - ma) \approx A - \frac{1}{K} \sum_{k=0}^{K-1} \tilde{f}(x + ka - ma) \qquad 0 \leq x \leq L. \tag{5.4-33}$$

Substituting Eq. (5.4-28) for $\tilde{f}$ yields[†]

$$\begin{aligned} \phi(x - ma) &\approx A - \frac{1}{K} \sum_{k=0}^{K-1} \sum_{j=0}^{k} g'(x + ka - ma - ja) \\ &\approx A - \frac{1}{K} \sum_{k=0}^{K-1} \sum_{j=0}^{k} g'[x - ma + (k - j)a]. \end{aligned} \tag{5.4-34}$$

[†] Note that the limit on the second summation is k instead of m. If we had started from Eq. (5.4-18) with $x + ka - ma$ instead of x, the limit in the summation of Eq. (5.4-28) would have been k because, from Eq. (5.4-18), $x + (ka - ma) = z + ma + (ka - ma) = z + ka$.

From Eqs. (5.4-28) and (5.4-29), we have the final result:

$$f(x) \approx A - \frac{1}{K} \sum_{k=0}^{K-1} \sum_{j=0}^{k} g'[x - ma + (k - j)a] + \sum_{j=0}^{m} g'(x - ja) \quad (5.4\text{-}35)$$

for $0 \leq x \leq L$. Reintroducing the suppressed variable y yields

$$f(x, y) \approx A - \frac{1}{K} \sum_{k=0}^{K-1} \sum_{j=0}^{k} g'[x - ma + (k - j)a, y] + \sum_{j=0}^{m} g'(x - ja, y) \quad (5.4\text{-}36)$$

for $0 \leq x, y \leq L$. As before, $f(x, y)$ is assumed to be a square image. Interchanging x and y in the right-hand side of Eq. (5.4-36) would give the reconstruction of an image that moves only in the y direction during exposure. The concepts presented can also be used to derive a deblurring expression that takes into account simultaneous uniform motion in both directions.

Example: The image shown in Fig. 5.4(a) was blurred by uniform linear motion in one direction during exposure, with the total distance traveled being approximately equal to $\frac{1}{8}$ the width of the photograph. Figure 5.4(b) shows the deblurred result obtained by using Eq. (5.4-36) with x and y interchanged because motion is in the y direction. The error in the approximation given by this equation is not objectionable. ❑

(a) (b)

Figure 5.4 *(a) Image blurred by uniform linear motion; (b) image restored by using Eq. (5.4-36). (From Sondhi [1972].)*

5.5 LEAST MEAN SQUARE (WIENER) FILTER

Let $\mathbf{R}_f$ and $\mathbf{R}_n$ be the correlation matrices of $\mathbf{f}$ and $\mathbf{n}$, defined respectively by the equations

$$\mathbf{R}_f = E\{\mathbf{f}\mathbf{f}^T\} \tag{5.5-1}$$

and

$$\mathbf{R}_n = E\{\mathbf{n}\mathbf{n}^T\} \tag{5.5-2}$$

where $E\{\cdot\}$ denotes the expected value operation, and $\mathbf{f}$ and $\mathbf{n}$ are as defined in Section 5.1.3. The ijth element of $\mathbf{R}_f$ is given by $E\{f_i f_j\}$, which is the correlation between the ith and the jth elements of $\mathbf{f}$. Similarly, the ijth element of $\mathbf{R}_n$ gives the correlation between the two corresponding elements in $\mathbf{n}$. Since the elements of $\mathbf{f}$ and $\mathbf{n}$ are real, $E\{f_i f_j\} = E\{f_j f_i\}$, $E\{n_i n_j\} = E\{n_j n_i\}$, and it follows that $\mathbf{R}_f$ and $\mathbf{R}_n$ are real symmetric matrices. For most image functions the correlation between pixels (that is, elements of $\mathbf{f}$ or $\mathbf{n}$) does not extend beyond a distance of 20 to 30 pixels in the image, so a typical correlation matrix has a band of nonzero elements about the main diagonal and zeros in the right upper and left lower corner regions. Based on the assumption that the correlation between any two pixels is a function of the distance between the pixels and not their position, $\mathbf{R}_f$ and $\mathbf{R}_n$ can be made to approximate block-circulant matrices and therefore can be diagonalized by the matrix $\mathbf{W}$ with the procedure described in Section 5.2.2 (Andrews and Hunt [1977]). Using $\mathbf{A}$ and $\mathbf{B}$ to denote matrices gives

$$\mathbf{R}_f = \mathbf{W}\mathbf{A}\mathbf{W}^{-1} \tag{5.5-3}$$

and

$$\mathbf{R}_n = \mathbf{W}\mathbf{B}\mathbf{W}^{-1}. \tag{5.5-4}$$

Just as the elements of the diagonal matrix $\mathbf{D}$ in the relation $\mathbf{H} = \mathbf{W}\mathbf{D}\mathbf{W}^{-1}$ correspond to the Fourier transform of the block elements of $\mathbf{H}$, the elements of $\mathbf{A}$ and $\mathbf{B}$ are the transforms of the correlation elements in $\mathbf{R}_f$ and $\mathbf{R}_n$, respectively. As indicated in Problem 3.4, the Fourier transform of these correlations is called the *power spectrum* (or *spectral density*) of $f_e(x, y)$ and $\eta_e(x, y)$, respectively and is denoted $S_f(u, v)$ and $S_\eta(u, v)$ in the following discussion.

Defining

$$\mathbf{Q}^T\mathbf{Q} = \mathbf{R}_f^{-1}\mathbf{R}_n \tag{5.5-5}$$

and substituting this expression in Eq. (5.3-9) gives

$$\hat{\mathbf{f}} = (\mathbf{H}^T\mathbf{H} + \gamma\mathbf{R}_f^{-1}\mathbf{R}_n)^{-1}\mathbf{H}^T\mathbf{g}. \tag{5.5-6}$$

Using Eqs. (5.2-21), (5.2-23), (5.5-3), and (5.5-4) yields

$$\hat{\mathbf{f}} = (\mathbf{W}\mathbf{D}^*\mathbf{D}\mathbf{W}^{-1} + \gamma\mathbf{W}\mathbf{A}^{-1}\mathbf{B}\mathbf{W}^{-1})^{-1}\mathbf{W}\mathbf{D}^*\mathbf{W}^{-1}\mathbf{g}. \tag{5.5-7}$$

Multiplying both sides by $\mathbf{W}^{-1}$ and performing some matrix manipulations reduces Eq. (5.5-7) to

$$\mathbf{W}^{-1}\hat{\mathbf{f}} = (\mathbf{D}^*\mathbf{D} + \gamma\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{D}^*\mathbf{W}^{-1}\mathbf{g}. \tag{5.5-8}$$

Keeping in mind the meaning of the elements of $\mathbf{A}$ and $\mathbf{B}$, recognizing that the matrices inside the parentheses are diagonal, and making use of the concepts developed in Section 5.2.3, allows writing the elements of Eq. (5.5-8) in the form

$$\begin{aligned}\hat{F}(u, v) &= \left[\frac{H^*(u, v)}{|H(u, v)|^2 + \gamma[S_\eta(u, v)/S_f(u, v)]}\right]G(u, v) \\ &= \left[\frac{1}{H(u, v)}\frac{|H(u, v)|^2}{|H(u, v)|^2 + \gamma[S_\eta(u, v)/S_f(u, v)]}\right]G(u, v)\end{aligned} \tag{5.5-9}$$

for $u, v = 0, 1, 2, \ldots, N-1$, where $|H(u, v)|^2 = H^*(u, v)H(u, v)$ and it is assumed that $M = N$.

When $\gamma = 1$, the term inside the outer brackets in Eq. (5.5-9) reduces to the so-called *Wiener filter*. If γ is variable this expression is called the *parametric Wiener filter*. In the absence of noise, $S_\eta(u, v) = 0$ and the Wiener filter reduces to the ideal inverse filter discussed in Section 5.4. However, when $\gamma = 1$, the use of Eq. (5.5-9) no longer yields an optimal solution in the sense defined in Section 5.3.2 because, as pointed out in that section, γ must be adjusted to satisfy the constraint $\|\mathbf{g} - \mathbf{H}\hat{\mathbf{f}}\|^2 = \|\mathbf{n}\|^2$. It can be shown however, that the solution obtained with $\gamma = 1$ *is* optimal in the sense that it minimizes the quantity $E\{[f(x, y) - \hat{f}(x, y)]^2\}$. Clearly, this is a statistical criterion that treats f and $\hat{f}$ as random variables.

When $S_\eta(u, v)$ and $S_f(u, v)$ are unknown (a problem often encountered in practice) approximating Eq. (5.5-9) by the relation

$$\hat{F}(u, v) \approx \left[\frac{1}{H(u, v)}\frac{|H(u, v)|^2}{|H(u, v)|^2 + K}\right]G(u, v) \tag{5.5-10}$$

where K is a constant, sometimes is useful. An example of results obtained with (5.5-10) follows. The problem of selecting the optimal γ for image restoration is discussed in some detail in Section 5.6.

Example: The first column in Fig. 5.5 shows three pictures of a domino corrupted by linear motion (at $-45°$ with respect to the horizontal) and noise whose variance at any point in the image was proportional to the brightness of the point. The three images were generated by varying the constant of proportionality so that the ratios of maximum brightness to noise amplitude were 1, 10, and 100, respectively, as shown on the left in Fig. 5.5. The Fourier spectra of the degraded images are shown in Fig. 5.5(b).

Since the effects of uniform linear motion can be expressed analytically, an equation describing $H(u, v)$ can be obtained without difficulty, as shown in Section 5.4.2. Figure 5.5(c) was obtained by direct inverse filtering following the procedure described in Section 5.4.1. The results are dominated by noise, but as the third image shows, the inverse filter successfully removed the degradation (blur) caused by motion. By contrast, Fig. 5.5(d) shows the results

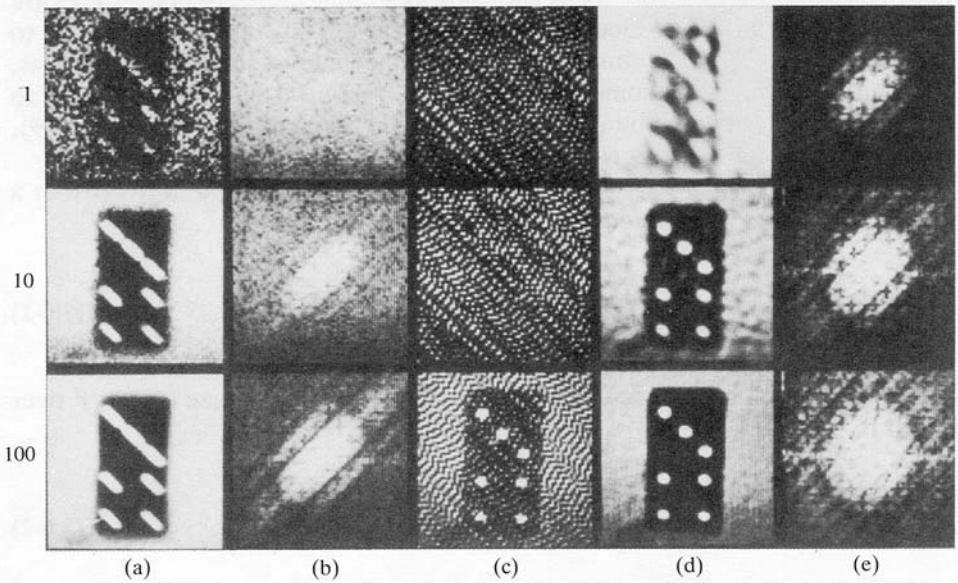


Figure 5.5 *Example of image restoration by inverse and Wiener filters: (a) degraded images and (b) their Fourier spectra; (c) images restored by inverse filtering; (d) images restored by Wiener filtering; (e) Fourier spectra of images in (d). (From Harris [1968].)*

obtained using Eq. (5.5-10) with $K = 2\sigma^2$, where σ^2 is the noise variance. The improvements over the direct inverse filtering approach are obvious, particularly for the third image. Figure 5.5(e) shows the Fourier spectra of the restored images. ❑

5.6 CONSTRAINED LEAST SQUARES RESTORATION

The least mean squares approach derived in Section 5.5 is a statistical procedure because the criterion for optimality is based on the correlation matrices of the image and noise functions. This implies that the results obtained by using a Wiener filter are optimal in an average sense. The restoration procedure developed in this section, however, is optimal for *each* given image and requires knowledge only of the noise mean and variance. Also considered is the problem of adjusting γ to satisfy the constraint leading to Eq. (5.3-9).

As indicated in Section 5.3.2, the restoration solution obtained by using Eq. (5.3-9) depends on the choice of the matrix $\mathbf{Q}$. Owing to ill-conditioning, that equation sometimes yields solutions that are obscured by large oscillating values. Therefore the feasibility of choosing $\mathbf{Q}$ so that these adverse effects are minimized is of interest. One possibility, suggested by Phillips [1962], is to formulate a criterion of optimality based on a measure of smoothness such as, for example, minimizing some function of the second derivative. In order to see how this criterion can be expressed in a form compatible with Eq. (5.3-9), let us first consider the 1-D case.

For a discrete function $f(x)$, $x = 0, 1, 2, \ldots$, the second derivative at a point x may be approximated by the expression

$$\frac{\partial^2 f(x)}{\partial x^2} \approx f(x + 1) - 2f(x) + f(x - 1). \tag{5.6-1}$$

A criterion based on this expression, then, might be to minimize $(\partial^2 f/\partial x^2)^2$ over x; that is,

$$\text{minimize} \left\{ \sum_x [f(x + 1) - 2f(x) + f(x - 1)]^2 \right\} \tag{5.6-2}$$

or, in matrix notation,

$$\text{minimize}\{\mathbf{f}^T\mathbf{C}^T\mathbf{C}\mathbf{f}\} \tag{5.6-3}$$

where

$$\mathbf{C} = \begin{bmatrix} 1 & & & & & & & \\ -2 & 1 & & & & & & \\ 1 & -2 & 1 & & & & & \\ & 1 & -2 & 1 & & & & \\ & & & \ddots & & & & \\ & & & & 1 & -2 & 1 & \\ & & & & & 1 & -2 & \\ & & & & & & 1 & \end{bmatrix} \tag{5.6-4}$$

is a "smoothing" matrix, and $\mathbf{f}$ is a vector whose elements are the samples of $f(x)$.

In the 2-D case we consider a direct extension of Eq. (5.6-1). In this case the criterion is to

$$\text{minimize}\left[\frac{\partial^2 f(x, y)}{\partial x^2} + \frac{\partial^2 f(x, y)}{\partial y^2}\right]^2 \tag{5.6-5}$$

where the derivative function is approximated by the expression

$$\begin{aligned} \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} &\approx [2f(x, y) - f(x + 1, y) - f(x - 1, y)] \\ &\quad + [2f(x, y) - f(x, y + 1) - f(x, y - 1)] \\ &\approx 4f(x, y) - [f(x + 1, y) + f(x - 1, y) + f(x, y + 1) \\ &\quad + f(x, y - 1)]. \end{aligned} \tag{5.6-6}$$

The derivative function given in Eq. (5.6-5) is the Laplacian operator discussed in Section 3.3.7.

Equation (5.6-6) can be implemented directly in a computer. However, the same operation can be carried out by convolving $f(x, y)$ with the operator

$$p(x, y) = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{bmatrix} \tag{5.6-7}$$

As indicated in Section 5.1.3, wraparound error in the discrete convolution process is avoided by extending $f(x, y)$ and $p(x, y)$. Having already considered

the formation of $f_e(x, y)$, we form $p_e(x, y)$ in the same manner:

$$p_e(x, y) = \begin{cases} p(x, y) & 0 \leq x \leq 2 \quad \text{and} \quad 0 \leq y \leq 2 \\ 0 & 3 \leq x \leq M - 1 \quad \text{or} \quad 3 \leq y \leq N - 1. \end{cases}$$

If $f(x, y)$ is of size $A \times B$, we choose $M \geq A + 3 - 1$ and $N \geq B + 3 - 1$, because $p(x, y)$ is of size 3×3.

The convolution of the extended functions then is

$$g_e(x, y) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} f_e(m, n) p_e(x - m, y - n) \tag{5.6-8}$$

which agrees with Eq. (5.1-23).

Following an argument similar to the one given in Section 5.1.3 allows expression of the smoothness criterion in matrix form. First, we construct a block-circulant matrix of the form

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_0 & \mathbf{C}_{M-1} & \mathbf{C}_{M-2} & \cdots & \mathbf{C}_1 \\ \mathbf{C}_1 & \mathbf{C}_0 & \mathbf{C}_{M-1} & \cdots & \mathbf{C}_2 \\ \mathbf{C}_2 & \mathbf{C}_1 & \mathbf{C}_0 & \cdots & \mathbf{C}_3 \\ \vdots & & & & \\ \mathbf{C}_{M-1} & \mathbf{C}_{M-2} & \mathbf{C}_{M-3} & \cdots & \mathbf{C}_0 \end{bmatrix} \tag{5.6-9}$$

where each submatrix $\mathbf{C}_j$ is an $N \times N$ circulant constructed from the jth row of $p_e(x, y)$; that is,

$$\mathbf{C}_j = \begin{bmatrix} p_e(j, 0) & p_e(j, N - 1) & \cdots & p_e(j, 1) \\ p_e(j, 1) & p_e(j, 0) & \cdots & p_e(j, 2) \\ \vdots & & & \\ p_e(j, N - 1) & p_e(j, N - 2) & \cdots & p_e(j, 0) \end{bmatrix} \tag{5.6-10}$$

Since $\mathbf{C}$ is block circulant, it is diagonalized by the matrix $\mathbf{W}$ defined in Section 5.2.2. In other words,

$$\mathbf{E} = \mathbf{W}^{-1}\mathbf{C}\mathbf{W} \tag{5.6-11}$$

where **E** is a diagonal matrix whose elements are given by

$$E(k, i) = \begin{cases} P\left(\left[\frac{k}{N}\right], k \bmod N\right) & \text{if } i = k \\ 0 & \text{if } i \neq k \end{cases} \tag{5.6-12}$$

as in Eq. (5.2-39). In this case $P(u, v)$ is the 2-D Fourier transform of $p_e(x, y)$. As with Eqs. (5.2-37) and (5.2-39), the assumption is that Eq. (5.6-12) has been scaled by the factor MN.

The convolution operation described above is equivalent to implementing Eq. (5.6-6), so the smoothness criterion of Eq. (5.6-5) takes the same form as Eq. (5.6-3):

$$\text{minimize}\{\mathbf{f}^T\mathbf{C}^T\mathbf{C}\mathbf{f}\} \tag{5.6-13}$$

where **f** is an MN-dimensional vector and **C** is of size $MN \times MN$. By letting $\mathbf{Q} = \mathbf{C}$, and recalling that $\|\mathbf{Qf}\|^2 = (\mathbf{Qf})^T(\mathbf{Qf}) = \mathbf{f}^T\mathbf{Q}^T\mathbf{Qf}$, this criterion may be expressed as

$$\text{minimize}\|\mathbf{Qf}\|^2 \tag{5.6-14}$$

which is the same form used in Section 5.3.2. In fact, if we require that the constraint $\|\mathbf{g} - \mathbf{H}\hat{\mathbf{f}}\|^2 = \|\mathbf{n}\|^2$ be satisfied, the optimal solution is given by Eq. (5.3-9) with $\mathbf{Q} = \mathbf{C}$:

$$\hat{\mathbf{f}} = (\mathbf{H}^T\mathbf{H} + \gamma\mathbf{C}^T\mathbf{C})^{-1}\mathbf{H}^T\mathbf{g}. \tag{5.6-15}$$

Using Eqs. (5.2-21), (5.2-23), and (5.6-11), allows Eq. (5.6-15) to be expressed as

$$\hat{\mathbf{f}} = (\mathbf{W}\mathbf{D}^*\mathbf{D}\mathbf{W}^{-1} + \gamma\mathbf{W}\mathbf{E}^*\mathbf{E}\mathbf{W}^{-1})^{-1}\mathbf{W}\mathbf{D}^*\mathbf{W}^{-1}\mathbf{g}. \tag{5.6-16}$$

Multiplying both sides by $\mathbf{W}^{-1}$ and performing some matrix manipulations reduces Eq. (5.6-16) to

$$\mathbf{W}^{-1}\hat{\mathbf{f}} = (\mathbf{D}^*\mathbf{D} + \gamma\mathbf{E}^*\mathbf{E})^{-1}\mathbf{D}^*\mathbf{W}^{-1}\mathbf{g}. \tag{5.6-17}$$

Keeping in mind that the elements inside the parentheses are diagonal and making use of the concepts developed in Section 5.2.3, allows expressing the elements of Eq. (5.6-17) in the form

$$\hat{F}(u, v) = \left[\frac{H^*(u, v)}{|H(u, v)|^2 + \gamma|P(u, v)|^2}\right]G(u, v) \tag{5.6-18}$$

for $u, v = 0, 1, 2, \ldots, N-1$, where $|H(u, v)|^2 = H^*(u, v)H(u, v)$, and we have assumed that $M = N$. Note that Eq. (5.6-18) resembles the parametric Wiener filter derived in Section 5.5. The principal difference between Eqs. (5.5-9) and (5.6-18) is that the latter does not require explicit knowledge of statistical parameters other than an estimate of the noise mean and variance.

The general formulation given in Eq. (5.3-9) requires that γ be adjusted to satisfy the constraint $\|\mathbf{g} - \mathbf{H}\hat{\mathbf{f}}\|^2 = \|\mathbf{n}\|^2$. Thus the solution given in Eq. (5.6-18) can be optimal only when γ satisfies this condition. An iterative procedure for estimating this parameter follows.

Define a residual vector $\mathbf{r}$ as

$$\mathbf{r} = \mathbf{g} - \mathbf{H}\hat{\mathbf{f}}. \tag{5.6-19}$$

Substituting Eq. (5.6-15) for $\hat{\mathbf{f}}$ yields

$$\mathbf{r} = \mathbf{g} - \mathbf{H}(\mathbf{H}^T\mathbf{H} + \gamma\mathbf{C}^T\mathbf{C})^{-1}\mathbf{H}^T\mathbf{g}. \tag{5.6-20}$$

Equation (5.6.20) indicates that $\mathbf{r}$ is a function of γ. In fact, it can be shown (Hunt [1973]) that

$$\begin{aligned}\phi(\gamma) &= \mathbf{r}^T\mathbf{r} \\ &= \|\mathbf{r}\|^2\end{aligned} \tag{5.6-21}$$

is a monotonically increasing function of γ. What we want to do is adjust γ so that

$$\|\mathbf{r}\|^2 = \|\mathbf{n}\|^2 \pm a, \tag{5.6-22}$$

where a is an accuracy factor. Clearly, if $\|\mathbf{r}\|^2 = \|\mathbf{n}\|^2$ the constraint $\|\mathbf{g} - \mathbf{H}\hat{\mathbf{f}}\|^2 = \|\mathbf{n}\|^2$ will be strictly satisfied, in view of Eq. (5.6-19).

Because $\phi(\gamma)$ is monotonic, finding a γ that satisfies Eq. (5.6-17) is not difficult. One simple approach is to

(1) specify an initial value of γ;

(2) compute $\hat{\mathbf{f}}$ and $\|\mathbf{r}\|^2$; and

(3) stop if Eq. (5.6-22) is satisfied; otherwise return to step 2 after increasing γ if $\|\mathbf{r}\|^2 < \|\mathbf{n}\|^2 - a$ or decreasing γ if $\|\mathbf{r}\|^2 > \|\mathbf{n}\|^2 + a$.

Other procedures such as a Newton–Raphson algorithm can be used to improve speed of convergence.

Implementation of these concepts requires some knowledge about $\|\mathbf{n}\|^2$. The variance of $\eta_e(x, y)$ is

$$\begin{aligned}\sigma_\eta^2 &= E\{[\eta_e(x, y) - \overline{\eta}_e]^2\} \\ &= E[\eta_e^2(x, y)] - \overline{\eta}_e^2\end{aligned} \tag{5.6-23}$$

where

$$\bar{\eta}_e = \frac{1}{(M-1)(N-1)} \sum_x \sum_y \eta_e(x, y) \tag{5.6-24}$$

is the mean value of $\eta_e(x, y)$. If a sample average is used to approximate the expected value of $\eta_e^2(x, y)$, Eq. (5.6-23) becomes

$$\sigma_\eta^2 = \frac{1}{(M-1)(N-1)} \sum_x \sum_y \eta_e^2(x, y) - \bar{\eta}_e^2. \tag{5.6-25}$$

The summation term simply indicates squaring and adding all values in the array $\eta_e(x, y)$, $x = 0, 1, 2, \ldots, M-1$, and $y = 0, 1, 2, \ldots, N-1$. This manipulation is simply the product $\mathbf{n}^T\mathbf{n}$, which, by definition, equals $\|\mathbf{n}\|^2$. Thus Eq. (5.6-25) reduces to

$$\sigma_\eta^2 = \frac{\|\mathbf{n}\|^2}{(M-1)(N-1)} - \bar{\eta}_e^2 \tag{5.6-26}$$

or

$$\|\mathbf{n}\|^2 = (M-1)(N-1)[\sigma_\eta^2 + \bar{\eta}_e^2]. \tag{5.6-27}$$

The importance of this equation is that it allows determination of a value for the constraint in terms of the noise mean and variance, quantities that, if not known, can often be approximated or measured in practice.

The constrained least squares restoration procedure can be summarized as follows.

Step 1. Choose an initial value of γ and obtain an estimate of $\|\mathbf{n}\|^2$ by using Eq. (5.6-27).

Step 2. Compute $\hat{F}(u, v)$ using Eq. (5.6-18). Obtain $\hat{\mathbf{f}}$ by taking the inverse Fourier transform of $\hat{F}(u, v)$.

Step 3. Form the residual vector $\mathbf{r}$ according to Eq. (5.6-19) and compute $\phi(\gamma) = \|\mathbf{r}\|^2$.

Step 4. Increment or decrement γ.

(a) $\phi(\gamma) < \|\mathbf{n}\|^2 - a$. Increment γ according to the algorithm given above or other appropriate method (such as a Newton–Raphson procedure).

(b) $\phi(\gamma) > \|\mathbf{n}\|^2 + a$. Decrement γ according to an appropriate algorithm.

Step 5. Return to step 2 and continue unless step 6 is true.

Step 6. $\phi(\gamma) = \|\mathbf{n}\|^2 \pm a$, where a determines the accuracy with which the constraint is satisfied. Stop the estimation procedure, with $\hat{\mathbf{f}}$ for the present value of γ being the restored image.

Example: Figure 5.6(b) was obtained by convolving the Gaussian-shaped point spread function

$$h(x, y) = \exp\left(-\frac{x^2 + y^2}{2400}\right)$$

with the original image shown in Fig. 5.6(a) and adding noise drawn from a uniform distribution in the interval [0, 0.5]. Figure 5.6(c) shows the result of

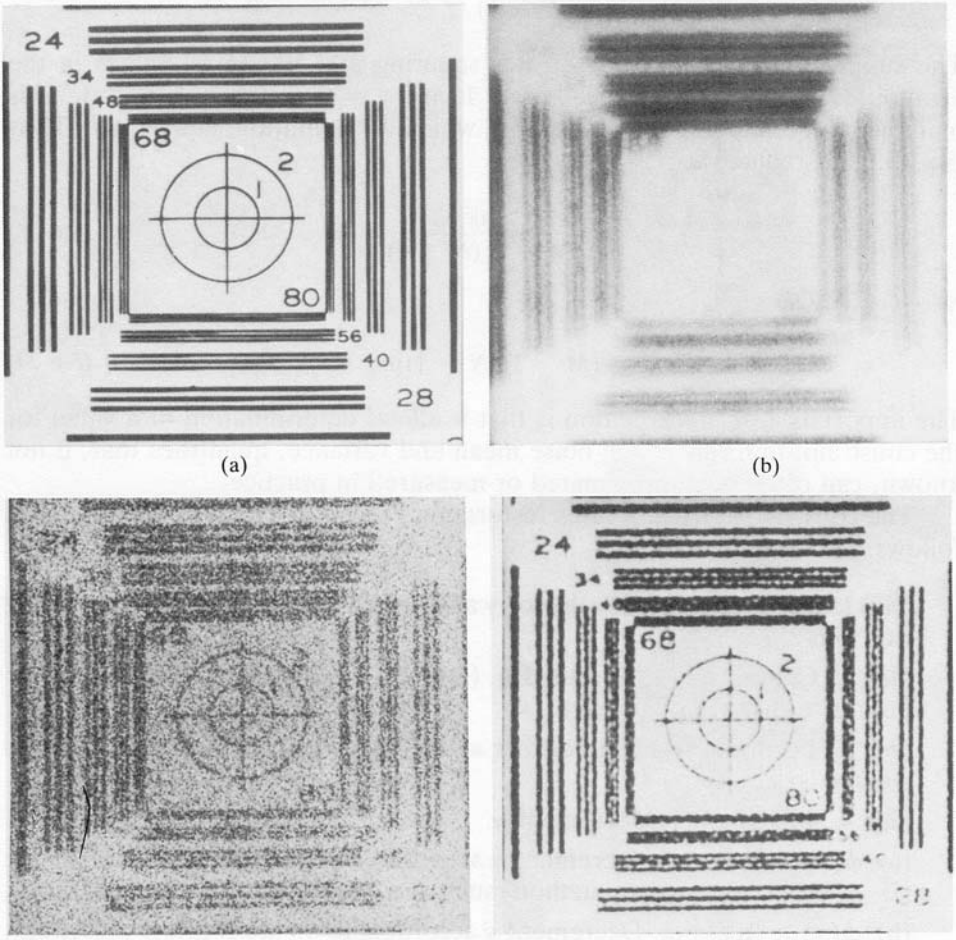


Figure 5.6 *(a) Original image; (b) image blurred and corrupted by additive noise; (c) image restored by inverse filtering; (d) image restored by the method of constrained least squares. (From Hunt [1973].)*

using the algorithm with $\gamma = 0$ (inverse filter). The ill-conditioned nature of the solution is evident by the dominance of the noise on the restored image. Figure 5.6(d) was obtained by using the preceding algorithm to seek a γ that would satisfy the constraint. The variance and mean of the uniform density in the interval [0, 0.5] were used to estimate $\|\mathbf{n}\|^2$, and the accuracy factor a was chosen so that $a = 0.025\|\mathbf{n}\|^2$. The improvement of the constrained solution over direct inverse filtering is clearly visible. ❑

5.7 INTERACTIVE RESTORATION

So far, we have focused on a strictly analytical approach to restoration. In many applications, the practical approach is to take advantage of human intuition, coupled with the versatility of a digital computer, to restore images interactively. In this case, the observer controls the restoration process and, by "tuning" the available parameters, is able to obtain a final result that may be quite adequate for a specific purpose.

One of the simplest cases of image corruption that lends itself well to interactive restoration is the occurrence of a 2-D sinusoidal interference pattern (often called *coherent noise*) superimposed on an image. Let $\eta(x, y)$ denote a sinusoidal interference pattern of amplitude A and frequency components (u_0, v_0); that is,

$$\eta(x, y) = A \sin(u_0 x + v_0 y). \tag{5.7-1}$$

Direct substitution of Eq. (5.7-1) into Eq. (3.1-9) yields the Fourier transform of $\eta(x, y)$:

$$N(u, v) = \frac{-jA}{2}\left[\delta\left(u - \frac{u_0}{2\pi}, v - \frac{v_0}{2\pi}\right) - \delta\left(u + \frac{u_0}{2\pi}, v + \frac{v_0}{2\pi}\right)\right]. \tag{5.7-2}$$

In other words, the Fourier transform of a 2-D sine function is a pair of impulses of strength $-A/2$ and $A/2$ located at coordinates $(u_0/2\pi, v_0/2\pi)$ and $(-u_0/2\pi, -v_0/2\pi)$, respectively, of the frequency plane. In this case the transform has only imaginary components.

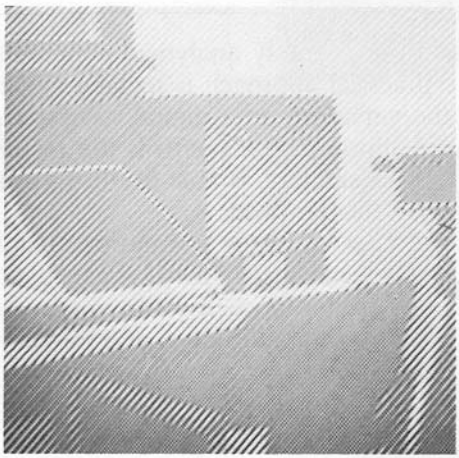
With the only degradation considered being additive noise, it follows from Eq. (5.2-40) that

$$G(u, v) = F(u, v) + N(u, v). \tag{5.7-3}$$

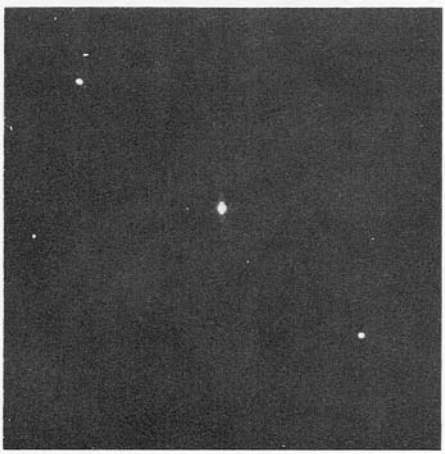
A display of the magnitude of $G(u, v)$ contains the magnitude of the sum of $F(u, v)$ and $N(u, v)$. If A is large enough, the two impulses of $N(u, v)$ usually appear as bright dots on the display, especially if they are located relatively far from the origin so that the contribution of the components of $F(u, v)$ is small.

If $\eta(x, y)$ were known completely, the original image, of course, could be recovered by subtracting the interference from $g(x, y)$. As this situation is seldom the case, a useful approach is to identify visually the location of impulse components in the frequency domain and use a bandreject filter (see Section 4.6.3) at these locations.

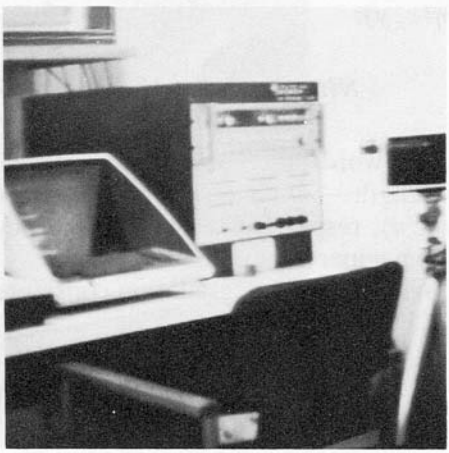
Example: The image shown in Fig. 5.7(a) was corrupted by a sinusoidal pattern of the form shown in Eq. (5.7-1). The Fourier spectrum of this image, shown



(a)

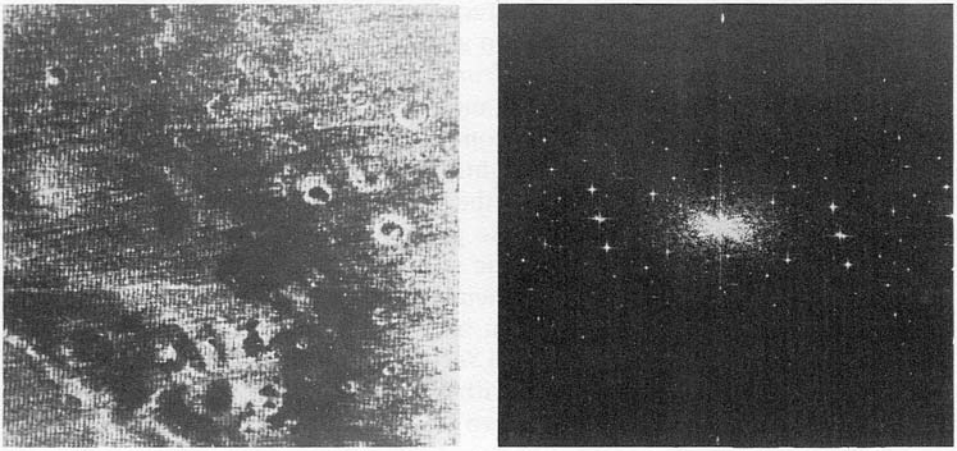


(b)



(c)

Figure 5.7 *Example of sinusoidal interference removal: (a) corrupted image; (b) Fourier spectrum showing impulses due to sinusoidal pattern; (c) image restored by using a band-reject filter with a radius of 1.*



(a) (b)

Figure 5.8 *(a) Picture of the Martian terrain taken by Mariner 6; (b) Fourier spectrum. Note the periodic interference in the image and the corresponding spikes in the spectrum. (Courtesy of NASA, Jet Propulsion Laboratory.)*

in Fig. 5.7(b), clearly exhibits a pair of symmetric impulses resulting from sinusoidal interference. Figure 5.7(c) was obtained by manually placing (from a computer console) two bandreject filters of radius 1 at the location of the impulses and then taking the inverse Fourier transform of the result. For all practical purposes, the restored image is free of interference. ❑

The presence of a single, clearly defined interference pattern, such as the one just illustrated, seldom occurs in practice. Notable examples are images derived from electro-optical scanners, such as those commonly used in space missions. A common problem with these sensors is interference caused by coupling and amplification of low-level signals in the electronic circuitry. As a result, images reconstructed from the scanner output tend to contain a pronounced, 2-D periodic structure superimposed on the scene data.

Figure 5.8(a), an example of this type of periodic image degradation, shows a digital image of the Martian terrain taken by the *Mariner 6* spacecraft. The interference pattern is quite similar to the one shown in Fig. 5.7(a), but the former pattern is considerably more subtle and, consequently, harder to detect in the frequency plane.

Figure 5.8(b) shows the Fourier spectrum of the image in question. The starlike components were caused by the interference, and several pairs of components are present, indicating that the pattern was composed of more than just one sinusoidal component. When several interference components are present, the method discussed above is not always acceptable because it may remove too much image information in the filtering process. In addition, these

components generally are not single-frequency bursts. Instead, they tend to have broad skirts that carry information about the interference pattern. These skirts are not always easily detectable from the normal transform background.

A procedure that has found acceptance in processing space-related scenes consists of first isolating the principal contributions of the interference pattern and then subtracting a variable, weighted portion of the pattern from the corrupted image. Although we develop the procedure in the context of a specific application, the basic approach is quite general and can be applied to other restoration tasks when multiple periodic interference is a problem.

The first step is to extract the principal frequency components of the interference pattern. This extraction can be done by placing a bandpass filter $H(u, v)$ at the location of each spike (see Section 4.6.3). If $H(u, v)$ is constructed to pass only components associated with the interference pattern, it follows that the Fourier transform of the pattern is given by the relation

$$P(u, v) = H(u, v)G(u, v) \tag{5.7-4}$$

where $G(u, v)$ is the Fourier transform of the corrupted image $g(x, y)$ and, for $N \times N$ digitization, u and v take on values in the range $0, 1, \ldots, N - 1$.

Formation of $H(u, v)$ requires considerable judgment about what is or is not an interference spike. For this reason, the bandpass filter generally is constructed interactively by observing the spectrum of $G(u, v)$ on a display. After a particular filter has been selected, the corresponding pattern in the spatial domain is obtained from the expression

$$p(x, y) = \mathfrak{F}^{-1}\{H(u, v)G(u, v)\}. \tag{5.7-5}$$

Because the corrupted image is formed by the addition of $f(x, y)$ and the interference, if $p(x, y)$ were known completely, subtracting the pattern from $g(x, y)$ to obtain $f(x, y)$ would be a simple matter. The problem, of course, is that this filtering procedure usually yields only an approximation of the true pattern. The effects of components not present in the estimate of $p(x, y)$ can be minimized by instead subtracting from $g(x, y)$ a weighted portion of $p(x, y)$ to obtain an estimate of $f(x, y)$:

$$\hat{f}(x, y) = g(x, y) - w(x, y)p(x, y) \tag{5.7-6}$$

where $\hat{f}(x, y)$ is the estimate of $f(x, y)$ and $w(x, y)$ is to be determined. The function $w(x, y)$ is called a *weighting* or *modulation* function, and the objective of the procedure is to select this function so that the result is optimized in some meaningful way. One approach is to select $w(x, y)$ so that the variance of $\hat{f}(x, y)$ is minimized over a specified neighborhood of every point (x, y).

Consider a neighborhood of size $(2X + 1)$ by $(2Y + 1)$ about a point (x, y). The "local" variance of $\hat{f}(x, y)$ at coordinates (x, y) is

$$\sigma^2(x, y) = \frac{1}{(2X + 1)(2Y + 1)} \sum_{m=-X}^{X} \sum_{n=-Y}^{Y} [\hat{f}(x + m, y + n) - \bar{\hat{f}}(x, y)]^2 \tag{5.7-7}$$

where $\bar{\hat{f}}(x, y)$ is the average value of $\hat{f}(x, y)$ in the neighborhood; that is,

$$\bar{\hat{f}}(x, y) = \frac{1}{(2X + 1)(2Y + 1)} \sum_{m=-X}^{X} \sum_{n=-Y}^{Y} \hat{f}(x + m, y + n). \tag{5.7-8}$$

Points on or near the edge of the image can be treated by considering partial neighborhoods.

Substituting Eq. (5.7-6) into Eq. (5.7-7) yields

$$\begin{aligned}\sigma^2(x,y) = \frac{1}{(2X+1)(2Y+1)} \sum_{m=-X}^{X} \sum_{n=-Y}^{Y} \{[g(x+m,y+n) \\ - w(x+m,y+n)p(x+m,y+n)] - [\bar{g}(x,y) - \overline{w(x,y)p(x,y)}]\}^2.\end{aligned} \tag{5.7-9}$$

Assuming that $w(x, y)$ remains essentially constant over the neighborhood gives the approximations

$$w(x + m, y + n) = w(x, y) \tag{5.7-10}$$

for $-X \leq m \leq X$ and $-Y \leq n \leq Y$; also

$$\overline{w(x, y)p(x, y)} = w(x, y)\bar{p}(x, y) \tag{5.7-11}$$

in the neighborhood. With these approximations, Eq. (5.7-9) becomes

$$\begin{aligned}\sigma^2(x, y) = \frac{1}{(2X + 1)(2Y + 1)} \sum_{m=-X}^{X} \sum_{n=-Y}^{Y} \{[g(x + m, y + n) \\ - w(x, y)p(x + m, y + n)] - [\bar{g}(x, y) - w(x, y)\bar{p}(x,y)]\}^2.\end{aligned} \tag{5.7-12}$$

To minimize $\sigma^2(x, y)$ we solve

$$\frac{\partial \sigma^2(x, y)}{\partial w(x, y)} = 0 \tag{5.7-13}$$

for $w(x, y)$. The result is

$$w(x, y) = \frac{\overline{g(x, y)p(x, y)} - \bar{g}(x, y)\bar{p}(x, y)}{\overline{p^2}(x, y) - \bar{p}^2(x, y)}. \tag{5.7-14}$$

To obtain the restored image $\hat{f}(x, y)$ we compute $w(x, y)$ from Eq. (5.7-14) and then make use of Eq. (5.7-6). As $w(x, y)$ is assumed to be constant in a neighborhood, computing this function for every value of x and y in the image is unnecessary. Instead, $w(x, y)$ is computed for *one* point in each nonoverlapping neighborhood (preferably the center point) and then used to process all the image points contained in that neighborhood.

Example: Figures 5.9 through 5.11 show the result of applying the above technique to the image shown in Fig. 5.8(a). In this case $N = 512$ and a neighborhood with $X = Y = 15$ was selected. Figure 5.9 shows the Fourier spectrum of the corrupted image, but the origin was not shifted to the center of the frequency plane. Figure 5.10(a) shows the spectrum of $P(u, v)$, where only the noise spikes are present. Figure 5.10(b) shows the interference pattern $p(x, y)$ obtained by taking the inverse Fourier transform of $P(u, v)$. Note the similarity between this pattern and the structure of the noise present in Fig. 5.8(a). Finally, Fig. 5.11 shows the processed image obtained by using Eq. (5.7-6). The periodic

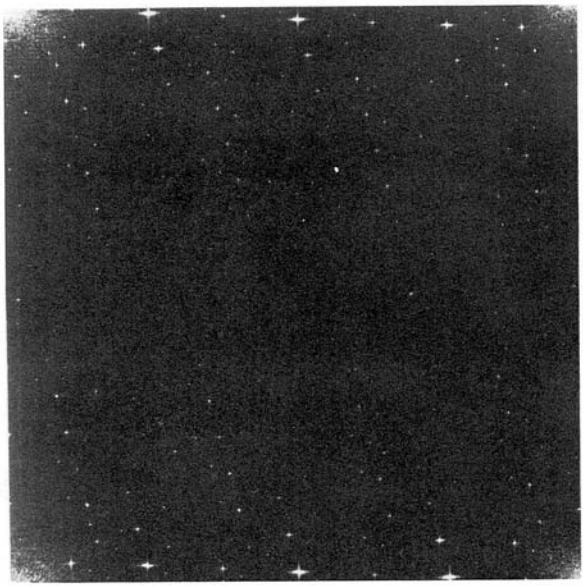


Figure 5.9 *Fourier spectrum (without shifting) of the image shown in Fig. 5.8(a). (Courtesy of NASA, Jet Propulsion Laboratory.)*

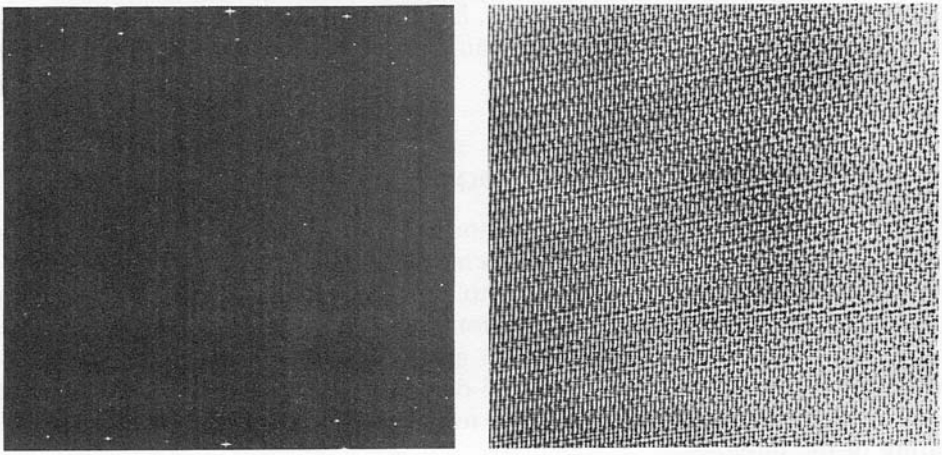


Figure 5.10 *(a) Fourier spectrum of $P(u, v)$; (b) corresponding interference pattern $p(x, y)$. (Courtesy of NASA, Jet Propulsion Laboratory.)*

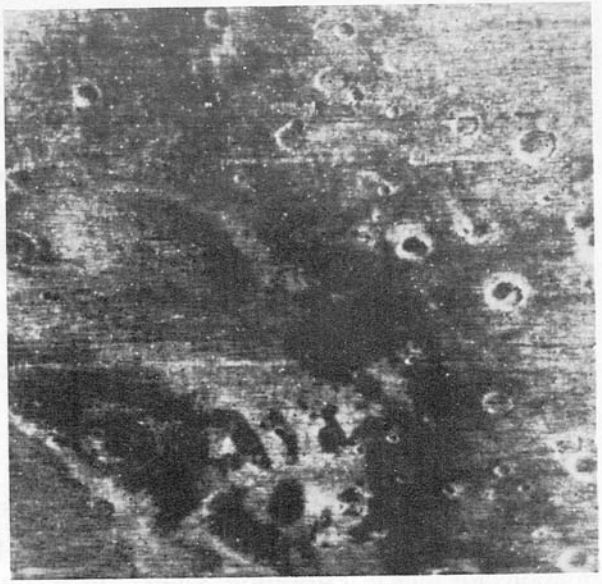


Figure 5.11 *Processed image. (Courtesy of NASA, Jet Propulsion Laboratory.)*

interference, for all practical purposes, has been removed, leaving only spotty noise that is not periodic. This noise can be processed by other methods, such as median filtering. ❑

5.8 RESTORATION IN THE SPATIAL DOMAIN

After a suitable frequency domain restoration filter has been obtained by any of the methods discussed earlier, implementing the solution in the spatial domain via a convolution mask in order to expedite processing (see Section 4.1) often is desirable. As indicated in Section 4.5, the coefficients of a convolution mask can be obtained directly from a given filter function via Eq. (4.5-12). Although the discussion in Section 4.5 deals with enhancement, the concepts developed there are equally applicable to restoration; the difference lies in the nature of the filter.

Example: Figure 5.12(a) shows an infrared image of a set of military targets in a field. The image is corrupted by nearly periodic scanner interference, visible as a "ripple" effect in the vertical direction. Because of its periodic nature, the interference produces bursts of concentrated energy in the vertical axis of the Fourier spectrum of the image, as shown in Fig. 5.13(a).

A simple approach for reducing the effect of the interference is to use a notch filter, $H(u, v)$, which attenuates the values of the Fourier transform in the vertical axis and multiplies all other values of the transform by 1, in a manner analogous to the procedure discussed in Section 5.7. Figure 5.13(b) shows such a filter superimposed on the spectrum, where the dark bands are the attenuated regions.

Figure 5.12(b) shows the result of using the notch filter and taking the inverse Fourier transform. Note that, for all practical purposes, the interference was eliminated from the image. The image shown in Fig. 5.12(c) was obtained by applying a 9×9 convolution mask (see Section 4.1) to the original, corrupted image. The coefficients of this mask were generated from the notch filter by using Eq. (4.5-12). This small mask is only an approximation of the Fourier filtering process, so some vertical lines are still visible in the processed image. A second pass of the mask further reduced the interference (at the cost of some noticeable blurring), as Fig. 5.12(d) shows. ❑

5.9 GEOMETRIC TRANSFORMATIONS

We conclude this chapter with an introductory discussion on the use of geometric transformations for image restoration. Unlike the techniques discussed so far, geometric transformations generally modify the spatial relationships

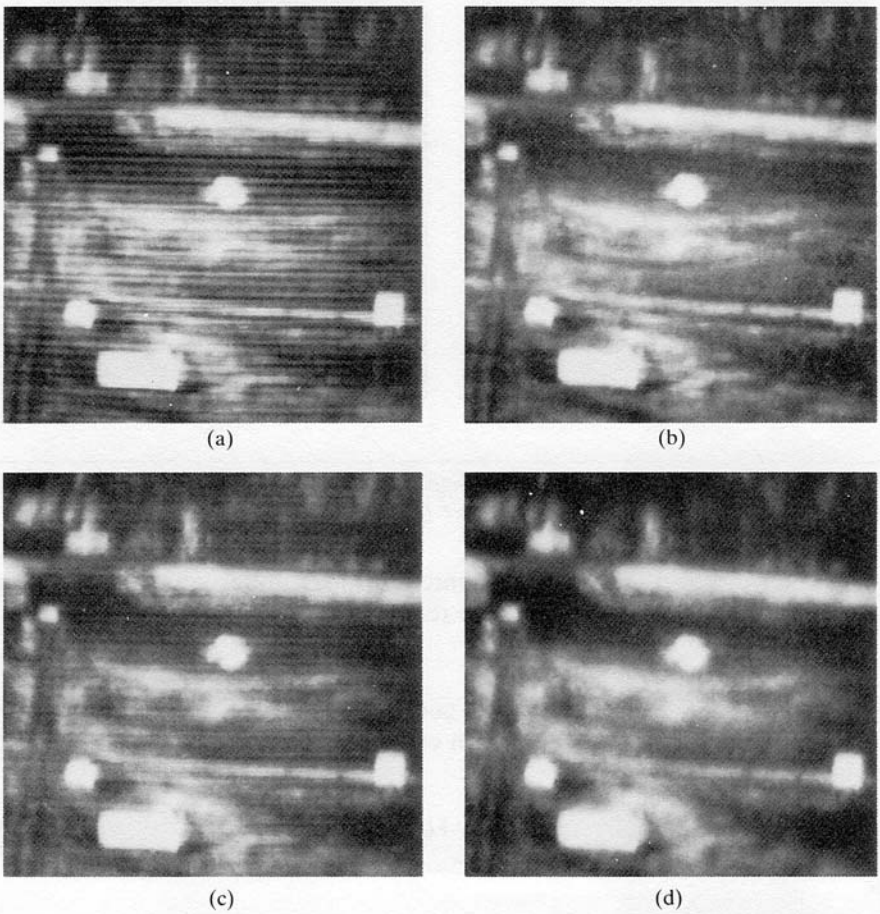
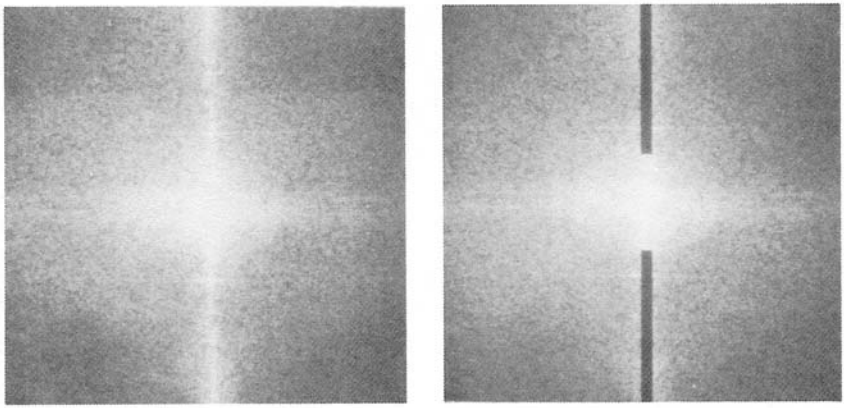


Figure 5.12 *(a) Infrared image showing interference; (b) image restored using a notch filter in the frequency domain; (c) image restored using a 9 × 9 convolution mask; (d) result of applying the mask a second time. (From Meyer and Gonzalez [1983].)*

between pixels in an image. Geometric transformations often are called *rubber-sheet transformations*, because they may be viewed as the process of "printing" an image on a sheet of rubber and then stretching this sheet according to some predefined set of rules.

In terms of digital image processing, a geometric transformation consists of two basic operations: (1) a *spatial transformation*, which defines the "re-arrangement" of pixels on the image plane; and (2) a *gray-level interpolation*, which deals with the assignment of gray levels to pixels in the spatially trans-



(a) (b)

Figure 5.13 *(a) Fourier spectrum of the image in Fig. 5.12(a); (b) Notch filter superimposed on the spectrum. (From Meyer and Gonzalez [1983].)*

formed image. We discuss the fundamental ideas underlying these concepts, and their use in the context of image restoration, in the following sections.

5.9.1 Spatial Transformations

Suppose that an image f with pixel coordinates (x, y) undergoes geometric distortion to produce an image g with coordinates $(\hat{x}, \hat{y})$. This transformation may be expressed as

$$\hat{x} = r(x, y) \tag{5.9-1}$$

and

$$\hat{y} = s(x, y) \tag{5.9-2}$$

where $r(x, y)$ and $s(x, y)$ represent the spatial transformations that produced the geometrically distorted image $g(\hat{x}, \hat{y})$. For example, if $r(x, y) = x/2$ and $s(x, y) = y/2$, the "distortion" is simply a shrinking of the size of $f(x, y)$ by one-half in both spatial directions.

If $r(x, y)$ and $s(x, y)$ were known analytically, recovering $f(x, y)$ from the distorted image $g(\hat{x}, \hat{y})$ by applying the transformations in reverse might be possible theoretically. In practice, however, formulating analytically a single set of functions $r(x, y)$ and $s(x, y)$ that describe the geometric distortion process over the entire image plane generally is not possible. The method used most frequently to overcome this difficulty is to formulate the spatial relocation of pixels by the use of *tiepoints*, which are a subset of pixels whose location in the input (distorted) and output (corrected) images is known precisely.

Figure 5.14 shows quadrilateral regions in a distorted and corresponding corrected image. The vertices of the quadrilaterals are corresponding tiepoints. Suppose that the geometric distortion process within the quadrilateral regions is modeled by a pair of bilinear equations so that

$$r(x, y) = c_1x + c_2y + c_3xy + c_4 \tag{5.9-3}$$

and

$$s(x, y) = c_5x + c_6y + c_7xy + c_8. \tag{5.9-4}$$

Then, from Eqs. (5.9-1) and (5.9-2),

$$\hat{x} = c_1x + c_2y + c_3xy + c_4 \tag{5.9-5}$$

and

$$\hat{y} = c_5x + c_6y + c_7xy + c_8. \tag{5.9-6}$$

Since there are a total of eight known tiepoints, these equations can be easily solved for the eight coefficients c_i, $i = 1, 2, \ldots, 8$. The coefficients constitute the model used to transform *all* pixels within the quadrilateral region characterized by the tiepoints used to obtain the coefficients. In general, enough tiepoints are needed to generate a set of quadrilaterals that cover the entire image, with each quadrilateral having its own set of coefficients.

The procedure used to generate the corrected image is straightforward. For example, to generate $f(0, 0)$, substitute $(x, y) = (0, 0)$ into Eqs. (5.9-5) and (5.9-6) and obtain a pair of coordinates $(\hat{x}, \hat{y})$ from those equations. Then, let $f(0, 0) = g(\hat{x}, \hat{y})$, where $\hat{x}$ and $\hat{y}$ are the coordinate values just obtained. Next, substitute $(x, y) = (0, 1)$ into Eqs. (5.9-5) and (5.9-6), obtain another pair of values $(\hat{x}, \hat{y})$, and let $f(0, 1) = g(\hat{x}, \hat{y})$ for those coordinate values. The procedure continues pixel by pixel and row by row until an array whose size does

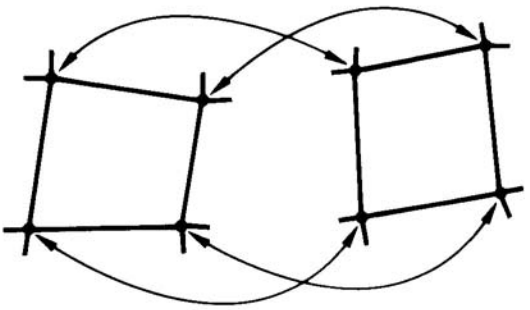


Figure 5.14 *Corresponding tiepoints in two image segments.*

not exceed the size of image g is obtained. A column (rather than a row) scan would yield identical results. Also, a bookkeeping procedure is essential to keep track of which quadrilaterals apply at a given pixel location in order to use the proper coefficients.

5.9.2 Gray-Level Interpolation

The method just discussed steps through integer values of the coordinates (x, y) to yield the corrected image $f(x, y)$. However, depending on the coefficients c_i, Eqs. (5.9-5) and (5.9-6) can yield noninteger values for $\hat{x}$ and $\hat{y}$. Because the distorted image g is digital, its pixel values are defined only at integer coordinates. Thus using noninteger values for $\hat{x}$ and $\hat{y}$ causes a mapping into locations of g for which no gray levels are defined. Inferring what the gray-level values at those locations should be, based only on the pixel values at integer coordinate locations, then becomes necessary. The technique used to accomplish this is called *gray-level interpolation*.

The simplest scheme for gray-level interpolation is based on a nearest neighbor approach. This method, also called *zero-order interpolation*, is illustrated in Fig. 5.15. This figure shows: (1) the mapping of integer coordinates (x, y) into fractional coordinates $(\hat{x}, \hat{y})$ by means of Eqs. (5.9-5) and (5.9-6); (2) the selection of the closest integer coordinate neighbor to $(\hat{x}, \hat{y})$; and (3) the assignment of the gray level of this nearest neighbor to the pixel located at (x, y).

Although nearest neighbor interpolation is simple to implement, this method often has the drawback of producing undesirable artifacts, such as distortion of straight edges in images of fine resolution. Smoother results can be obtained by using more sophisticated techniques, such as *cubic convolution interpolation* (Bernstein [1976]), which fits a surface of the $(\sin x)/x$ type through a much larger number of neighbors (say, 16) in order to obtain a smooth estimate of the gray level at any desired point. However, from a computational point of view this technique is costly, and a reasonable compromise is to use

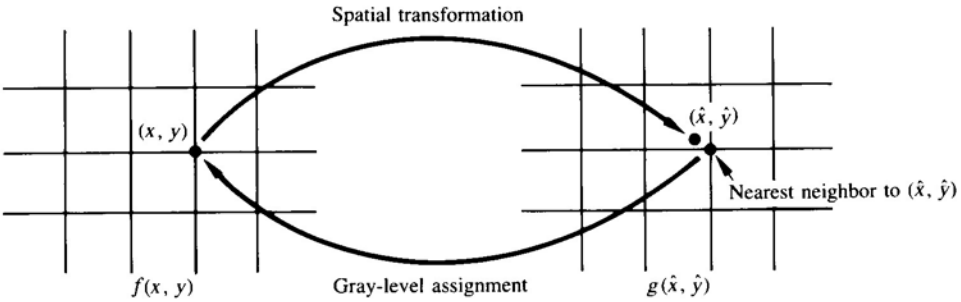
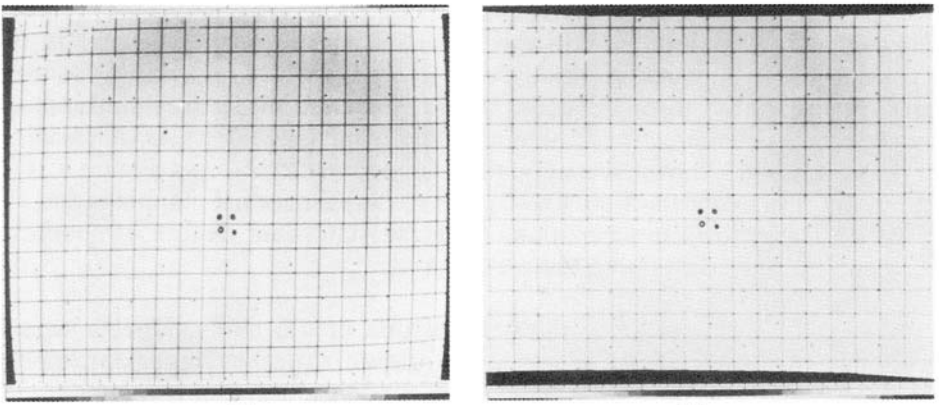


Figure 5.15 *Gray-level interpolation based on the nearest neighbor concept.*



(a) (b)

Figure 5.16 *(a) Distorted image; (b) image after geometric correction. (From O'Handley and Green [1972].)*

a *bilinear interpolation* approach that uses the gray levels of the four nearest neighbors. In other words, the idea is that the gray level of each of the four integral nearest neighbors of a nonintegral pair of coordinates $(\hat{x}, \hat{y})$ is known. The gray-level value of $(\hat{x}, \hat{y})$, denoted $v(\hat{x}, \hat{y})$, can then be interpolated from the values of its neighbors by using the relationship

$$v(\hat{x}, \hat{y}) = a\hat{x} + b\hat{y} + c\hat{x}\hat{y} + d \tag{5.9-7}$$

where the four coefficients are easily determined from the four equations in four unknowns that can be written using the four known neighbors of $(\hat{x}, \hat{y})$. When these coefficients have been determined, $v(\hat{x}, \hat{y})$ is computed and this value is assigned to the location in $f(x, y)$ which yielded the spatial mapping into location $(\hat{x}, \hat{y})$. It is easy to visualize this procedure with the aid of Fig. 5.15. The exception is that, instead of using the gray-level value of the nearest neighbor to $(\hat{x}, \hat{y})$, we actually interpolate a value at location $(\hat{x}, \hat{y})$ and use this value for the gray-level assignment at (x, y).

Example: The methods developed in this section and Section (5.9.1) can be illustrated by applying these techniques to the problem of correcting an image that has been distorted geometrically. The image in question is shown in Fig. 5.16(a). This image exhibits the "barrel" distortion found in many vidicon-based imaging cameras. The rectilinear grid shown in Fig. 5.16(a) is severely distorted, particularly near the edges of the image. Note also that the distortion is not uniform and that the degree of distortion increases nonlinearly as a function of distance from the center of the image.

As indicated in Section 5.9.1, the use of Eqs. (5.9-5) and (5.9-6) requires knowledge of tiepoints in both the distorted and corrected images. In this particular case, tiepoints are the *reseau marks* visible in Fig. 5.16(a) as small dark dots scattered throughout the image. (Reseau marks are small metallic squares embedded directly on the surface of the imaging tube.) As the locations of these marks are known precisely, they serve as ideal tiepoints. Figure 5.16(b) shows the result of using Eqs. (5.9-5) and (5.9-6) for spatial mappings and Eq. (5.9-7) for gray-level interpolation. Note the significant degree of geometric correction achieved by using these equations. ❑

The preceding example indicates but one of the many possible uses of geometric transformations for image restoration. Another important application is *image registration*, or finding correspondence between two images. The procedure for image registration is the same as the method just illustrated for geometric correction. However, the emphasis is on transforming an image so that it will correspond with another image of the same scene but viewed perhaps from another perspective. Other applications of the techniques discussed in this section include rectification of display distortions, map projections, and cartographic projections. The books by Castleman [1979] and Green [1983] contain numerous examples of these applications.

Establishing corresponding tiepoints in two images in many cases can be a rather difficult task. Not every situation is characterized by the availability of controlled artifacts such as reseau marks. When marks are not known a priori, tiepoints are usually established by using correlation techniques (see Chapter 9) to find corresponding features in two images. However, correlation measures are affected by factors such as noise and image rotation and thus generally yield less precise spatial correspondences between tiepoints.

5.10 CONCLUDING REMARKS

The principal concepts developed in this chapter are a formulation of the image restoration problem in the framework of linear algebra and the subsequent simplification of algebraic solutions based on the properties of circulant and block-circulant matrices.

Most of the restoration techniques derived in preceding sections are based on a least squares criterion of optimality. The use of the word *optimal* in this context refers strictly to a mathematical concept, not to optimal response of the human visual system. In fact, the present lack of knowledge about visual perception precludes a general formulation of the image restoration problem that takes into account observer preferences and capabilities. In view of these limitations, the advantage of the procedures followed in this chapter is the development of a basic approach from which a set of previously known (but

not unified) results can be derived. Thus the power of the algebraic approach is evident in the simplicity by which methods such as the Wiener and constrained least squares filters can be obtained, starting from the same basic principles.

The key points leading to the results in the first eight sections of this chapter are based on the assumption of linear, space invariant degradations. This assumption leads immediately to the convolution integral, whose discrete formulation can be expressed in terms of the basic degradation model given in Eq. (5.1-24). The assumed periodicity of the input functions further simplified the problem by producing circulant and block-circulant matrices. In terms of implementation, these matrices allow all the derived restoration techniques to be carried out in the frequency domain by means of a 2-D FFT algorithm, thus greatly reducing the computational complexity posed by the original matrix formulation of the degradation process.

The material in Section 5.8 provides a convenient way to implement in the spatial domain an approximation of the results in Sections 5.2–5.7. Finally, the discussion in Section 5.9 introduces the problem of restoring images that have been distorted geometrically.

REFERENCES

The definitions given in Section 5.1 were adapted from Schwarz and Friedland [1965]. Background for most of the basic matrix operations used in this chapter is contained in Deutsch [1965], Noble [1969], and Bellman [1970]. Development of the discrete degradation model in terms of circulant and block-circulant matrices is based on two papers by Hunt [1971, 1973]. These papers and the book by Bellman [1970] also consider the diagonalization properties discussed in Section 5.2. For additional information on the material of Section 5.3, as well as the algebraic derivation of the various restoration techniques used in this chapter, see Andrews and Hunt [1977]. That book, devoted entirely to the topic of image restoration, treats in detail other restoration techniques in addition to those developed here.

Numerous investigators have considered the inverse filtering approach. References for the material in Section 5.4 are McGlamery [1967], Sondhi [1972], Cutrona and Hall [1968], and Slepian [1967]. Additional references on the least squares restoration approach discussed in Section 5.5 are Helstrom [1967], Slepian [1967], Harris [1968], Rino [1969], Horner [1969], and Rosenfeld and Kak [1982]. Comparison of the classical derivations in these references with the algebraic approach in Section 5.5 is interesting. The material in Section 5.6 is based on a paper by Hunt [1973]. Other references related to the topics discussed in Sections 5.1–5.7 are Slepian and Pollak [1961], Phillips [1962], Twomey [1963], Shack [1964], Lohman and Paris [1965], Harris [1966], Mueller and Reynolds [1967], Blackman [1968], Huang [1968], Rushforth and Harris [1968], MacAdam [1970], Falconer [1970], Som [1971], Frieden [1972, 1974], Habibi [1972], Sawchuck [1972], Robbins and Huang [1972], Andrews [1974], Jain and Angel [1974], and Anderson and Netravali [1976]. The material in Section 5.8 is from Meyer and Gonzalez [1983]. Additional reading for the topics in Section 5.9 may be found in O'Handley and Green [1972], Bernstein [1976], Castleman [1979], and Green [1983].

For further reading on the general topic of image restoration, see Bates and McDonnell [1986], Stark [1987], Jain [1989], and Kak and Slaney [1988]. The latter reference deals almost exclusively with computerized tomographic imaging, a topic that although beyond our discussion, is of considerable interest in medical imaging.

PROBLEMS

5.1 Consider a linear, position invariant image degradation system with impulse response $h(x - \alpha, y - \beta) = e^{-[(x-\alpha)^2 + (y-\beta)^2]}$. Suppose that the input to the system is an image consisting of a line of infinitesimal width located at $x = a$, and modeled by $f(x, y) = \delta(x - a)$. Assuming no noise, what is the output image $g(x, y)$?

5.2 Show the validity of Eq. (5.2-8).

5.3 A professor of archeology doing research on currency exchange practices during the Roman empire recently became aware that four Roman coins crucial to his research are listed in the holdings of the British Museum in London. Unfortunately, he was told after arriving there that the coins recently had been stolen. Further research on his part revealed that the museum keeps photographs of every item for which it is responsible. Unfortunately, the photos of the coins in question are blurred to the point where the date and other small markings are not readable. The blurring was caused by the camera being out of focus when the pictures were taken. You are hired as a consultant to determine whether computer processing can be utilized to restore the images to the point where the professor can read the markings. You are told that the original camera used to take the photos is still available, as are other representative coins of the same era. Propose a step-by-step solution to this problem.

5.4 Derive an equation analogous to Eq. (5.4-13), but for arbitrary uniform velocity in both the x and y directions.

5.5 Consider the problem of image blurring caused by uniform acceleration in the x direction. If the image is at rest at time $t = 0$ and accelerates with a uniform acceleration $x_0(t) = at^2/2$ for a time T, find the transfer function $H(u, v)$.

5.6 A space probe is designed to transmit images from a planet as it approaches it for landing. During the last stages of landing, one of the control thrusters fails, resulting in rapid rotation of the craft about its vertical axis. The images sent during the last two seconds prior to landing are blurred as a consequence of this circular motion. The camera is located in the bottom of the probe, along its vertical axis, and pointing down. Fortunately, the rotation of the craft is also about its vertical axis, so the images are blurred by uniform rotational motion. In addition, during the acquisition time of each image the craft rotation was limited to $\pi/8$ radians. The image acquisition process can be modeled as an ideal shutter that is open only during the time the craft rotated the $\pi/8$ radians. You may assume that vertical motion was negligible during image acquisition. How would you use the concepts you have learned in this chapter to restore the images? You are not being asked to provide a specific solution. Rather, you are asked to provide a basic approach to the solution.

5.7 Provide a specific solution (in the form of equations) to Problem 5.6, listing any assumptions that you made in arriving at that solution.

5.8 **a)** Show how Eq. (5.5-8) follows from Eq. (5.5-7).

b) Show how Eq. (5.5-9) follows from Eq. (5.5-8).

5.9 Image blurring caused by long-term exposure to atmospheric turbulence can be modeled by the transfer function $H(u, v) = \exp[-(u^2 + v^2)/2\sigma^2]$. Assume negligible noise. What is the equation of the Wiener filter you would use to restore an image blurred by this type of degradation?

5.10 Assume that the model in Fig. 5.1 is linear and position-invariant and show that the power spectrum of the output is given by $|G(u, v)|^2 = |H(u, v)|^2|F(u, v)|^2 + |N(u, v)|^2$. Refer to Eq. (5.2-40).

5.11 Cannon [1974] suggested a restoration filter $R(u, v)$ satisfying the condition $|\hat{F}(u, v)|^2 = |R(u, v)|^2|G(u, v)|^2$ and based on the premise of forcing the power spectrum of the restored image, $|\hat{F}(u, v)|^2$, to equal the power spectrum of the original image, $|F(u, v)|^2$.

a) Find $R(u, v)$ in terms of $|F(u, v)|^2$, $|H(u, v)|^2$, and $|N(u, v)|^2$. (*Hint:* Refer to Fig. 5.1, Eq. (5.2-40), and Problem 5.10.)

b) Use your result in (a) to state a result in the form of Eq. (5.5-9).

5.12 Suppose that each element of an image is normalized to the range [0, 1]. Then it is possible to interpret each such element as the probability of a certain number of photons hitting that particular element location in the image. Entropy is defined as $E = -p \ln p$, where p is a probability, and ln is the natural logarithm (see Chapter 6). We define $E = -\mathbf{f}^T \ln \mathbf{f}$ to be the entropy of an image that has been normalized and has been expressed in vector form. In this notation, the vector $\ln \mathbf{f}$ is formed by taking the natural logarithm of each component of $\mathbf{f}$. A useful filter for addressing degradations based on a random grain model (similar to modeling film grain) is obtained by performing a constrained least squares minimization of the negative of the entropy. Show that the resulting restored image is given by the transcendental equation $\hat{\mathbf{f}} = \exp[-\mathbf{1} - 2\alpha\mathbf{H}^T(\mathbf{g} - \mathbf{H}\hat{\mathbf{f}})]$.

5.13 A linear approximation to the maximum entropy solution given in Problem 5.12 can be obtained by expanding the exponential in a Taylor series and then keeping only the linear part of the expansion. Show that this approach results in the constrained least squares formulation in Eq. (5.3-9) but with $\mathbf{Q} = \mathbf{I}$.

5.14 A certain x-ray imaging geometry produces a blurring degradation that can be modeled as the convolution of the sensed image with the spatial, circularly symmetric function $h(r) = [(r^2 - 2\sigma^2)/\sigma^4]\exp[-r^2/2\sigma^2]$, where $r^2 = x^2 + y^2$. Obtain the transfer function of a constrained least squares filter you could use to deblur the images produced by this x-ray system. You may assume that the images are square.

5.15 Start with Eq. (5.7-12) and derive Eq. (5.7-14).

5.16 Suppose that, instead of using quadrilaterals, you used triangular regions in Section 5.9 to establish a spatial transformation and gray-level interpolation. What would be the equations analogous to Eqs. (5.9-5), (5.9-6), and (5.9-7) for triangular regions?