Journal of Mathematical Imaging and Vision, 1, 257–278 (1992).

Hole-Spectrum Representation and Model-Based Optimal Morphological Restoration of Binary Images Degraded by Subtractive Noise

EDWARD R. DOUGHERTY
Center for Imaging Science, Rochester Institute of Technology, Rochester, NY 14623

ROBERT M. HARALICK
Department of Electrical Engineering, University of Washington, Seattle, WA 98105

Abstract. A shape-based image representation grounded on the distribution of holes within an image is developed, and the manner in which this representation can be used to design optimal morphological filters to restore images suffering from subtractive-noise degradation is investigated. The image and noise models are predicated on the existence of some class of shape primitives into which both image and noise can be decomposed (relative to union), and this decomposition is developed within the framework of a general algebraic paradigm for component-based filtering that does not depend on the linear-space structure typically used in spectral representations. Both deterministic and nondeterministic models are considered, and in each case the necessary model constraints are fully explored. Moreover, the type of filters that are naturally compatible with the image-noise models are analyzed. Specifically, optimal morphological filter design is studied in terms of the shape-based hole spectrum (as linear filter design is studied in terms of the frequency spectrum). Various forms of a design algorithm are discussed, the particulars depending on a symmetric-difference error analysis yielding approximate error expressions in terms of the spectral decomposition and the geometry of the underlying shape primitives. Finally, the statistical estimation procedures required for practical implementation of the entire spectrum-filter paradigm are explained.

Key Words: mathematical morphology, restoration, optimal filter, image representation, spectrum

1. Introduction

In the present paper we address the problem of finding an image representation applicable to restoration of binary images degraded by subtractive noise. Our goal is to provide a paradigm for model-based restoration that is compatible with the very general statistical characterization of optimal morphological filters given by Dougherty [1], [2]. As a general framework, the latter is applicable to image-processing problems concerned with estimation—for instance, restoration and compression. Although the optimal linear filter can best be used by relating it to frequency content, as is done by the Wiener filter, there remains to be found relevant image decompositions facilitating discovery of optimal morphological filters.

The problem we wish to address can be appreciated by briefly considering the Wiener filter and its relation to optimal linear filtering. Given a collection of observation random variables, the optimal filtering problem is to find an estimation rule (filter) involving the observation variables that best predicts the outcome of some other random variable relative to some measure of goodness. If the goodness measure is mean-square error, the optimal (best) filter is the conditional expectation of the variable to be estimated given the observation variables. However, in many situations the estimation rule is constrained to some class of functions, thereby constraining the optimization. In linear filtering the estimation rule must be a linear combination of the observations; in morphological filtering it must be a morphological function of the observations. In linear filtering, design of the optimal filter reduces to the solution of a set of linear equations; in morphological filtering, design involves a search over certain classes of structuring elements [1], [2]. If we consider restoring an image, then the observations lie in some pixel neighborhood about the pixel whose value is to be estimated, and in both the linear and morphological cases the general estimation problem is pixel dependent, so that the optimal filter is spatially variant and there is (possibly) a different estimation rule at each pixel. For linear filtering, wide-sense stationarity of the image-noise process yields a spatially invariant (pixel-independent) filter; for morphological filtering, strict-sense stationarity of the image-noise process yields a spatially invariant filter [1], [2]. In the linear case and under wide-sense stationarity, optimal-filter design can be simplified by transforming the problem into the frequency domain by means of the Fourier transform, thereby reducing design to an algebraic problem involving the spectra of the image and noise. In effect, the task is to weight the spectral components of the degraded image so as to produce, upon inverse transforming, the optimal linear estimate of the uncorrupted image. The salient point relative to the present paper is that spectral decomposition leads to componentwise, frequency-based optimal-linear-filter design. The general morphological spectral problem is to find image representations leading to component-based optimal-morphological-filter design, the utility being the facilitation of filter design (the Wiener filter facilitates linear-filter design under certain conditions). This requires appropriate modeling of the morphological problem, since here we do not have recourse to linear-space theory.

In the absence of linear-space methods, a key aspect of the current study is the presentation of a general algebraic formulation of component-pass-type filtering that serves to frame the model subsequently discussed. This formulation is, in essence, an algebraic abstraction of lowpass/highpass filtering (and has been presented in a partial report on the current investigation [3]).

Having found an appropriate spectral (component) decomposition, we proceed to analyze signal and noise relative to it, not only to design optimal filters operating in the spectral domain, but also to design optimized spatial-domain filters. For frequency decompositions, such spatial filtering is achieved by means of the convolution theorem for Fourier transforms; here we use shape-based decompositions grounded on hole distributions and derive efficient morphological filters from these.

We have used a similar spectral approach in a different framework [4], [5] by employing an opening decomposition to discover an optimal filter for a certain image-noise model. There, however, although the Wiener analogy holds quite well, the resulting filters do not fit the framework of [1]. Specifically, the resulting optimal filters are not necessarily monotonic. The present model fits the framework and leads to a class of filters that are closely related to the image and noise.

Before proceeding, we will mention two areas of current investigation that bear relation to the present work. First, morphological optimization has been studied by Schonfeld and Goutsias [6], [7]. There the setting is not as general as that discussed in [1], [2], and the optimization criteria are different. Nonetheless, a key aspect of [6], [7] is the manner in which the optimization is applied to alternating sequential filters (see Lougheed [8] or Serra [9]). Since these filters are increasing and translation invariant, they could be fit into the general binary paradigm of [1]. As for the second area of investigation, one more closely related to the present study, a good deal of effort has gone into making the general binary-filter-design paradigm of [1] more tractable; the problem is to mitigate the computational burden entailed in the search for optimal structuring elements. We mention two approaches used to date. Dougherty, Mathew, and Swarnakar [10] have developed an algorithm to derive the optimal binary mean-square morphological filter from the conditional expectation, and in cases for which the optimal morphological filter is near to being overall optimal, the algorithm works extremely fast. In addition, Loce and Dougherty [11]–[13] have facilitated filter design by using constraints that, in return for some slight loss in filter performance, eliminate vast numbers of structuring elements from consideration. A fundamental facet of our model-based approach to filter optimization is that it *ipso facto* greatly improves design tractability.

2. Morphological preliminaries

An image is a subset of the Cartesian grid $Z \times Z$, and we use two elementary morphological operations, erosion and dilation, defined by $S \ominus A = \{z : A + z < S\}$ and $S \oplus A = \cup\{A + x : x \in S\}$, respectively, where $A + z = \{a + z : a \in A\}$. We leave the basic properties of the these operations to the literature [14]–[18] (see also [19]–[22] for lattice-based approaches to mathematical morphology). More generally, a binary *morphological filter* is a set-to-set mapping Ψ that is *increasing* [$S < T$ implies $\Psi(S) < \Psi(T)$] and *translation invariant* [$\Psi(S + x) = \Psi(S) + x$]. The *kernel*, Ker[$\Psi$], is the class of all images S such that $\Psi(S)$ contains the origin. (Note that Serra [23], in the context of lattices, defines a morphological filter differently: the mapping must be increasing and *idempotent* [$\Psi\Psi = \Psi$]; translation invariance is not relevant in the abstract lattice setting he uses.)

A fundamental proposition of mathematical morphology is the Matheron representation [14]: every morphological filter can be expressed as a union of erosions by its kernel elements. As was noticed by Dougherty and Giardina [18], [24] and Maragos and Schafer [25], [26], if certain pathological cases are excluded, a morphological filter Ψ has a *basis*, Bas[Ψ], of structuring elements such that the Matheron expansion is taken over Bas[Ψ] instead of Ker[Ψ]: $\Psi(S) = \cup\{S \ominus E : E \in \text{Bas}[\Psi]\}$. The basis is a minimal class of structuring elements within the kernel: for any $E \in$ Ker[Ψ] there exists $E' \in$ Bas[Ψ] such that $E' < E$ and there does not exist a pair of structuring elements in Bas[Ψ] properly related by set inclusion. If **B** is the basis for a filter, we will often denote the filter by $\Psi_{\mathbf{B}}$. (For gray-scale and lattice extensions of the Matheron theorem, see [18], [22], [23], [25], [26].

In [1], mean-square-error (MSE) filter optimization is characterized by the Matheron representation. Consider N binary observation random variables $X[1], X[2], \ldots, X[N]$. Each realization of the random vector $X = (X[1], X[2], \ldots, X[N])$ is a 0-1 N-tuple. If we let 1 and 0 denote points of the domain that lie within or without the point set $\{1, 2, \ldots, N\}$, respectively, then each realization x of X constitutes a subset of $\{1, 2, \ldots, N\}$, and we can erode x by a deterministic structuring element $A = (a[1], a[2], \ldots, a[N])$, where $a[j]$ is 0 or 1. The erosion $x \ominus A$ is a binary functional, and its value is either 0 or 1. For a random vetor X and fixed structuring element A, erosion defines an estimator $X \ominus A$ that can be used to estimate another random variable Y. The optimal MS erosion filter is the one defined by the structuring element A minimizing $\text{MSE}\langle A\rangle = E[|Y - (X \ominus A)|^2]$.

Using the Matheron representation as a guide, [1] defines an *N-observation digital morphological filter* as a functional of the form $\Psi(x) = \max\{x \ominus A(i)\}_i$, where x and $A(i)$ are deterministic binary N-vectors. $\{A(i)\}$ is called the *basis* of Ψ. Extension of optimality to N-observation morphological filters involves minimizing $\text{MSE}\langle\Psi\rangle = E[|Y - \Psi(X)|^2]$ over all possible choices of N-observation morphological filters Ψ. Since Ψ is fully determined by its basis, finding the optimal N-observation filter reduces to selecting an optimal basis.

3. An algebraic paradigm for component filtering

We require a class **S** of signals, and with each signal f in **S** there must be an associated n-tuple $\Phi(f) = [f_k]$, called the *spectrum* of f, whose components come from some set **C** of objects (which need not be numbers). This association forms a mapping

$$\Phi : f \to \Phi(f) = [f_1, f_2, \ldots, f_n] \qquad (1)$$

from **S** into $\mathbf{C}^n$, and, since $[f_k]$ must serve as a representation of f, we demand that there exist an inverse mapping Φ^{-1}.

For any two signals f and g in **S**, there is defined a binary operation $\langle\rangle$, not necessarily commutative, such that $f\langle\rangle g$ need not lie in **S**. If $f\langle\rangle g$ lies in **S**, then we say g is *weakly compatible* with f. Weak compatibility takes the place of linear-space closure, but more is required: some stronger condition must serve in lieu of basis-coefficient linearity.

To proceed further, we assume there exists an operation $\otimes$ in **C**, and for any two elements in $\mathbf{C}^n$ we induce $\otimes$ componentwise by

$$[a_k] \otimes [b_k] = [a_k \otimes b_k]. \tag{2}$$

We say g is *strongly compatible* with f if it is weakly compatible and

$$\Phi(f\langle\rangle g) = [f_k] \otimes [g_k]. \tag{3}$$

$\mathbf{S}\langle f \rangle$ denotes the class of strongly f-compatible elements of **S**.

We now have sufficient structure to provide an algebraic characterization of component filtering. First suppose g is strongly f-compatible and that f and g do not share nonnull components, i.e., that the signal and noise components are separated. After rearranging the components, if necessary, there exists an index q such that $f_k = 0$ for $k > q$ and $g_k = 0$ for $k \leq q$ (note that we could have interchanged the ordering so that f_k is null for small k and g_k is null for large k). We define the function Q on $\mathbf{C}^n$ by

$$Q([a_1, a_2, \ldots, a_n]) = [a_1, a_2, \ldots a_q, 0, 0, \ldots, 0]. \tag{4}$$

By assuming $f\langle\rangle g$ to be a degradation of f, total restoration is achieved by the mapping $\Phi^{-1}Q\Phi$.

Rather than filter in the component domain, we might, as with convolution and frequency, desire a spatial-domain filter that accomplishes the same task as Q. Such a filter would be a mapping Ψ that completes the commutative diagram.

$$\begin{array}{ccc} & \Phi & \\ f\langle\rangle g & \rightarrow & [f_k \otimes g_k] \\ \Psi \downarrow & & \downarrow \ Q \\ f & \rightarrow & [f_k] \\ & \Phi & \end{array} \tag{5}$$

that is, $\Psi = \Phi^{-1}Q\Phi$. From a strictly algebraic perspective, Ψ always exists as $\Phi^{-1}Q\Phi$; however, what we really desire is that Ψ come from some class of filters. Specifically, we require Ψ to be a morphological filter of some given kind. Even with signal–noise separation, Ψ in the desired form might not be obtainable exactly, so we will have to make do with some approximate completion of the diagram.

More generally, the nonnull components of f and g are not disjoint. Thus, the mapping Q of (4) does not exist (as formulated). Instead it takes the form $Q = [q_k]$, and filtering is accomplished by $Q([a_k]) = [q_k(a_k)]$, where q_k is the kth component function of Q. This is analogous to the frequency form of the Wiener filter, in which case $q_k(a_k) = w_k a_k$, where w_k is the weight and a_k is the frequency component. To measure the goodness of Q, there must exist some *error measure* e between signals. If signals are assumed to be random functions, goodness of Q is measured by the expected value $E[e(f, \Phi^{-1}Q\Phi(f\langle\rangle g))]$.

4. The hole spectrum

Our goal is the development of a generalized spectral (canonical) representation that models holes in an image in a manner conducive to filling noise-created holes by a union of erosions. We begin by postulating a list of primitive shapes that will be assumed to generate both noise and generic holes. We consider a family of images, called *shape primitives*: $\mathbf{P} = \{N_1, N_2, \ldots, N_n\}$. Each N_j is assumed to be path connected. If S is any image, $\langle S \rangle$ denotes the minimal rectangle containing S. S is said to be **P**-*representable* if there exist points x_{jk} such that

(i)

$$\begin{aligned} S &= \langle S\rangle - \cup\{N_j + x_{jk} : j = 1, 2, \ldots, n; \\ &\qquad k = 1, 2, \ldots, s(j)\} \\ &= \langle S \rangle - \cup N_j \oplus X_j, \end{aligned}$$

where $X_j = \{x_{j1}, x_{j2}, \ldots, x_{j,s(j)}\}$.

(ii) If C is any connected component of $\langle S \rangle - S$, then there exists a unique pair (j, k) in the index set of the union such that $C = N_j + x_{jk}$. Moreover, for any other

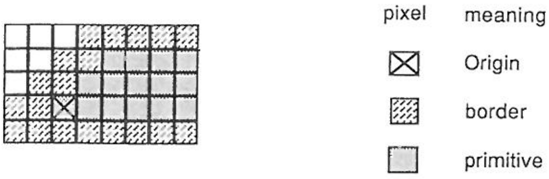


Fig. 1.

index pair (r, s), $N_r + x_{rs}$ is disjoint from C.

Condition (ii) imposes a minimality condition on the union in (i), thereby making the representation of (i) unique. **S** is the class of **P**-representable images, and henceforth we assume all images to be **P**-representable. A *hole* is any connected component of $\langle S \rangle - S$. It need not be enclosed by pixels in S; rather, it may be a penetration of the minimal containing rectangle $\langle S \rangle$.

We postulate the existence of a subclass of **P**, $\mathbf{N} = \{N_1, N_2, \ldots, N_p\}$, $1 \leq p \leq n$. The purpose of **N** is to be a subfamily of shape primitives that serve as noise generators; specifically, we limit the noise holes of S to be those holes in $\langle S \rangle$ created by terms in (i) of the form $N_j + x_{jk}$, where N_j lies in **N**. It might be that there is no limitation and $p = n$. We call **N** the *noise class*. Note that *generic* holes, those belonging to S properly and not having been created by noise, can arise from shapes in **N**.

For each j, we assume there exists some *enclsoure* set M_j containing N_j. The purpose of M_j is to form a *border* $M_j - N_j$ (see figure 1), so that the filling of a hole created by N_j can be examined in terms of erosions applied to the border. The extent to which M_j is greater than N_j is a modeling question having to do with the nature of generic holes and constraints on the noise.

In regard to representation (i), the *hole spectrum* of a **P**-representable image S is defined by

$$H[S] = [X_1, X_2, \ldots, X_n]. \quad (6)$$

$H[S]$ and $\langle S \rangle$ uniquely characterize S. The n-tuple

$$h[S] = [s(1), s(2), \ldots, s(n)] \quad (7)$$

is called the *hole-amplitude spectrum* (HAS) of S. It gives the number of times each shape primitive N_j is translated to form the representation (i). The set $\langle S \rangle - S$ consists of "hole-like" components, is given by the union representation in (i), and is characterized by the hole spectrum.

Our filtering concern is with the image S corrupted by subtractive noise, i.e., new holes are created by the noise. If all shape primitives are assumed to lie in **P**, a noise image N is postulated to be of the form $N = \langle S \rangle - N^*$, where

(iii) $N^* = \cup\{N_j + z_{jk} : j = 1, 2, \ldots, p;\ k = 1, 2, \ldots, t(j)\} = \cup N_j \oplus Z_j$, where $Z_j = \{z_{j1}, z_{j2}, \ldots, z_{j,t(j)}\}$.

The noise-corrupted image is defined to be $S \cap N$. Since

$$S \cap N = S - N^*, \quad (8)$$

the noisy image results from subtracting translated noise patterns from the uncorrupted image.

We wish to view $S \cap N$ as the noisy version of S; however, we are faced with a serious hurdle: we must be assured that $S \cap N$ is **P**-representable. Thus we must put constraints on the union forming N^*. Put crudely, we must be careful that the intersection does not create holes that are not fit tightly by single translates of the N_j, a condition required for **P**-representation. The problem here is the lack of a vector-space closurelike property. A practical way around the difficulty is to define a noise image of the type in (iii) to be *weakly compatible* with S if $\langle S \cap N \rangle = \langle S \rangle$ and $S \cap N$ is **P**-representable and to limit our formal analysis to weakly S-compatible noise images. Relative to the algebraic paradigm of the preceding section, intersection corresponds to the abstract operation $\langle\,\rangle$, with **S** being the **P**-representable images. The purpose of the requirement $\langle S \cap N \rangle = \langle S \rangle$, which says that the minimal enclsoing rectangle is not diminished, is to avoid reformulation of the representation (i), which might be required to maintain the uniqueness of the hole-spectrum representation if $\langle S \cap N \rangle$ were a proper subset of $\langle S \rangle$.

Even if N is weakly S-compatible, there is another potential problem with our representation theory. In subtracting N^* from S we may create *interaction holes*, which are created by enlarging generic holes during noise intersection. To appreciate the undesirability of such holes, consider the addition of a signal with noise in the frequency setting. Owing to the linearity of the basis coefficients, the spectrum of the sum is obtained by adding basis coefficients componentwise. We would like to have an analogous process with regard to hole spectra: coefficients for $S \cap N$ should be formed by unioning signal and nosie coefficients. Relative to the algebraic paradigm, we would like (3) to hold with intersection in place of $\langle\rangle$ and union in place of $\otimes$. The following conditions on the noise will guarantee both weak compatibility and the type of noisy-image spectra we desire:

(iv) For each N_j in $\mathbf{N}$, the border $M_j - N_j$ is strongly path connected.

(v) If $N_j + z_{jk}$ and $N_r + z_{rs}$ are translates in the union forming N^*, then

$$(N_r + z_{rs}) \cap (M_j + z_{jk}) = \emptyset.$$

(vi) For each translate $N_j + z_{jk}$ forming N^*, $M_j + z_{jk} < S$.

Condition (iv) makes the borders sufficiently impenetrable so that conditions (v) and (vi) accomplish their desired ends. Condition (v) is a separation condition assuring that the $\mathbf{P}$-representability of $S \cap N$ is not destroyed by overlapping or touching noise holes. Condition (vi) says that we ignore noise-primitive translates that miss S altogether (certainly realistic) and also that noise enclosures cannot hit the boundary of S, which could destroy the $\mathbf{P}$-representability of $S \cap N$ or create interaction holes. Note another key consequence of the noise conditions: no translate $N_r + z_{rs}$ forming N can "get lost" as a subset of some translate $N_j + x_{jk}$ in (i), or vice versa. If a noise image of the form given in (iii) satisfies conditions (iv), (v), and (vi), it is said to *conform* to S. Under these conditions, the hole spectrum of the noise is

$$H[N] = [Z_1, Z_2, \ldots, Z_n]. \tag{9}$$

Consider the hole spectrum of an image $S \cap N$ obtained from S by comformable intersection noise. With N^* given by (iii), we need a representation of the generic holes. By (i), these holes comprise $\langle S \rangle - S$, which is of the form

$$\langle S \rangle - S = \cup N_j \oplus X_j. \tag{10}$$

If we assume that S is $\mathbf{P}$-representable and that N conforms to S,

$$\begin{aligned} S \cap N &= [\langle S \rangle - (\langle S \rangle - S)] - N^* \\ &= \langle S \rangle - \cup N_j \oplus X_j - \cup N_j \oplus Z_j \\ &= \langle S \rangle - \cup N_j \oplus (X_j \cup Z_j). \end{aligned} \tag{11}$$

Its hole spectrum is

$$H[S \cap N] = [X_1 \cup Z_1, X_2 \cup Z_2, \ldots, X_n \cup Z_n]. \tag{12}$$

Relative to (3), (12) can be rewritten as

$$H[S \cap N] = H[S] \cup H[N], \tag{13}$$

so that N is strongly compatible with S. In summary, if N conforms to S, then N is strongly compatible with S and we can proceed with component filtering. Of great practical importance is that the hole-amplitude spectrum components are combined by addition:

$$\begin{aligned} h[S \cap N] = {} & [s(1) + t(1), s(2) + t(2), \ldots, \\ & s(n) + t(n)]. \end{aligned} \tag{14}$$

Let us now consider the spectral approach to noise filtering. Suppose there exists a *separation index* q such that $s(j) = 0$ for $j \leq q$ and $t(j) = 0$ for $j > q$, so that the signal is separated from the noise and the relevant spectra take the forms

$$\begin{aligned} H[S] &= [\emptyset, \emptyset, \ldots, \emptyset, X_{q+1}, X_{q+2}, \ldots, X_n], \\ H[N] &= [Z_1, Z_2, \ldots, Z_q, \emptyset, \emptyset, \ldots, \emptyset]. \end{aligned} \tag{15}$$

In the spectrum domain the noise is fully filtered by the function

$$\begin{aligned} Q([Y_1, Y_2, \ldots, Y_n]) = {} & [\emptyset, \emptyset, \ldots, \emptyset, Y_{q+1}, \\ & Y_{q+2}, \ldots, Y_n] \end{aligned} \tag{16}$$

since $Q(H[S \cap N]) = H[S]$.

When signal and noise are not separated, we desire a function Q that best restores the hole spectrum of S from $S \cap N$, where bestness is defined relative to some criterion of goodness.

Since our desire is restoration in the spatial domain, a good measure of error is the symmetric difference

$$e[Q] = c[H^{-1}[Q(H[S \cap N])] - S] + c[S - H^{-1}[Q(H[S \cap N])]], \quad (17)$$

where c denotes the number of pixels in a set. This approach is well defined because we have assumed an image is reconstructable from its hole spectrum, thereby making H^{-1} well defined.

5. Spatial-domain filters

In analogy to the frequency-spectrum setting, we would like some spatial-domain filter that can, to one degree or another (and in a manner analogous to convolution), accomplish the task of Q. In the present setting we desire a morphological filter Ψ such that $\Psi(S \cap N) = S$. The problem is to find Ψ so as to complete the following commutative diagram:

$$\begin{array}{ccc} & H & \\ S \cap N & \rightarrow & H[S \cap N] \\ \Psi \downarrow & & \downarrow Q \\ S & \rightarrow & H[S] \\ & H & \end{array} , \quad (18)$$

which corresponds to diagram (5). Whether we are in the separated or nonseparated case, no filter Ψ can be expected to complete the diagram exactly, so we use the measure of goodness

$$e[\Psi] = c[S - \Psi(S \cap N)] + c[\Psi(S \cap N) - S], \quad (19)$$

which is simply a reformulation of the spectral-restoration error $e[Q]$ of (17). Since we cannot expect to complete the diagram exactly, we should not expect $e[\Psi]$ to equal $e[Q]$, although ideally they would agree.

The most general problem, whether or not the noise and signal are separated by some index, is to seek Ψ among all morphological filters; however, on the basis of our model, we will seek Ψ from a subclass of filters. We proceed to outline the model-based paradigm suitable to observed images $S \cap N$.

Although we could search for a suitable filter Ψ among all possible morphological filters, this would mitigate the purpose of constructing an appropriate noise model. Thus, along with the class $\mathbf{P}$ of shape primitives, we postulate the existence of a class $\mathbf{E} = \{E_1, E_2, \ldots, E_m\}$ of structuring elements. Since these elements must restore images suffering subtractive-noise degradation, to be of use they must satisfy some conditions that make them suitable for the task at hand. Hence we require the elements of $\mathbf{E}$ to satisfy the following two conditions:

(vii) For any N_j lying in $\mathbf{N}$,

$$\cup\{(M_j - N_j) \ominus E_i : i = 1, 2, \ldots, m\} > N_j.$$

(viii) For $i = 1, 2, \ldots, m$, E_i does not contain the origin and there exist activated pixels p_1 and p_2 in E_i such that the origin lies between p_1 and p_2.

Condition (vii) guarantees that holes created by subtracting noise primitives can be filled by eroding by elements of $\mathbf{E}$, and condition (viii) assures that eroding S or a subset of S by an element of $\mathbf{E}$ yields an output that lies within $\langle S \rangle$. This latter condition is reasonable since $S \cap N < S$.

For any image S and family $\mathbf{B}$ of structuring elements, we define the morphological filter $\Psi_{\mathbf{B}}(S)$ by

$$\Psi_{\mathbf{B}}(S) = S \cup [\cup\{S \ominus E_k : E_k \in \mathbf{B}\}]. \quad (20)$$

Since S is equal to S eroded by the origin, $\Psi_{\mathbf{B}}$ is a union of erosions, and as such it is a morphological filter. There is no basis minimality condition imposed on $\mathbf{E}$; that is, it may be that there is an element of $\mathbf{E}$ that is a proper subset of another element of $\mathbf{E}$. Thus if $\mathbf{B}$ is selected from $\mathbf{E}$, it may not be a basis; if not, simply eliminate redundant structuring elements. Define an image S to be $\mathbf{B}$-*closed* if $\mathbf{B}$ is a family of structuring elements for which $\Psi_{\mathbf{B}}(S) = S$. Whether S is $\mathbf{B}$-closed or not, we refer to $\Psi_{\mathbf{B}}(S)$ as the $\mathbf{B}$-*closure* of S. $\mathbf{E}$-closed images are of special interest because if $S \cap N$ is a noisy image, $\mathbf{B}$ is a subclass of $\mathbf{E}$ and we apply $\Psi_{\mathbf{B}}$, then $\Psi_{\mathbf{B}}(S \cap N) < S$, so that $\Psi_{\mathbf{B}}$ cannot create error pixels outside of S so long as S is $\mathbf{E}$-closed.

Note that by (viii) the **E**-closure of S is a subset of $\langle S \rangle$, so that $S < \Psi_{\mathbf{E}}(S) < \langle S \rangle$. The **B**-closure of S need not be **B**-closed: $\Psi_{\mathbf{B}}(S)$ is not necessarily equal to $\Psi_{\mathbf{B}}(\Psi_{\mathbf{B}}(S))$; that is, $\Psi_{\mathbf{B}}$ is not generally idempotent. Nonetheless, for any integral power k, $\Psi_{\mathbf{B}}^k(S) < \langle S \rangle$. If we construct **E** in a manner compatible with filling holes created by noise primitives, filters of the form $\Psi_{\mathbf{B}}$, $\mathbf{B} < \mathbf{E}$, comprise a potentially fertile class.

If we use the symmetric-difference restoration error, then error is basis dependent and is given by

$$e[\mathbf{B}] = c[S - \Psi_{\mathbf{B}}(S \cap N)] + c[\Psi_{\mathbf{B}}(S \cap N) - S]. \quad (21)$$

Our problem is combinatoric in that we must search for the best among all subsets of **E** that are filter bases. As will be subsequently discussed, such a search fits neatly into the general filter methodology developed in [1], [2]. If **E** has been selected in a manner highly compatible with filling holes created by primitives in **N**, then we have accomplished a good deal because the general search problem of [1] is dramatically reduced.

6. The nondeterministic setting

In moving to the nondeterministic setting, we need to reinterpret the foundations of the hole-spectrum theory in a probabilistic light. Regarding the signal and noise assumptions, $S \cap N$ is a random images formed from intersecting two random image, S and N. To apply our theory we make the modeling assumption that for each realization of $S \cap N$ the realizatioin of N conforms to the corresponding realization of S. In the random model the hole spectrum is an n-tuple of random sets, and the HAS is a random n-vector. For a given basis **B** we consider the expected value of the error (21), namely, $E[e(\mathbf{B})]$. Optimality is achieved by finding **B** that minimizes $E[e(\mathbf{B})]$. Note that $E[e(\mathbf{B})]$ corresponds to $E[e(\Psi)]$, where $e(\Psi)$ is given in (19), and hence to $E[e(Q)]$, where $e(Q)$ is given in (17). The latter point is important, since, just as with linear frequency-component optimization, there is a relation between $E[e(Q)]$ and $E[e(\Psi)]$. Ideally, if the commuting diagram (18) were completed, then we would have $E[e(Q)] = E[e(\Psi)]$.

We now consider the manner in which our theory relates to the general MSE optimization theory developed in [1]. Because that theory treats window observations as a collection of random variables to estimate another random variable, the "true" value of the image at a given pixel, it yields morphological filters given by erosion expansions for which the basis elements are pixel dependent; that is, the optimal filter is spatially variant. As discussed in [1], the filter becomes spatially invariant if the image-noise process is stationary.

To proceed with a comparison of the hole-spectrum approach and the general MSE approach of [1], we suppose $S \cap N$ to be stationary as a random process. Applying the methodology of [1] to our noise model, at an arbitrary pixel x mean-square error is

$$\text{MSE}[x] = E[|\Psi(S \cap N)(x) - S(x)|^2], \quad (22)$$

where $S(x) = 1$ if x lies in S and $S(x) = 0$ otherwise. If we let $P[|]$ denote conditional probability, MSE can be rewritten as

$$\begin{aligned} \text{MSE}[x] &= P[\Psi(S \cap N)(x) = 1 \,|\, S(x) = 0] \\ &\quad P[S(x) = 0] \\ &\quad + P[\Psi(S \cap N)(x) = 0 \,|\, S(x) = 1] \\ &\quad P[S(x) = 1] \\ &= P[x \in \Psi(S \cap N) - S] \\ &\quad + P[x \in S - \Psi(S \cap N)]. \quad (23) \end{aligned}$$

If we let A denote the number of pixels in the image, by stationarity we obtain $A(\text{MSE}[x]) = E[e(\Psi)]$. Thus minimization of $E[e(\Psi)]$ is equivalent to MSE minimization.

Because we do not require stationarity to obtain a spatially invariant filter, herein we will always obtain a global basis. Moreover, whereas in [1] the restriction was that the structuring elements must lie in some observation window about x (in analogy to a windowed convolution), here we restrict our attention to a set of model elements. Such constraint is discussed in [1] relative to the general statistical approach. It should be noted that although we have dropped a local-estimation approach common to basic statistical estimation, thereby circumventing stationarity questions, we have paid a price. First, we have given up hope of obtaining general probabilistic distributional characterizations. Second, we have placed strict constraints on the signal, the noise, and their interaction. Third, we have restricted our attention to filtering a certain type of noise, rather than presenting a theory that applies equally well to any type of noise filtering, compression, or general restoration. Our gain, however, is that when the model is (sufficiently) appropriate we have a filter-design methodology that is intuitive, avoids the taxing statistical analysis required in the more general approach, is not sensitive to deviations from stationarity, and is computationally more efficient than the general approach in [1].

7. The fill matrix

A principle motivation for component representation by the hole spectrum is the development of fast algorithms to find statistically efficient morphological filters. Relative to the hole spectrum, our technique is to find "good" structuring elements in **E**. When a structuring element E_i is used to help fill a noise hole created by a noise primitive N_j, it accomplishes its end because $[(M_j - N_j) + z_{jk}] \ominus E_i$ contains pixels of $S - (N_j + z_{jk})$. If S is $\{E_i\}$-closed, then there is no overfilling; otherwise, $[S - (N_j + z_{jk})] \ominus E_i$ quite possibly contains pixels not originally in S. Thus if S is **E**-closed, no overfilling can arise; typically, however, S will not be **E**-closed. To help analyze overfill errors we introduce a fill measure for structuring elements.

For $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$, define the *fill ratio* r_{ij} by

$$r_{ij} = c[N_j]^{-1}c[((M_j - N_j) \ominus E_i) \cap N_j]. \quad (24)$$

These ratios form the *fill matrix* $R = [r_{ij}]$ with respect to **P** and **E**. The rows correspond to $E_1, E_2, \ldots, E_m$ and the columns to $N_1, N_2, \ldots, N_n$. Each r_{ij} gives the proportion of a "hole" in the enclosure M_j created by N_j that is filled when eroding by E_i.

Consider a noise-corrupted image $T = S \cap N$, and suppose there exists a separation index q. Then, relative to T the fill matrix takes the form $R_T = [N_T : G_T]$. There are q columns in the submatrix N_T corresponding to the shape primitives contributing to N, and there are $n - q$ columns in the submatrix G_T corresponding to shape primitives contributing generic holes. If $p = q$, then N_T corresponds exactly to **N**, whereas G_T corresponds to **P** − **N**. Appreciation of the fill matrix is straightforward in the separated case. If $j \leq q$ and $r_{ij} > r_{kj}$, then more of a noise hole created by N_j is filled by E_i than by E_k; thus if all other factors are ignored, E_i is a better contributor to a filter basis than is E_k. If $j > q$ and $r_{ij} > r_{kj}$, then more of a generic hole created by N_j is filled by E_i than by E_k; thus if all other factors are ignored, E_k is a better contributor to a filter basis than is E_i.

8. Filter design in the E-closed case – signal and noise separated

In developing algorithms we first consider the separated case under the assumption that S is **E**-closed. Since there is no overfilling, the fill matrix takes the block form $R_S = [N_S : 0]$, and total restoration of S results from the filter $\Psi_{\mathbf{E}}$. Rather than stop here, we desire a convenient method for using fewer than all of the structuring elements of **E** in a fully restoring basis.

We proceed in the following manner. Find a subset $E[1]$ of **E** such that

$$\cup\{(M_1 - N_1) \ominus E_i : E_i \in E[1]\} > N_1. \quad (25)$$

If we let $N[j]$ denote the union of the N_j translates in the union forming N^*, then under the model conditions

$$\Psi_{E[1]}(S - N[1]) = S. \quad (26)$$

We next extend $E[1]$ to $E[2]$ by adjoining elements so that

$$\cup\{(M_2 - N_2) \ominus E_i : E_i \in E[2]\} > N_2. \quad (27)$$

Under the model conditions

$$\Psi_{E[2]}(S - (N[1] \cup N[2])) = S. \quad (28)$$

Note that it is certainly possible that $E[2] = E[1]$. We proceed inductively to form an increasing set of structuring-element classes $E[1] < E[2] < \cdots < E[n]$ such that $\Psi_{E[n]}(S - N) = S$ for total restoration.

Because there is no assumption that the structuring elements of **E** are not redundant (and, indeed, we do not want to make such an assumption), there can exist E_i and $E_{i'}$ in **E** such that E_i is a proper subset of $E_{i'}$. Hence $E[n]$ might not form a legitimate filter basis. This problem can be corrected at each step of the algorithm by making certain that $E[k]$, $k = 1, 2, \ldots, n$, is nonredundant, or it can be corrected at the end by eliminating redundancy from $E[n]$ to form a basis **B**, and the resulting filter is $\Psi_{\mathbf{B}}$.

A second point concerns the size of $E[n]$ or of the resulting basis **B**. Because there is no "looking back" in the algorithm as posed, there is no guarantee that the resulting basis will contain the smallest possible number of elements. Since, for the present, a central aim is to avoid combinatorics, we will content ourselves with a stepwise algorithm.

9. Filter design when the image is not E-closed – signal and noise separated

Let us now drop the assumption that S is **E**-closed. No longer can we be assured of finding an ideal filter, one that gives full restoration. The data of the fill matrix facilitate a detailed study of the problem. Consider a subclass (potential basis) **B** from **E**. Let

$$\begin{aligned} e_j &= c[N_j - \Psi_{\mathbf{B}}(M_j - N_j)] \\ &= c[N_j] - c[\Psi_{\mathbf{B}}(M_j - N_j) \cap N_j] \quad (29) \end{aligned}$$

for $j \leq q$. Then e_j, to be called a *noise-error term*, counts the number of pixels in N_j that are erroneously not filled by $\Psi_{\mathbf{B}}$. For $j > q$, define the *generic-error term*

$$e_j = c[\Psi_{\mathbf{B}}(M_j - N_j) \cap N_j], \quad (30)$$

which counts the number of pixels in N_j erroneously filled. If we assume the enclosures M_j are sufficiently large, so that filling the hole created by $N_j + z_{jk}$ is accomplished only by structuring elements fitting in $(M_j - N_j) + z_{jk}$, then

$$\begin{aligned} e[\mathbf{B}] &= h(1)e_1 + \cdots + h(q)e_q \\ &\quad + h(q+1)e_{q+1} + \cdots + h(n)e_n \\ &= e_N[\mathbf{B}] + e_G[\mathbf{B}], \quad (31) \end{aligned}$$

where $h[S \cap N] = [h(1), \ldots, h(n)]$ is the hole-amplitude spectrum of $S \cap N$. A key point is that the e_j error terms depend only on the basis, the shape primitives, and the enclosures. The coefficients in the expansion are the HAS components.

In the **E**-closed case, $e_j = 0$ for $j > q$. When S is not **B**-closed, there is a price to pay when constructing the basis **B**: some or all elements E_i create overfilling that results in some of the e_j, $j > q$, error terms being nonzero. Because the problem is finite, we could proceed combinatorically to minimize $e[\mathbf{B}]$ exactly; however, our goal is a satisfactory stepwise algorithm.

A conflict regarding the enclosures occurs in minimizing $e[\mathbf{B}]$. The error term e_j depends not only on N_j but also on the enclosure M_j. If the enclosure hypothesis stated before (31) is not satisfied, then occasionally structuring elements in **B** can fill pixels in an N_j-created hole by fitting in the larger set $S \cap N$, whereas they do not fit in the translated copy of $M_j - N_j$. In such instances the actual error terms may differ: those for noise holes are less and those for generic holes are greater. Thus the spectral equation for $e[\mathbf{B}]$, (31), is only an approximation. Accepting this, let us assume that no pixel in a hole is filled by two structuring elements. If the elements of **E** are selected in such a manner that each corresponds to a given noise pattern and is not adept at fixing other noise patterns, then the assumption is appropriate. Practically, the assumption imposes a design constraint on **E**.

Given the preceding codicil, we can find useful operational formulae. First, consider a generic-error term e_j, $j > q$. For notational ease assume that **B** consists of $E_1, E_2, \ldots, E_r$. Then

$$\begin{aligned} e_j &= \sum_{i=1}^{r} c[((M_j - N_j) \ominus E_i) \cap N_j] \\ &= \sum_{i=1}^{r} r_{ij} c[N_j]. \quad (32) \end{aligned}$$

Thus

$$e_G[\mathbf{B}] = \sum_{j=q+1}^{n} h(j) \sum_{i=1}^{r} r_{ij} c[N_j]$$

$$= \sum_{j=q+1}^{n} h(j) r_{\mathbf{B}}(j) c[N_j], \quad (33)$$

where we have let $r_{\mathbf{B}}(j)$ denote the sum down the jth column of the fill matrix over the r_{ij} corresponding to structuring elements in **B** that are in the jth column. Next, consider a noise-error term e_j, $j \leq q$:

$$\begin{aligned} e_j &= c[N_j] - \sum_{i=1}^{r} c[((M_j - N_j) \ominus E_i) \cap N_j] \\ &= c[N_j] \left(1 - \sum_{i=1}^{r} r_{ij}\right). \quad (34) \end{aligned}$$

Thus

$$\begin{aligned} e_N[\mathbf{B}] &= \sum_{j=1}^{q} h(j) c[N_j] \left(1 - \sum_{j=1}^{r} r_{ij}\right) \\ &= \sum_{j=1}^{q} h(j) c[N_j] (1 - r_{\mathbf{B}}(j)). \quad (35) \end{aligned}$$

In sum, the error accruing because of overfilling and underfilling by the basis **B** is expressed in terms of the HAS, elements of the fill matrix, and the areas of the shape primitives.

The preceding expressions for $e_G(\mathbf{B})$ and $e_N(\mathbf{B})$ have been obtained under a strong non-interactive filling assumption. If we propose to use them in an approximate sense, we must be careful. On the basis of the noninteractive assumption, $0 \leq r_{\mathbf{B}}(j) \leq 1$ for all j. But if more than a single structuring element can fill pixels in a hole, then the column sums $r_{\mathbf{B}}(j)$ overestimate the filling by $\Psi_{\mathbf{B}}$. Thus we get an overestimate of the error $e_G(\mathbf{B})$ and an underestimate of the error $e_N(\mathbf{B})$. Keeping this in mind, we can use the error-estimate formulae for filter design; their strong advantages are their simplicity and their close relation to the HAS.

We demonstrate application of the formulae under the assumption that all noise holes are fully filled; this assumption is apt if the image does not possess too many generic holes and those that it does possess are large in comparison to the noise holes (and structuring elements in **E**). To proceed we simply eliminate $e_N[\mathbf{B}]$ by using a basis-finding algorithm that fills all noise holes. Thus our optimality is relative to only allowing overfill errors, and total error is simply $e[\mathbf{B}] = e_G[\mathbf{B}]$. Note that $e_G[\mathbf{B}]$ is bounded in terms of the **B**-closure of S, and more generally by the **E**-closure of S, namely,

$$e_G[\mathbf{B}] \leq c[\Psi_{\mathbf{B}}(S) - S] \leq c[\Psi_{\mathbf{E}}(S) - S], \quad (36)$$

thus quantifying our comment regarding the relative sizes of noise and generic holes. As we have seen previously, if S is **E**-closed, then restoration is complete.

We describe a stepwise algorithm. In all steps, $e_G[\mathbf{B}]$ is the approximation of (33). Select a subset $E[1]$ of **N** so that inclusion relation (25) is satisfied. Do this with the constraint that $e_G(E[1])$ is minimized. Next, adjoin enough elements to $E[1]$, thereby forming $E[2]$, so that inclusion relation (27) is satisfied and $e_G(E[2])$ is minimized. Continue in a stepwise fashion until $E[n]$ is obtained. Redundancy elimination to obtain a basis **B** can be treated analogously to the **E**-closed case. The stepwise form of the algorithm with no looking back may not produce the best filter, but the savings in design time over a combinatoric approach can be significant.

10. Filter design in the nondeterministic setting

Thus far we have considered filter design only in the deterministic setting, a modeling assumption that is often not realistic. In the nondeterministic case, we will assume S is a random image possessing L realizations $S_1, S_2, \ldots, S_L$, where the probability of S_k is p_k and $h[S_k] = [h_{Sk}(j)]$ is the HAS of S_k. S has the random-vector HAS $h[S] = [h_S(j)]$. Note that

$$E[h_S(j)] = \sum_{k=1}^{L} p_k h_{Sk}(j). \quad (37)$$

The error decomposition (31) now takes the expected-value form

$$E[e(\mathbf{B})] = E[e_N(\mathbf{B})] + E[e_G(\mathbf{B})]. \quad (38)$$

Since e_G depends on the signal realization, (33) needs to be reinterpreted in this light. Let $e_G[\mathbf{B}; S_k]$ denote the overfill error relative to basis $\mathbf{B}$ and the signal realization S_k. Equation (33) applies directly to $e_G[\mathbf{B}; S_k]$. Thus

$$\begin{aligned} E[e_G(\mathbf{B})] &= \sum_{k=1}^{L} p_k e_G[\mathbf{B}; S_k] \\ &= \sum_{k=1}^{L} p_k \sum_{j=q+1}^{n} h_{Sk}(j) r_{\mathbf{B}}(j) c[N_j] \\ &= \sum_{j=q+1}^{n} r_{\mathbf{B}}(j) c[N_j] \sum_{k=1}^{L} p_k h_{Sk}(j) \\ &= \sum_{j=q+1}^{n} E[h_S(j)] r_{\mathbf{B}}(j) c[N_j]. \quad (39) \end{aligned}$$

With N random, analysis of the underfilling error e_N is a bit more subtle. Even if the physical noise process is independent of the signal process, the manner in which we have defined the canonical representation depends on the noise realization conforming to the signal realization. Thus, probabilistically, the noise process is not independent of the signal. To that end, let $N_k = N|S_k$, the noise process given S_k, and let $h[N_k] = [h_{Nk}(j)]$ denote the HAS of N_k. Because the noise is random, its HAS $h[N] = [h_N(j)]$ is a random vector. Moreover,

$$E[h_N(j)] = \sum_{k=1}^{L} p_k E[h_{Nk}(j)]. \quad (40)$$

Let $e_N[\mathbf{B}; S_k]$ denote the underfill error relative to basis $\mathbf{B}$ and the signal realization S_k. With $E[h_{Nk}(j)]$ in place of $h(j)$, (35) applies to $e_N[\mathbf{B}; S_k]$, the difference being that it now gives $E[e_N(\mathbf{B}; S_k)]$. Proceeding similarly to the $E[e_G(\mathbf{B})]$ case, we obtain

$$E[e_N(\mathbf{B})] = \sum_{j=1}^{q} E[h_N(j)] c[N_j](1 - r_{\mathbf{B}}(j)). \quad (41)$$

There are a number of ways to use the fill matrix in the nondeterministic setting. Two obvious options are filling all noise holes and using the fill-matrix data to fill or not to fill a noise hole. Other options involve the amount of information used and the degree of looking back. At this point we give an algorithm that makes a fill decision but, to achieve tractability, ignores interaction and does not look back.

Begin by considering the first column of the fill matrix, corresponding to the noise primitive N_1. Read down the column, and find a subset $E[1]$ of $\mathbf{E}$ that fills N_1-created holes, while at the same time minimizing $E[e_G(E[1])]$; however, delete from $E[1]$ any structuring element E_i for which the expected underfill error resulting from not using E_i does not exceed the expected overfill error resulting from using E_i. Having found $E[1]$, proceed to the second column and adjoin enough structuring elements to $E[1]$ to form a set $E[2]$ that fills N_2-created holes, while minimizing $E[e_G(E[2] - E[1])]$; however, delete from $E[2]$ any E_i in $E[2] - E[1]$ for which the expected underfill error (relative to N_2) resulting from not using E_i does not exceed the expected overfill error from using E_i. Proceed inductively: having found $E[j]$, $j < q$, form the set $E[j+1]$ by adjoining enough structuring elements to $E[j]$ so that $E[j+1]$ fills N_{j+1}-created holes, at the same time minimizing

$$E[e_G(E[j+1] - (E[1] \cup E[2] \cup \cdots \cup E[j]))]. \quad (42)$$

Delete from $E[j+1]$ any newly adjoined E_i for which the expected underfill error (relative to N_{j+1}) resulting from not using E_i does not exceed the expected overfill error resulting from using E_i. If necessary, redundancy can be eliminated from the terminal class $E[n]$ to obtain a basis $\mathbf{B}$.

11. Nonseparated signal and noise

The algebraic filtering paradigm represented by (4) and the commuting diagram (5) depend on signal and noise separation. As noted subsequently in section 3, in the nonseparated situation the filter Q takes the form $Q[(a_k)] = [q_k(a_k)]$, which is analogous to the frequency form of the Wiener filter. In [4], [5] we discuss a binary approach to the Wiener filter, which, put simply, results in binary filter weights, so that a frequency is either left in or removed. Thus if the frequencies are reordered, one can view the filter as lowpass: if there are q nonzero weights, reorder the orthonormal basis with the frequencies corresponding to these nonzero weights listed first. Once this is done, the binary Wiener filter takes the form of (4), with $a_k = 1$ for $k = 1, 2, \ldots, q$.

In [4], [5] we show that under a general filter paradigm optimal binary filtering can be viewed in the manner just described for the binary Wiener filter. To adapt the Weiner paradigm to the abstract setting of (2) and (3) we need to assume there exists a second operation Δ in **C** and that with respect to Δ there exists a null element 0 such that $0 \,\Delta\, a = 0$ and a unit element 1 such that $1 \,\Delta\, a = a$. Interpreted with reference to (3) and (4), where $f\langle\rangle g$ is the degraded image, we desire a filter Q given componentwise by $Q([a_k]) = [q_k \,\Delta\, a_k]$, where q_k is either 0 or 1. If we make the supposition that the signal error e satisfies the relation

$$e(f, g) = \sum_{k=1}^{n} e(f_k, g_k), \qquad (43)$$

then restoration error is measured by

$$E[e(f, \Phi^{-1}Q\Phi(f\langle\rangle g))] = \sum_{k=1}^{n} E[e(f_k, q_k\Delta(f_k \otimes g_k))]. \qquad (44)$$

The error is minimized by selecting

$$q_k = \begin{cases} 1 & \text{if } E[e(f_k, f_k \otimes g_k)] \leq E[e(f_k, 0)], \\ 0 & \text{otherwise.} \end{cases} \qquad (45)$$

In the usual Hilbert space setting, e is the inner product.

Relative to the degraded-image hole spectrum $H[S \cap N]$ of (12), the function $Q = [Q_j]$ is defined by

$$Q([X_j \cup Z_j]) = [Q_j \cap (X_j \cup Z_j)], \qquad (46)$$

where Q_j is a subset of $Z \times Z$. Under binary optimization with Δ being intersection, $Q_j = Z \times Z$ or $Q_j = \emptyset$. According to (45), $Q_j = Z \times Z$ and if and only if

$$E[c[(X_j \cup Z_j) \oplus N_j - X_j \oplus N_j]] \leq E[c[X_j \oplus N_j]], \qquad (47)$$

which under the modeling assumptions reduces to

$$E[c[Z_j \oplus N_j]] \leq E[c[X_j \oplus N_j]]. \qquad (48)$$

Since $c[Z_j \oplus N_j] = h_N[j]c[N_j]$ and since a similar expression applies to $X_j \oplus N_j$, we deduce that

$$Q_j = \begin{cases} Z \times Z & \text{if } E[h_N(j)] \leq E[h_S(j)], \\ \emptyset & \text{otherwise.} \end{cases} \qquad (49)$$

Having arrived at an optimal solution relative to the spectral representation, we need to find a filter to produce it. In the frequency case the Wiener filter can be implemented either by using the Fourier transform to enter the frequency domain or by applying a convolution. In [5] we show that an operation involving morphological opening and set subtraction can be used to obtain optimization relative to the opening spectrum and that this operation gives the precise result under the modeling assumptions. Here, however, the problem is much more difficult. Unless we place unreasonable restrictions on the model, we cannot escape the overfill–underfill conundrum. Nonetheless, we can proceed by reordering **P** so that $E[h_N(j)] \leq E[h_S(j)]$ if and only if $j \leq q$ and then proceed under this reordering as though $h_N(j) = 0$ for $j \leq q$ and $h_S(j) = 0$ if $j > q$. The net effect will be to implement the binary Wiener paradigm in the spectral domain and then apply the foregoing separated-image-noise filter construction to achieve a morphological (and implementable) form of the filter. Optimal reconstruction will be achieved only if the filter-design algorithm produces no error, but this is precisely the problem we have already confronted.

12. Statistics

In the foregoing hole-spectrum methodology various statistical estimation procedures must be used. We will discuss these in the context of the nondeterministic setting with the understanding that relevant simplications can be made whereever there is determinism. Note that as part of the estimation procedure we need to estimate **P** and **N**, although in many practical circumstances it might be reasonable to assume we know **P** − **N** and therefore need to estimate only **N** to obtain both **P** and **N**. Here we will not make such a modeling assumption, so that we might explain statistical estimation in full. In a similar vein, it might be presumed that we know the L realizations of S; however, here again we assume not and will provide a procedure to obtain these. Throughout the statistical discussion we will assume the modeling assumptions hold.

Beginning with the signal S, let us randomly observe it W times to obtain the random sample $\Omega_S = \{S^1, S^2, \ldots, S^W\}$. Given that W is sufficiently large, we can assume that all realizations $S_1, S_2, \ldots, S_L$ of S occur at least one in Ω_S. By matching we can deduce these from the observations. Moreover, by counting occurrences in Ω_S and dividing by W we obtain estimates for the realization probabilities $p_1, p_2, \ldots, p_L$.

A more interesting problem is to obtain the hole-amplitude spectrums of the S-realizations, as well as the expectation of the random HAS $h[S]$ of S. In fact, the procedure to obtain these involves finding the generic-hole primitives of $\mathbf{P}$. Let S_k be any realization of S. Given our assumption regarding separation of generic holes, we can apply a customary morphological component-filling algorithm to find the generic holes in $\langle S \rangle - S_k$ (see [18], Example 4.6, p. 119). If we do this for $k = 1, 2, \ldots, L$, we will have the primitives in $\mathbf{P} - \mathbf{N}$. Moreover, we will have found $h_{Sk}(j)$ for $k = 1, 2, \ldots, L$ and for all j, as well as $c[N_j]$. By equation (37) we will have also found the expectation of the random HAS for S. What of the enclosures M_j pertaining to $\mathbf{P}-\mathbf{N}$? If we have copies of $S_1, S_2, \ldots, S_L$, these can easily be obtained by finding boundaries of the relevant N_j in the relevant S_k. Three points should be noted: (a) the depth of the boundary, whether it is one, two, three, or more pixels wide, is a modeling question; (b) boundaries can be found by using conditional dilation relative to $\langle S \rangle$; (c) if a particular N_j requires different enclosures in different realizations, or even in the same realization, $\mathbf{P}$ must be reorganized (expanded) so that each of the hole-enclosure pairs are represented in $\mathbf{P}$ and the probabilities p_k need be adjusted in accordance with the reorganization.

The problem of noise estimation is even more interesting. Although the shape primitives comprising $\mathbf{N}$ can be obtained by observing the noise independently of the signal, the expectation of the noise HAS cannot be so estimated because, as may be seen in (40), $E[h_N(j)]$ is expressed in terms of the conditional expectations $E[h_{Nk}(j)]$. Thus we need to observe the conditional processes $N_k = N|S_k$. Given that we estimate $E[h_{Nk}(j)]$ for $k = 1, 2, \ldots, L$ and for all j by making the appropriate observations, we can then find $E[h_N(j)]$ by (40).

Unfortunately, proceeding in strict accordance with the probabilistic assumptions when we are dealing with noise estimation is highly problematic. The difficulty is that the modeling assumptions cannot be expected to be completely satisfied in practice, thereby making strict observation of $N_k = N|S_k$ practically impossible. Consequently, pragmatic estimation will likely require foregoing (40) and estimating $E[h_N(j)]$ by simply counting components in independent observations of the noise. Such an experimental design should provide reasonable estimation so long as the noise components are not too large relative to the generic holes and they are not too dense.

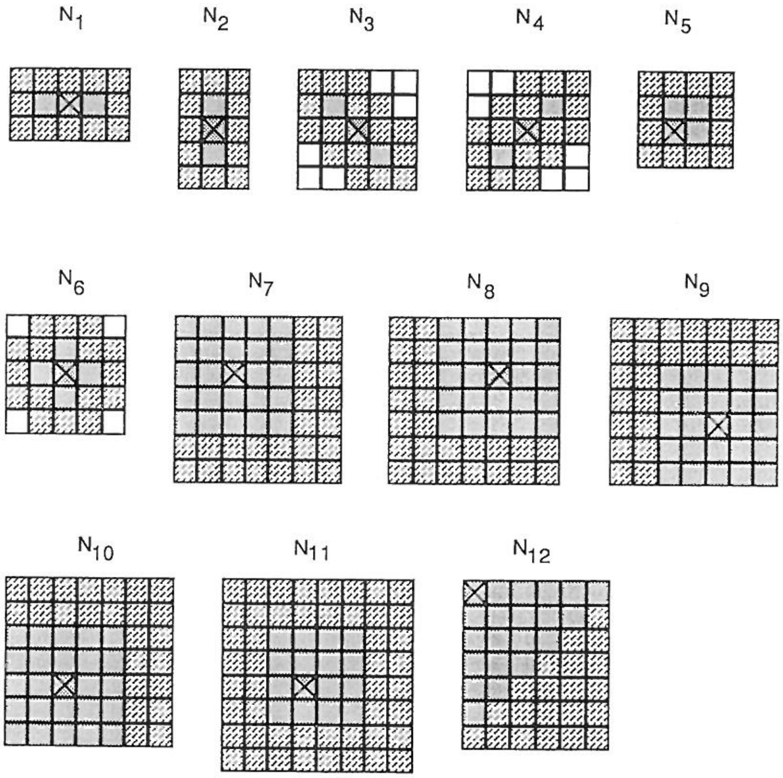
13. Examples

In this section we illustrate various aspects of the hole-spectrum theory with some numerical examples.

Example 1

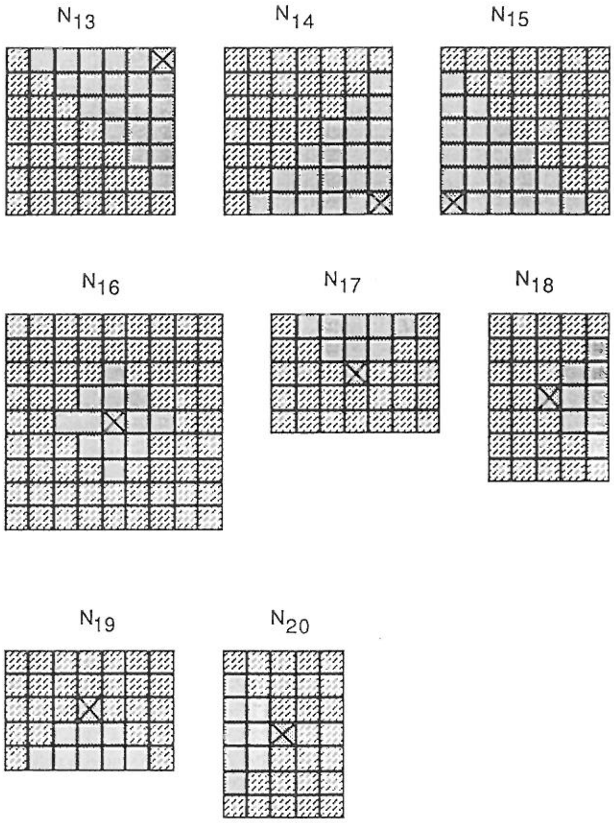
In this example we illustrate the spectra concepts. Let $\mathbf{P} = \{N_1, N_2, \ldots, N_{20}\}$ in figure 2, and let S_1 and S_2 be the images in figures 3 and 4, respectively. The hole-spectra and hole-amplitude spectra of these images are

$$\begin{aligned} H[S_1] &= [\emptyset, \emptyset, \emptyset, \emptyset, \emptyset, \emptyset, \{(-7, 8)\}, \\ &\quad \{(8, 8)\}, \{(8, -7)\}, \\ &\quad \{(-7, -7)\}, \{(0, 0)\}, \\ &\quad \emptyset, \emptyset, \emptyset, \emptyset, \emptyset, \emptyset, \emptyset, \emptyset, \emptyset], \\ H[S_2] &= [\emptyset, \emptyset, \emptyset, \emptyset, \emptyset, \emptyset, \emptyset, \emptyset, \emptyset, \emptyset, \emptyset, \\ &\quad \{(-10, 10)\}, \{(10, 10)\}, \\ &\quad \{(10, -10)\}, \{(-10, -10)\}, \\ &\quad \{(0, 0)\}, \{(0, 8)\}, \{(8, 0)\}, \\ &\quad \{(0, -8)\}, \{(-8, 0)\}], \end{aligned}$$



(a)

Fig. 2.



(b)

Fig. 2. (continued)

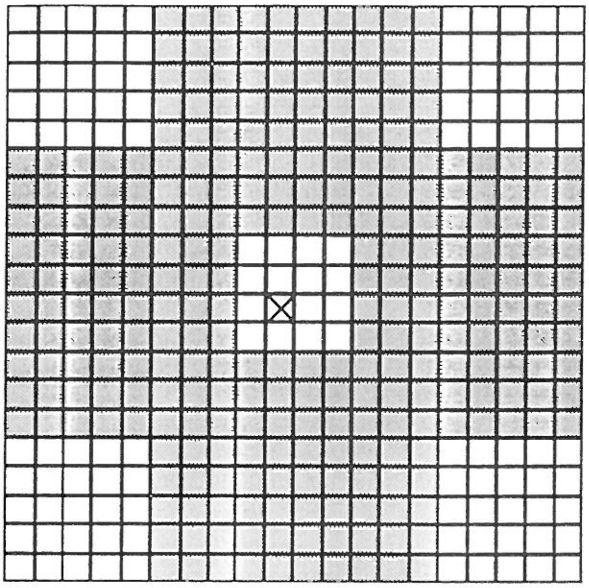


Fig. 3.

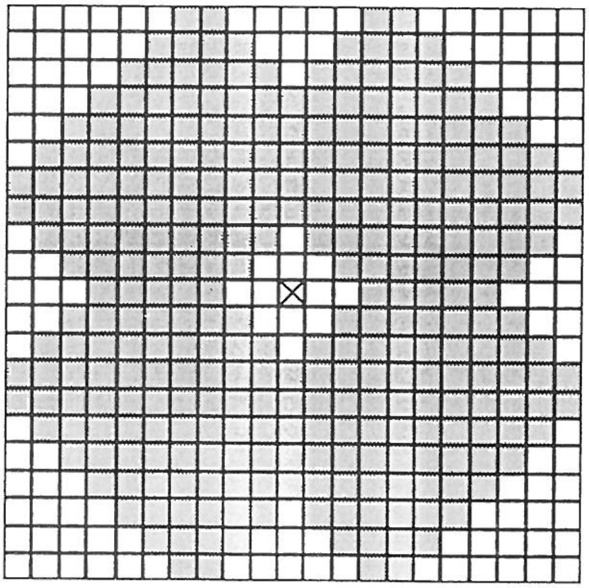


Fig. 4.

$$h[S_1] = [0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0],$$
$$h[S_2] = [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1].$$

The noise image N in figure 5 conforms to S_1, and its spectra are

$$H[N] = [\{(3, 4), (3, 8), (-4, 1)\}, \{(7, 1)\}, \{(-1, 6)\}, \{(3, -3), (1, -7), (-7, 2)\}, \{(5, -2)\}, \emptyset, \emptyset, \emptyset, \emptyset, \emptyset, \emptyset, \emptyset, \emptyset, \emptyset, \emptyset, \emptyset, \emptyset, \emptyset, \emptyset, \emptyset],$$
$$h[N] = [3, 1, 1, 3, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0].$$

Figure 6 shows the noise-corrupted image $S_1 \cap N$. Its spectra are

$$H[S_1 \cap N] = H[S_1] \cup H[N],$$
$$h[S_1 \cap N] = h[S_1] + h[N] = [3, 1, 1, 3, 1, 0, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0].$$

Example 2

In this example we illustrate the fill matrix and the various forms of the filtering algorithm. We use structuring-element set $\mathbf{E} = \{E_1, \ldots, E_8\}$ of figure 7, and we let S be a random image with the realizations S_1 and S_2 of figures 3 and 4; their spectra are given in Example 1. The fill matrix for S relative to $\mathbf{P}$ and $\mathbf{E}$ of figures 2 and 7 is given in Table 1. Note that it is partitioned because, relative to N, $q = 6$ is a separation index for both S_1 and S_2. We now examine the filter-design algorithm under various assumptions on S and N, in each case assuming the noise N is random and conformable.

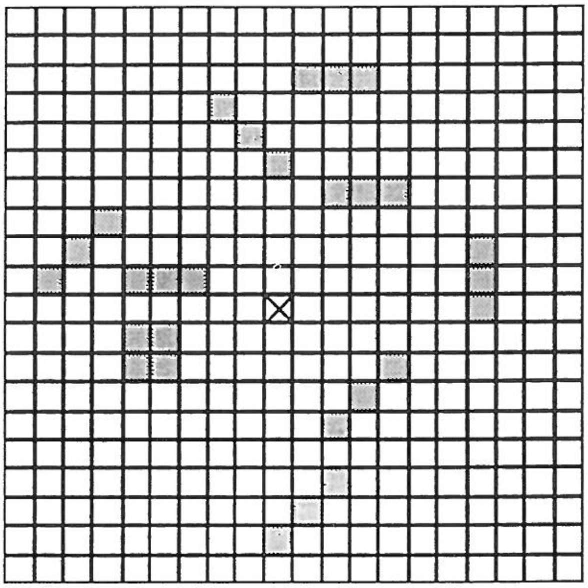


Fig. 5.

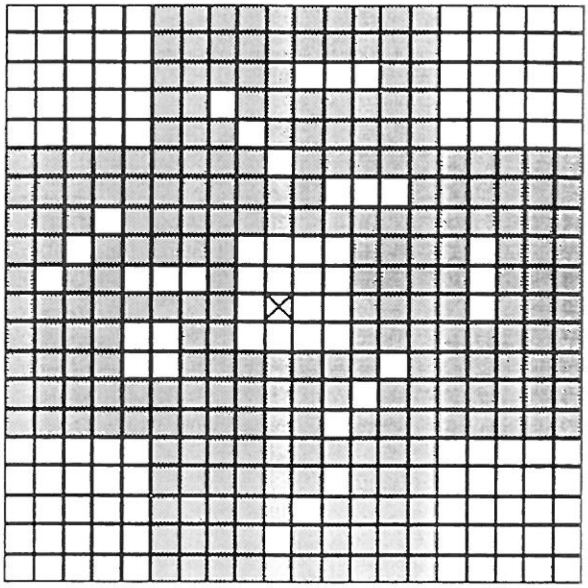


Fig. 6.

Table 1.

	1	2	3	4	5	6
1	1	0	0	0	0	0
2	0	1	0	0	0	0
3	0	0	1	0	0	1/5
4	0	0	0	1	0	1/5
5	1	0	1	1	0	2/5
6	0	1	1	1	0	2/5
7	1	1	1	0	1/2	1/5
8	1	1	0	1	1/2	1/5

	7	8	9	10	11	12	13
1	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0
7	1/25	0	1/25	0	1/8	0	0
8	0	1/25	0	1/25	1/8	0	0

	14	15	16	17	18	19	20
1	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0
5	0	0	2/13	0	1/9	0	1/9
6	0	0	2/13	1/9	0	1/9	0
7	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0

Case (*a*). S is deterministic, $S = S_1$, and

$$E[h[N]] = [2, 2, 2, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0].$$

Scanning down column 1 of the fill matrix, we see that, since $r_{11} = 1$, E_1 fills N_1 holes, and, since $r_{1j} = 0$ for $j > 6$, there is no overfilling. Thus $E[1] = \{E_1\}$. Scanning down column 2, we see that similar reasoning applies to $E_2 (r_{22} = 1$ and $r_{2j} = 0$ for $j > 6)$, so that $E[2] = \{E_1, E_2\}$. Proceeding, we find that $r_{33} = r_{44} = 1$, $E[3] = \{E_1, E_2, E_3\}$,

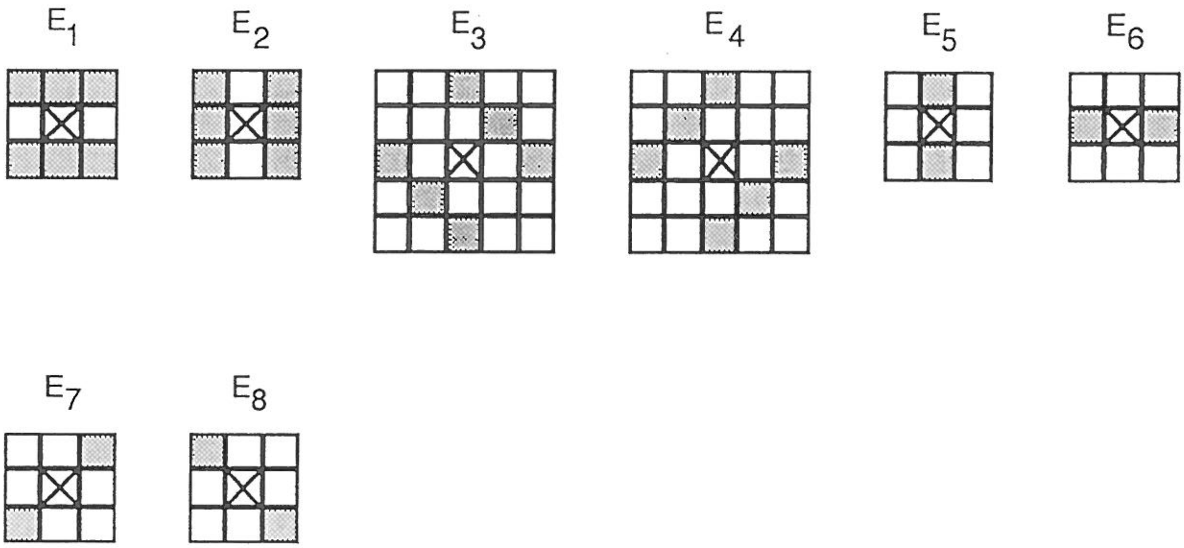


Fig. 7.

$E[4] = \{E_1, E_2, E_3, E_4\}$, and there has been no overfilling. Since $E[h_N(j)] = 0$ for $j > 4$, the algorithm can terminate with

$$\mathbf{B} = E[6] = E[5] = E[4] = \{E_1, E_2, E_3, E_4\}$$

and full restoration (MSE = 0) under the modeling assumptions. Note that $\{E_2, E_5\}$ is also a fully restoring basis; however, it does not result from the algorithm because there is no looking back to find a basis with a minimal number of elements.

Case (*b*). S is deterministic, $S = S_1$, and

$$E[h[N]] = [2, 2, 2, 2, 1, \frac{1}{2}, 0, 0, \ldots, 0].$$

The algorithm proceeds exactly as in case (a) up through $E[4]$, but now $E[h_N(5)] = 1$. Although we could fix any noise hole created by N_5 by adjoining E_7 and E_8 to $E[4]$, by examining $r_{7,7}$, $r_{7,9}$, and $r_{7,11}$, we see that adjoining E_7 creates more overfilling error (in the expected-value sense) than it corrects noise error. Indeed, the overfill error due to E_7 is

$$\sum_{j=7,9,11} h_S(j) r_{7,j} c[N_j] = 4,$$

whereas the expected restoration of N_5-created holes by E_5 is only $h_N(5) r_{7,5} c[N_5] = 2$. Similar reasoning applies to using E_8, so that $E[5] = E[4]$. N_6-created holes are $\frac{4}{5}$ filled by using E_5 and E_6. Although using E_7 in addition will fully fix N_6 holes, again the cost is too high, and so too it is for E_8. Thus we adjoin E_5 and E_6 to $E[5]$ to form

$$\mathbf{B} = E[6] = \{E_1, E_2, E_3, E_4, E_5, E_6\}.$$

As constructed, $\Psi_{\mathbf{B}}$ does no overfilling and yields MSE= 0.0125.

Case (*c*). S is deterministic, $S = S_1$, and

$$E[h[N]] = [2, 2, 2, 2, 3, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0].$$

Again the algorithm proceeds as in case (a) through $E[4]$. It differs from case (b) because the underfill error from not using E_7 to fill N_5-created holes is now 6, which is greater than the overfill error, which is still 4. Thus it proceeds with $E[5] = \{E_1, E_2, E_3, E_4, E_7, E_8\}$ and $\mathbf{B} = E[6] = \mathbf{E}$. Note that N_6-created holes are now fully fixed, since E_5 finishes the job of E_7 and E_8. The resulting mean-square error is MSE= 0.02.

Case (*d*). S is nondeterministic, and realizations S_1 and S_2 possess equal probability. Thus its

HAS is given by

$$h[S] = [0, 0, 0, 0, 0, 0, \frac{1}{2}, \frac{1}{2}, \dots, \frac{1}{2}].$$

If the noise HAS is used from case (a), then the algorithm proceeds exactly as in case (a).

Case (e). S is the random image of case (d), and N has the HAS of case (b). Thus the algorithm runs $E[1] = \{E_1\}$, $E[2] = \{E_1, E_2\}$, $E[3] = \{E_1, E_2, E_3\}$, and

$$\mathbf{B} = E[6] = E[5] = E[4] = \{E_1, E_2, E_3, E_4\}$$

with MSE= 0.015. Note that, as opposed to the deterministic situation in case (b), here we must be concerned with S_2, so that using E_5 and E_6 to fix holes created by N_6 can create overfilling. Since the expected number of error pixels fixed by E_5 is 1 and the expected number of overfilled pixels is

$$P(S_2) \sum_{j=16,17,20} h_{S2}(j) r_{5,j} c[N_j] = 2,$$

E_5 is not adjoined to $E[5]$ in forming $E[6]$. Note also that E_7 is not adjoined to $E[4]$ in forming $E[5]$ because the expected underfilling errors due to not adjoining equals the expected number of overfilling errors due to adjoining, i.e., 2.

14. Conclusion

The image representation developed in the present paper facilitates model-based optimization for a certain class of image-noise processes over a relevant class of morphological filters, namely, those that fill holes. Relative to the general theory of morphological optimization developed in [1], [2], the present approach is akin to the Wiener-frequency design pertaining to optimal linear filters. In place of design by means of frequency components, there is design by means of shape-based hole spectra. The loss of generality relative to the theory of [1] is compensated for by a resulting design procedure that is both intuitive and more computationally tractable. A drawback of the present theory, as opposed to the linear orthonormal-system-based approach, is that there needs to be a greater number of model constraints for the theory to apply with full rigor. Owing to the nonlinearity inherent in shape-based decomposition, this should not be surprising.

Based on the shape-based spectral decomposition, various algorithms for designing spatial morphological filters have been suggested; the most general one applies to nondeterministic signal and noise. These algorithms employ the fill matrix and are facilitated by overfill and underfill error estimates. They also require statistical estimates for various parameters, and we have discussed appropriate estimators. Actual implementation of the algorithms will depend on a number of tradeoffs. For instance, there must be a suitable method of constructing **E** based on the noise primitives because the goodness of any optimal filter designed with the hole-spectrum paradigm will depend on **E**. Moreover, the various tradeoffs between overfilling, underfilling, and design time require investigation.

References

1. E.R. Dougherty, "Optimal Mean-Square N-Observation Digital Morphological Filters—Part 1: Optimal Binary Filters," *J. Comput. Vis., Graph., Image Process.*, vol. 55, 1992, pp. 36-54.
2. E.R. Dougherty, "Optimal Mean-Square N-observation Digital Morphological Filters—Part II: Optimal Gray-Scale Filters," *J. Comput. Vis., Graph., Image Process.*, vol. 55, 1992, pp. 55-72.
3. E.R. Dougherty and R.M. Haralick, "The Hole Spectrum—Model-Based Optimization of Morphological Filters," *Proc. Soc. Photo-Opt. Instrum. Eng.*, vol. 1568, 1991.
4. R.M. Haralick, E.R. Dougherty, and P. Katz, "Model-Based Morphology," *Proc. Soc. Photo-Opt. Instrum. Eng.*, vol. 1472, 1991.
5. R.M. Haralick, E.R. Dougherty, and P. Katz, "Model-Based Morphology: The Opening Spectrum," Morphological Imaging Laboratory Report MIL-91-12, Rochester Institute of Technology, Rochester, NY, August 1991 (submitted for publication).
6. D. Schonfeld and J. Goutsias, "Optimal Morphological Filters for Pattern Restoration," *Proc. Soc. Photo-Opt. Instrum. Eng.*, vol. 1199, 1989.
7. D. Schonfeld and J. Goutsias, "Optimal Morphological Pattern Restoration from Noisy Binary Images," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 13, 1991.
8. R. Lougheed, "Lecture Notes for Summer Course in Mathematical Morphology," University of Michigan, Ann Arbor, MI, 1983.

9. J. Serra, "Alternating Sequential Filters," in *Image Analysis and Mathematical Morphology*, vol. 2, J. Serra, ed., Academic Press, New York, 1988.
10. E.R. Dougherty, A. Mathew, and V. Swarnakar, "A Conditional-Expectation-Based Implementation of the Optimal Mean-Square Binary Morphological Filter," *Proc. Soc. Photo-Opt. Instrum. Eng.*, vol. 1451, 1991.
11. E.R. Dougherty and R.P. Loce, "Constrained Optimal Digital Morphological Filters," in *Proc. 25th Annual Conference on Information Sciences and Systems*, Baltimore, March 1991.
12. R.P. Loce and E.R. Dougherty, "Using Structuring-Element Libraries to Design Suboptimal Morphological Filters," *Soc. Photo-Opt. Instrum. Eng.*, vol. 1568, 1991.
13. R.P. Loce and E.R. Dougherty, "Facilitation of Optimal Binary Morphological Filter Design via Structuring Element Libraries and Design Constraints," J. Opt. Engrg., vol.31, May, 1992, pp. 1008-1025.
14. G. Matheron, *Random Sets and Integral Geometry*, John Wiley, New York, 1975.
15. J. Serra, *Image Analysis and Mathematical Morphology*, Academic Press, New York, 1983.
16. R.M. Haralick, S. Sternberg, and X. Zhuang, "Image Analysis Using Mathematical Morphology," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 9, 1987.
17. E.R. Dougherty and C.R. Giardina, *Image Processing: Continuous to Discrete*, Prentice-Hall, Englewood Cliffs, NJ, 1987.
18. C.R. Giardina and E.R. Dougherty, *Morphological Methods in Image and Signal Processing*, Prentice-Hall, Englewood Cliffs, NJ, 1988.
19. J. Serra and L. Vincent, "Lecture Notes on Morphological Filtering," Center de Morphologie Mathematique de Fontainebleau, Ecole National Superieure des Mines de Paris, Fontainbleau, 1989.
20. J. Serra, "Mathematical Morphology for Complete Lattices," in *Image Analysis and Mathematical Morphology*, vol. 2, J. Serra, ed., Academic Press, New York, 1988.
21. J. Serra, "Mathematical Morphology for Boolean Lattices," in *Image Analysis and Mathematical Morphology*, vol. 2, J. Serra, ed., Academic Press, New York, 1988.
22. H.J. Heijmans and C. Ronse, "The Algebraic Basis of Mathematical Morphology—Part I: Dilations and Erosions," *J. Comput. Vis., Graph., Image Process.*, vol. 50, 1990.
23. J. Serra, "Introduction to Morphological Filters," in *Image Analysis and Mathematical Morphology*, vol. 2, J. Serra, ed., Academic Press, New York, 1988.
24. E.R. Dougherty and C.R. Giardina, "A Digital Version of Matheron's Theorem for Increasing Tau-Mappings in Terms of a Basis for the Kernel," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, Miami, June 1986.
25. P. Maragos and R.W. Schafer, "Morphological Filters—Part I: Their Set-Theoretic Analysis and Relations to Linear Shift-Invariant Filters," *IEEE Trans. Acous. Speech, Signal Process.*, vol. 35, 1987.
26. P. Maragos, "A Representation Theory for Morphological Image and Signal Processing," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 11, 1989.

Edward R. Dougherty is an associate professor at the Center for Imaging Science of the Rochester Institute of Technology and also serves as an industrial consultant. He holds an M.S. in computer science from Stevens Institute of Technology and a Ph.D. in mathematics from Rutgers University. He has written numerous papers in the area of mathematical morphology and image algebra and has authored or coauthored six books: *An Introduction to Morphological Image Processing, Matrix Structured Image Processing, Image Processing—Continuous to Discrete, Morphological Methods in Image and Signal Processing, Mathematical Methods for Artificial Intelligence and Autonomous Systems*, and *Probability and Statistics for the Engineering, Computing, and Physical Sciences.* He regularly teaches conference short courses in morphological image processing, including those for SPIE, and serves as a chair for two SPIE conferences, *Image Algebra and Morphological Image Processing* and *Nonlinear Image Processing.* His current research involves several principal areas: design of statistically optimal morphological filters, development of model-based approaches to morphological image analysis, granulometric characterization of texture, and development of instrinsically fuzzy approaches to mathematical morphology.

Robert M. Haralick is the Boeing Clairmont Egtvedt Professor in Electrical Engineering at the University of Washington. His recent work is in shape analysis and extraction using mathematical morphology, robust pose estimation, and techniques for making geometric inference from perspective projection information.

Professor Haralick has made a series of contributions in the high-level vision area, specifically dealing with the consistent labeling problem, arrangements, relation homomorphism, matching, and tree search. In the low-level and mid-level areas, Professor Haralick has worked in image texture analysis using spatial gray-tone co-occurrence texture features, and in the feature detection area he has developed the facet model for image processing.

Professor Haralick is a Fellow of IEEE for his contributions in computer vision and image processing. He serves on the Editorial Board of *IEEE Transactions on Pattern Analysis and Machine Intelligence.* He is the computer vision area editor for *Communications of the ACM*. He also serves as associate editor for *Computer Vision, Graphics, and Image Processing and Pattern Recognition*.

He received a B.A. degree in mathematics in 1964, a B.S. degree in electrical engineering in 1966, an M.S. degree in electrical engineering in 1967, and his Ph.D. in 1969, all from the University of Kansas.