Optimal Filtering of Digital Binary Images Corrupted by Union/Intersection Noise

N. D. Sidiropoulos, *Member, IEEE,* John S. Baras, *Fellow, IEEE,* and Carlos A. Berenstein

Abstract—**We model digital binary image data as realizations of a uniformly bounded discrete random set (or discrete random set, for short), which is a mathematical object that can be directly defined on a finite lattice. We consider the problem of estimating realizations of discrete random sets distorted by a degradation process that can be described by a union/intersection noise model. Two distinct optimal filtering approaches are pursued. The first involves a class of "mask" filters, which arises quite naturally from the set-theoretic analysis of optimal filters. The second approach involves a class of morphological filters. We prove that under i.i.d noise morphological openings, closings, unions of openings, and intersections of closings can be viewed as MAP estimators of morphologically smooth signals. Then, we show that by using an appropriate (under a given degradtion model) expansion of the optimal filter, we can obtain universal characterizations of optimality that do not rely on strong assumptions regarding the spatial interaction of geometrical primitives of the signal and the noise. The results generalize to gray-level images in a fairly straightforward manner.**

I. Introduction

An important problem in digital image processing and analysis is the development of optimal filtering procedures that attempt to restore a binary image ("signal") from its degraded version [25], [8]. Here, the degradation mechanism usually models the combined effect of two distinct types of distortion, namely, image object obscurations because of clutter and sensor/channel noise. It is typically assumed that the degraded image can be accurately modeled as the union of the uncorrupted binary image with an independent realization of the noise process, which is a binary image itself [15]. This degradation model is known as the union noise model. Other models exist, such as the intersection noise model, and the union/intersection noise model, which are defined in the obvious fashion. The assumption of independence is crucial for the theoretical analysis of optimal filters, and it is plausible in many practical situations. These models are rather general in that they can be tailored to describe most popular types of signal-independent noise, e.g., salt-and-pepper noise (also known as binary symmetric channel (BSC) transmission noise), burst channel errors, noise with a geometric structure [15], occlusion, etc.

This research has been largely motivated by the works of Haralick *et al.* [15] and Schonfeld and Goutsias [25]. Their approach is model based in that they assume specific probabilistic/geometrical models that govern the behavior of both signal and noise "patterns," i.e., the elementary geometrical primitives from which the signal and noise images are constructed. Haralick *et al.* assume that the signal and noise patterns are "noninterfering" with one another, meaning that each signal or noise pattern is disconnected from all remaining signal and noise patterns. Schonfeld and Goutsias make a stronger assumption concerning the separability of noise patterns. These assumptions are reasonable if the image is sparse, i.e., the signal and noise patterns are most likely to remain uncluttered. Haralick *et al.* adopt the area of the symmetric set difference between the ideal image and its reconstruction as their choice of distance metric and work with a union noise model to derive the optimal (in the sense of minimizing the expected distance between the signal and its reconstruction) value of a "size" parameter that determines the optimal filter within a restricted family of morphological opening filters [22], [26], [7]. In their work, the signal and noise patterns are all assumed to be of the same basic shape, and only their size varies. Schonfeld and Goutsias consider morphological alternating sequential filters (ASF's) [22], [26], [7] and work with the union/intersection noise model. They adopt an implicit least mean difference "uniform" optimality criterion (i.e., the best filter, within a family of filters, is defined to be the one that minimizes an average (over the family) distance metric between the *outputs* of all the filters in the family for a given class of inputs). They derive the "optimal" ASF by means of minimizing an upper bound on their cost function. Related work can also be found in a series of papers by Dougherty *et al.* [8]–[10].

This work focuses on a different viewpoint. As it turns out, by restricting our attention to suitable classes of filtering operations and uniformly bounded discrete random sets (defined below), we can obtain optimal filtering results *under considerably milder assumptions on the signal and noise patterns*, i.e., results that are applicable for all signal and noise models under the assumption of mutual independence of the signal and the noise. Specifically, one need not assume that signal and noise patterns are "noninterfering." Furthermore, it is possible to obtain simple, closed characterizations of the optimal filter. The resulting formulas are intuitively appealing, and directly amenable to design and implementation.

Some motivating comments are in order. A fair question to ask is whether it is necessary for the purposes of filtering to model binary image data as random sets. We feel that it is for two reasons. First, this enables us to formulate

Manuscript received July 2, 1992; revised April 19, 1993. This work was supported by NSF grant NSFD CDR 8803012, through the Engineering Research Center's Program.

The authors are with the Institute of Systems Research, University of Maryland, College Park, MD 20742.

IEEE Log Number 9400312.

the optimal filtering problem within a rigorous statistical framework. Second, random set theory is closely related to mathematical morphology, which is a nonlinear image algebra that effectively addresses the problem of quantitative shape description. Thus, random set theory allows the simultaneous modeling of two important aspects of binary images: geometric structure and statistical behavior. In and by itself, neither one of the two can provide a complete summary of the images under consideration. In the terminology of nonlinear filtering, our unified approach allows the joint optimization of both the *syntactical* and the *statistical* properties of a filter structure.

Another important question is how much common ground exists between the optimal filtering problem for discrete random sets and standard optimal filtering theory for the case of real or vector-valued random variables. To what extent can we translate well established results (e.g., the orthogonality principle) in the discrete random set setting? The answer is that the analogy is rather superficial. The major difference is that our problem does not have the nice vector space structure that underscores classical optimal filtering theory. We will explain this in detail.

It is important to note that our results can be extended to finite-gray-level digital images of compact support and sup/inf noise via threshold decomposition of functions and/or by treating functions as sets via their umbrae[1] (cf. [20], [7], [6], [27] among others), which allow for a rigorous treatment of random finite-gray-level digital images of compact support as uniformly bounded discrete random sets defined on a 3-D finite lattice $B \subset \mathbf{Z}^3$. This is achieved as follows. Since the number of gray levels is finite, we can assume, without loss of generality, that the range of gray levels is $\mathcal{A} = 1, \ldots, N$. Then the umbra of a gray-level digital image f of compact support $D \subset \mathbf{Z}^2$ is the subset of $B = D \times \mathcal{A} \subset \mathbf{Z}^3$ beneath the graph of f. It is easy to see that f completely determines, and is completely determined by, its umbra.

The rest of this paper is organized as follows. Section II contains some discrete random set fundamentals, whereas Section III contains a formalization of the optimal filtering problem, including a discussion of our choice of distance metric. Some connections with classical optimal filtering theory are also made, and the fundamental differences are pointed out. Section IV contains results on optimal mask filtering, which is motivated by the set-theoretic analysis of optimal filters. Section V takes on a morphological filtering approach, which is motivated by the widespread use of morphological filters, their well-known shape-preservation properties, and a fresh statistical insight into some "folk theorems" of applied morphological filtering. Some simulation experiments are also included. Finally, Section VI offers some conclusions.

II. Discrete Random Set Fundamentals

Intuitively, a random digital binary image X is an object that takes on "values" in the set of all subsets of a finite (e.g., 512 $\times$ 512) pixel lattice, which we will denote by B, according to some probability law P_X on a suitable σ algebra. Each instance $K \subseteq B$ of image data[2] is thus viewed as a realization of the random object under consideration, and it is assigned a probability $P_X(X = K)$ by the law P_X. Knowledge of $P_X(X = K)$ for all $K \subseteq B$ completely specifies the random object X because it allows the computation of the probabilities of all conceivable events of interest (e.g., What is the probability of $X \cap L = \emptyset$ for some (given) $L \subseteq B$? What is the expected area of X? etc.). However, for all practical purposes, this specification is not convenient to work with. There exist alternative (but equivalent) means to accomplish this specification (in terms of the so-called *generating functional* or other suitable functionals of the law P_X), which are naturally more convenient for modeling and design purposes. We will argue for this point as we move on.

Mathematically, a random digital binary image of compact support can be formally defined as a *uniformly bounded discrete random set.*

Definition 1: Let B be a bounded subset of $\mathbf{Z}^2$. Assume that B contains the origin. Let $\Sigma(\Omega)$ denote the σ algebra on Ω. Let $\Sigma(B)$ denote the power set (i.e. the set of all subsets) of B, and let $\Sigma(\Sigma(B))$ denote the power set of $\Sigma(B)$. A uniformly bounded discrete random set, or, for brevity, **discrete random set (DRS)** X on B is a measurable mapping of a probability space $(\Omega, \Sigma(\Omega), P)$ into the measurable space $(\Sigma(B), \Sigma(\Sigma(B)))$. A DRS X on B induces a unique probability measure P_X on $\Sigma(\Sigma(B))$.

Definition 2: The functional

$$T_X(K) = P_X(X \cap K \neq \emptyset), K \in \Sigma(B)$$

is called the **capacity functional** of the DRS X.

Definition 3: The functional

$$Q_X(K) = P_X(X \cap K = \emptyset) = 1 - T_X(K), \ K \in \Sigma(B)$$

is called the **generating functional** of the DRS X.

In the context of DRS's, the generating functional plays a role analogous to the one played by the cumulative distribution function (cdf) in the context of scalar discrete random variables. The following lemma will be useful. Its proof can be found in the Appendix. See [1] for basic Moebius inversion.

Lemma 1: (Variant of Moebius inversion for Boolean algebras) Let v be a function on $\Sigma(B)$. Then, v can be represented as

$$v(A) = \sum_{S \subseteq A^c} u(S) \quad \text{"external decomposition."}$$

The function u is uniquely determined by v, namely

$$u(S) = \sum_{C \subseteq S} (-1)^{|C|} v(S^c \cup C)$$

where ^{c} denotes complement with respect to B.

The capacity functional T_X (or, equivalently, the generating functional Q_X) carries all the information about a DRS X. This is clearly stated in the following theorem.

[1] The *umbra* of a function of n variables is the set of all points beneath the graph of the function in $n + 1$ space.

[2] By convention, K is taken to be the set of all black pixels in the image.

Theorem 1: Given $Q_X(K), \forall K \in \Sigma(B)$, $P_X(A), \forall A \in \Sigma(\Sigma(B))$ is uniquely determined and, in fact, can be recovered via the measure reconstruction formulas

$$P_X(A) = \sum_{K \in A} P_X(X = K)$$

with

$$P_X(X = K) = \sum_{K' \subseteq K} (-1)^{|K'|} Q_X(K^c \cup K').$$

Proof: The functional Q_X can be expressed in terms of P_X as

$$Q_X(K) = \sum_{K' \subseteq K^c} P_X(X = K').$$

This observation, along with Lemma 1, establishes the validity of the theorem. □

The *uniqueness* part of this theorem is originally due to Choquet [3], and it has been independently introduced in the context of continuous-domain random set theory by Kendall [17] and Matheron [21], [22]. Related results can also be found in Ripley [24]. However, the measure reconstruction formulas are essentially only applicable within a uniformly bounded discrete random set setting. In the case of (uncountably or countably) infinite observation sites, the uniqueness result relies heavily on Kolmogorov's extension theorem, which is nonconstructive. See [12]–[14], [28]–[33] for some other interesting results on DRS's. (In [13] and [14], DRS's are defined on the infinite lattice $\mathbf{Z}^2$, and the results evolve in different directions than ours.)

The capacity functional plays an important role in the study of statistical inference problems for DRS's. This is especially true for a class of DRS models known as *germ-grain* models and the Boolean model in particular, whose capacity functional has a simple, tractable form. We will see that the capacity functional has an equally fundamental role in the study of optimal filtering.

III. Formulation of the Optimal Filtering Problem

Let X, N, Y be DRS's on B. X models the "signal," whereas N models the noise. Let $g : \Sigma(B) \times \Sigma(B) \mapsto \Sigma(B)$ be a mapping that models the degradation (measurability is automatically satisfied here since the domain of g is finite). The observed DRS is $Y = g(X, N)$. Let $d : \Sigma(B) \times \Sigma(B) \mapsto \mathbf{Z}_+$ be a distance metric between subsets of B. In this context, the optimal filtering problem is to find a mapping $f : \Sigma(B) \mapsto \Sigma(B)$ such that the cost (expected error)

$$E(e) \triangleq \mathrm{E}d(X, \hat{X}), \qquad \hat{X} = f(Y) = f(g(X, N))$$

is minimized over all possible choices of the mapping ("filter") f. This problem is in general intractable. The main difficulty is the lack of structure on the search space. The family of all mappings $f : \Sigma(B) \mapsto \Sigma(B)$ is just too big. It is common practice to *impose* structure on the search space, i.e., constrain f to lie in $\mathcal{F}$, which is a suitably chosen subcollection of *admissible* mappings (family of filters), and optimize within this subcollection. The resulting filter is the best among its peers, but it is not guaranteed to be globally optimal.

We adopt the following distance metric (area of the symmetric set difference)

$$\begin{aligned} d(X, \hat{X}) &= |(X \backslash \hat{X}) \cup (\hat{X} \backslash X)| \\ &= |(X \backslash \hat{X})| + |(\hat{X} \backslash X)| \\ &= |(X \cup \hat{X}) \backslash (X \cap \hat{X})| \\ &= |(X \cup \hat{X})| - |(X \cap \hat{X})| \end{aligned}$$

where $|\ |$ stands for set cardinality, $\backslash$ stands for set difference, i.e., $X \backslash Y = X \cap Y^c$, and ^{c} stands for complementation with respect to the base frame B. This distance metric is essentially the *Hamming distance* [23] when $X, \hat{X}$ are viewed as vectors in $\{0,1\}^{|B|}$. Since the component variables are binary, it can also be interpreted as the square of the L_2 distance of vectors in $\{0,1\}^{|B|}$, i.e., with some abuse of notation

$$d(X, \hat{X}) = (X - \hat{X})^T (X - \hat{X})$$

where on the left-hand side symbols are interpreted as sets, whereas on the right-hand side, symbols are interpreted as column vectors in $\{0,1\}^{|B|}$, and ^{T} stands for transpose. In this setting, the *sufficiency part* of the orthogonality principle (OP) [23] applies. It states that a sufficient condition for the existence of an $f^* \in \mathcal{F}$ such that

$$\begin{aligned} &E\big[(X - f^*(Y))^T (X - f^*(Y))\big] \\ &\quad \le E\big[(X - f(Y))^T (X - f(Y))\big], \forall f \in \mathcal{F} \end{aligned}$$

is that

$$E\big[(X - f^*(Y))^T (f^*(Y) - f(Y))\big] = 0, \forall f \in \mathcal{F}.$$

However, unlike the case of vectors in R^n, where $\mathcal{F}$ is a vector space over the field of reals (known as the space of *square integrable* estimators), here, $\mathcal{F}$ is not a vector space. The proof of the necessity part of the OP strongly depends on $\mathcal{F}$ having a vector space structure. When $\mathcal{F}$ does not have a vector space structure, the key notion is that of *conditional expectation*. Here, however, we run into trouble defining what we mean by conditional expectation of a DRS X given a DRS Y, let alone computing it. Fortunately, it turns out that it is often possible to write down an expression for $\mathrm{E}[\mathrm{d}(X, \hat{X})]$ and optimize over $\mathcal{F}$ by brute force.

Technically speaking, $d(X, \hat{X})$ can be considered as a quadratic distance measure when we view $X, \hat{X}$ as vectors in $\{0,1\}^{|B|}$. From a set-theoretic point of view, $d(X, \hat{X})$ is clearly not a quadratic distance measure since it penalizes errors in a linear fashion. However, the squared area of the symmetric set difference (which is a quadratic distance measure in the set-theoretic sense) does not yield useful optimality conditions. This is partly due to the lack of a meaningful and tractable definition for the *expectation of a DRS* X. In the continuous case, there exists such a definition. The expectation of a random compact set X can be defined via the expectation of *random selections*, i.e., random vectors that are a.s. contained in X. The expectation of X is defined as the union of expectations of all its random selections. Random

selections exist, and the resulting notion of expectation of a random compact set is well defined. This definition is adapted to the development of strong limit theorems [40], [41], [2], [37], [36], [39], [38], [4]. However, it is not convenient for our purposes. Consider a uniformly bounded DRS X defined on a finite lattice $B \subset \mathbf{Z}^2$. A random selection from X is a random vector taking values in B; however, its expectation is not necessarily a point in B. Thus, the resulting expectation of X will not necessarily be a subset of B. Furthermore, in contrast to the expectation of a random variable or random vector, this notion of expectation of a random compact set (measurable mapping) depends not only on the induced distribution over the space of realizations but also on the mapping itself. This surprising fact has interesting ramifications [41]. However, in our case, it introduces unnecessary complications. Finally, here we are not interested in asymptotics when the lattice goes to infinity; instead, we focus on filtering. For this, we need an alternative definition of expectation. From a quadratic estimation-theoretic point of view, a proper *formal* definition of the expectation of a DRS X would be as follows:

$$EX \triangleq \arg \min_{W \in \Sigma(B)} E[d(X,W)]^2.$$

However, there exist several flaws with this formal definition. It can be shown [28] that

$$\begin{aligned}\arg &\min_{W \in \Sigma(B)} E[d(X,W)]^2 \\ &= \arg \min_{W \in \Sigma(B)} \Bigg\{ |W|^2 \\ &\quad + 2|W| \Bigg(\sum_{z \in W^c} \Pr(z \in X) - \sum_{z \in W} \Pr(z \in X) \Bigg) \\ &\quad - 4 \sum_{z \in W^c} \sum_{y \in W} \Pr(z \in X, y \in X) \Bigg\}.\end{aligned}$$

If we assume that $\Pr(z \in X) = p, \forall z \in B$, $\Pr(z \in X, y \in X) = \Pr(z \in X)\Pr(y \in X) = p^2, \forall z, y \ s.t. \ z \neq y$, and $p < 0.5$, then $EX = \emptyset$, regardless of the specific value of p. If $p = 0.5$, then *any* $W \in \Sigma(B)$ will do. However, the single most important problem is that, given a specification of the first and second-order statistics of X, it is not clear how to find an explicit solution to the above minimization problem. On the other hand, the *median of a DRS* X, which is formally defined as

$$MX \triangleq \arg \min_{W \in \Sigma(B)} \mathrm{E[d}(X,W)]$$

is much easier to deal with. Although the solution to this latter minimization problem is not (in general) unique, it can be forced to be unique by means of a simple regularization. Let $C(z)$ be a Boolean proposition that, for each point $z \in B$, is either true or false. Define the *indicator function*

$$1(C(z)) \triangleq \begin{cases} 1, & \text{if } C(z) \text{ is true at } z \\ 0, & \text{otherwise} \end{cases}$$

Let supp $1(C(z))$ stand for the *support set* of the indicator function $1(C(z))$, i.e., the subset of B where $C(z)$ is true. Then, it can be shown [28] that

$$M_R X \triangleq \text{supp } 1(1 - T_X(\{z\}) < T_X(\{z\}))$$

is the unique minimum cardinality solution to the minimization problem

$$\min_{W \in \Sigma(B)} \mathrm{E[d}(X,W)].$$

These considerations essentially dictate our choice of distance metric. In terms of the degradation, we assume that N is independent of X and that the mapping g is given by

$$g(X,N) = X \cup N \quad \text{(union noise model)}$$

or

$$g(X,N) = X \cap N \quad \text{(intersection noise model)}.$$

Although we shall be mainly concerned with either union *or* intersection noise, on some occasions, we will allow g to be a mapping from $\Sigma(B) \times \Sigma(B) \times \Sigma(B)$ to $\Sigma(B)$

$$g(X,N_1,N_2) = (X \cap N_1) \cup N_2 \quad \text{(union/intersection noise model)}$$

where X, N_1, N_2 will be assumed to be mutually independent.

IV. Optimal Mask Filters

In the case of union noise, we can assume, without loss of generality, that the optimal filter is of the form

$$f(Y) = f_W(Y) = Y \cap W = (X \cup N) \cap W, \quad \text{for some } W \in \Sigma(B).$$

Similarly, in the case of intersection noise, we can assume that the optimal filter is of the form

$$f(Y) = f^W(Y) = Y \cup W = (X \cap N) \cup W, \quad \text{for some } W \in \Sigma(B).$$

Finally, in the case of combined union/intersection noise, we can assume that the optimal filter is of the form

$$\begin{aligned} f(Y) = f^{W_1}_{W_2}(Y) &= (Y \cap W_2) \cup W_1 \\ &= ((((X \cap N_1) \cup N_2) \cap W_2) \cup W_1), \\ &\quad \text{for some } W_1, W_2, \text{both in } \Sigma(B). \end{aligned}$$

We will collectively refer to these filters as *mask filters*. For example, in the case of union noise, the optimal filter should retain a subset of the observation points and reject the rest; this should be done in some sort of statistically optimal fashion. This is achieved by intersecting the observation with a suitably chosen "mask" W, which, in general, depends on the observation.

As a first step, we might want to investigate how much we can achieve using a fixed mask W, i.e., one that does not depend on the observation, and optimize the choice of this fixed mask over all possible observations. We will call the resulting constrained optimal filter the optimal *fixed mask filter*. The second step would be to allow W to depend on the observation via some suitable adaptation strategy. The ideal situation would be to optimize the mask pointwise,

i.e., construct a mapping $W(\cdot):\Sigma(B) \mapsto \Sigma(B)$, which takes an observation and maps it to the best mask for the given observation. However, it seems that in general, this optimization is intractable. Furthermore, the implementation of such a pointwise optimal strategy requires a realization of the mapping $W(\cdot)$, which seems impractical. Nevertheless, we will show that explicit optimization is possible under some restrictions on the adaptation strategy. We will call the resulting constrained optimal filter the optimal *adaptive mask filter*.

Let us first consider fixed mask filtering. Here, we only work with the union/intersection noise model. The other two noise models are special cases. We have the following proposition.

Proposition 1: Under the expected symmetric set difference measure, an optimal fixed pair of masks (W_1, W_2) is given by

$$W_2 = \text{supp } 1(T_X(\{z\}) > \max(T_1(\{z\}), T_2(\{z\})))$$
$$W_1 = \text{supp } 1(T_2(\{z\}) \leq \min(T_X(\{z\}), T_1(\{z\})))$$

whereas the associated minimum cost achieved by such an optimal pair of masks is

$$E(e^*) = \sum_{z \in B} \min\left(T_X(\{z\}), T_1(\{z\}), T_2(\{z\})\right)$$

with

$$T_1(\{z\}) = T_X(\{z\})(1 - T_{N_1}(\{z\})) \times (1 - T_{N_2}(\{z\})) + (1 - T_X(\{z\}))T_{N_2}(\{z\})$$

and

$$T_2(\{z\}) = T_X(\{z\})(1 - T_{N_1}(\{z\})) + (1 - T_X(\{z\})).$$

Proof: **See the Appendix.**

If the first-order statistics (pixel hitting probabilities) of the signal and noise DRS's are spatially invariant, then obviously, the optimal pair of masks is either $(\emptyset, B)$, $(\emptyset, \emptyset)$, or (B, B). In this case, fixed mask filtering is not appropriate. It is exactly when the signal and/or the noise statistics are highly nonstationary (meaning not even first-order stationary) that this filtering approach makes sense. In such a highly nonstationary environment, traditional shift-invariant neighborhood filtering operations (e.g., local mean, median, order statistics) typically fail to provide adequate filtering, and their optimization is very difficult. On the other hand, the optimization of the masks is based on simple statistics, which can be efficiently estimated from training data. A potentially big gain in quality of restoration rests exactly with proper use of the nonstationary nature of the signal and/or the noise.

An obvious drawback of fixed mask filtering is that it does not exploit the autocorrelation structure of the signal and the noise. Furthermore, it is nonadaptive. Whenever higher order statistics are available, we would like to use them. We would also like to allow for an adaptation of the mask using information extracted from the given input. Adaptive mask filtering fits both bills. The tradeoff is an increase in design/implementation complexity.

Let us first consider the case of union noise. Assume that we are presented with a specific input L, i.e. we are given that $Y = X \cup N = L$. One adaptation strategy is to incorporate this information into the cost function. This is done by considering the *conditional* expectation of $d(X, \hat{X})$ conditioned on $Y = X \cup N = L$. However, this does not lead to a closed-form solution for the optimal filter. The reason is that the minimization of this conditional expectation requires the explicit computation of a pseudo-convolution of likelihoods on the lattice of realizations. This computation is, in general, intractable. Instead, suppose we can *upper bound* the observable DRS Y, i.e., suppose that $Y = X \cup N \subseteq K$ for some $K \subseteq B$. Given this information, we would like to pick W to minimize the conditional expectation

$$E(e|Y = X \cup N \subseteq K).$$

The following proposition addresses this minimization problem.

Proposition 2: Given that $X \cup N \subseteq K$, an optimal intersection mask W for filtering out the noise component N is given by

$$W = K \cap \text{supp } 1([1 - T_X(K^c \cup \{z\})][T_N(K^c \cup \{z\}) - T_N(K^c)] \leq [T_X(K^c \cup \{z\}) - T_X(K^c)][1 - T_N(K^c)]).$$

The corresponding minimum conditional cost achieved by such an optimal choice of W is given by[3]

$$E(e^*|X \cup N \subseteq K) = \frac{1}{(1 - T_X(K^c))(1 - T_N(K^c))} \times \sum_{z \in K} \min\{[1 - T_X(K^c \cup \{z\})] \times [T_N(K^c \cup \{z\}) - T_N(K^c)], [T_X(K^c \cup \{z\}) - T_X(K^c)][1 - T_N(K^c)]\}.$$

Proof: **See the Appendix.**

Observe how information about the higher order statistics of the signal and the noise is incorporated into the filter structure by means of the capacity functionals of the signal and the noise. Note that the minimum conditional cost achieved by an optimal choice of W is not necessarily increasing in K; in the expression for the minimum conditional cost, we can show that the sum is increasing in K, but the normalizing factor $1/((1 - T_X(K^c))(1 - T_N(K^c)))$ is decreasing in K. Given *an observation*, i.e., given that $Y = X \cup N = L$, the obvious (and tightest) upper bound is the observation itself. However, by virtue of the previous remark, this tightest upper bound is not necessarily the best choice. Nevertheless, we will see (cf. Section IV-A) that it is a good choice under a low-noise scenario.

The case of intersection noise can be addressed by appealing to duality. One can simply take the complement of all the sets and operations involved and apply the result that has been obtained for the case of union noise. This is clear because

$$d(X_1, X_2) = d(X_1^c, X_2^c)$$
$$\text{and} \quad ((X \cap N) \cup W)^c = (X^c \cup N^c) \cap W^c.$$

Hence, by conditioning on the event $X^c \cup N^c \subseteq K^c$, i.e., $(X^c \cup N^c) \cap K = \emptyset$, we obtain the following result:

[3] We assume that $\Pr(X \cup N \subseteq K) > 0$. Note that this, in turn, implies that $T_X(K^c) < 1$, and $T_N(K^c) < 1$.

Proposition 3: Given that $X^c \cup N^c \subseteq K^c$, an optimal union ("fill") mask W for filtering out the intersection noise component N is specified by

$$W^c = K^c \cap \operatorname{supp} 1([1-T_{X^c}(K\cup\{z\})][T_{N^c}(K\cup\{z\})-T_{N^c}(K)] \le [T_{X^c}(K\cup\{z\})-T_{X^c}(K)][1-T_{N^c}(K)]).$$

The corresponding minimum conditional cost achieved by such an optimal choice of W is given by[4]

$$E(e^*|X^c\cup N^c\subseteq K^c) = \frac{1}{(1-T_{X^c}(K))(1-T_{N^c}(K))} \times \sum_{z\in K^c} \min\{[1-T_{X^c}(K\cup\{z\})] \times [T_{N^c}(K\cup\{z\})-T_{N^c}(K)], [T_{X^c}(K\cup\{z\})-T_{X^c}(K)] \times [1-T_{N^c}(K)]\}$$

A. An Example: The Discrete Radial Boolean Random Set (DRBRS)

The *Boolean* random set is by far the most important random set model to date. Its importance stems from its power to model many interesting phenomena as well as its analytical tractability. This model has received considerable attention in the literature (see [35] for a review). In a sense, the Boolean random set is a generalization of white noise for the case of spatial processes. Therefore, it is well suited to model random noise and obscuration under any of the degradation models which have been adopted so far. Here, we develop a restricted discrete-case analog of the Boolean model and compute its capacity functional.

The theory of *mathematical morphology* has been developed mainly by Serra [26], [11], Matheron [22], and their collaborators during the 1970's and early 1980's. Since then, mathematical morphology and its applications have become very popular. The theory is concerned with the quantitative analysis of shape with an emphasis on geometric structure. It is founded on certain elementary set-to-set mappings, namely, set dilation/erosion, which are inherently nonlinear. These mappings are defined in terms of a *structuring element*, which is a "small" primitive shape (set of points) that interacts with the input image to transform it and, in the process, extract useful information about its geometrical and topological structure. Let

$$W^s \triangleq \{z\in\mathbf{Z}^2 | -z\in W\}.$$

The dilation of a set $Y\subset\mathbf{Z}^2$ by a structuring element W is defined as[5]

$$Y\oplus W^s = \{z\in\mathbf{Z}^2 | W_z\cap Y\neq\emptyset\}$$

whereas the erosion of a set $Y\subset\mathbf{Z}^2$ by a structuring element W is defined as

$$Y\ominus W^s = \{z\in\mathbf{Z}^2 | W_z\subseteq Y\}.$$

Erosion and dilation are *dual* operators in the sense that $Y\ominus W^s = (Y^c\oplus W^s)^c$, where here, ^{c} stands for complementation with respect to $\mathbf{Z}^2$. Two fundamental composite morphological operators are opening and closing. The opening $Y\circ W$ of a set $Y\subset\mathbf{Z}^2$ by a structuring element W is defined as

$$Y\circ W \triangleq (Y\ominus W^s)\oplus W = \bigcup_{z\in\mathbf{Z}^2 | W_z\subseteq Y} W_z.$$

Similarly, the closing $Y\bullet W$ of a set $Y\subset\mathbf{Z}^2$ by a structuring element W is defined as

$$Y\bullet W \triangleq (Y\oplus W^s)\ominus W.$$

By duality of erosion/dilation, it follows that opening and closing are dual operators. Both can be viewed as nonlinear smoothing operators. Opening and closing are *idempotent (stable)* operators in the sense that $(Y\circ W)\circ W = Y\circ W$, and $(Y\bullet W)\bullet W = Y\bullet W$. A set Y is said to be (morphologically) *open* (*closed*) with respect to the structuring element W iff $Y\circ W = Y$ ($Y\bullet W = Y$). We shall say that a set Y is *smooth with respect to* W iff Y can be expressed as a union of shifted replicas of W. Y is open with respect to W iff Y is smooth with respect to W. Y is closed with respect to W iff Y^c is smooth with respect to W.

Let H be a convex[6] structuring element that contains the origin. In the discrete case, the notion of size of a convex structuring element can be formalized via the $\oplus$ operation. Let $\{\bar{0}\}$ denote the origin, and define

$$rH \triangleq \begin{cases} \{\bar{0}\}\oplus H\oplus H\oplus\cdots\oplus H, & (r \text{ times}), \quad r=1,2,\ldots \\ \{\bar{0}\}, & r=0 \end{cases}$$

Definition 4: Let Ψ be a generalized Bernoulli lattice process (or Bernoulli DRS or binary Bernoulli random field) on B constructively defined in the following manner: Each point $z\in B$ is contained in Ψ with probability $\lambda_s(z)$ independently of all others. Let $\{G_1, G_2, \ldots\}$ be a set of nonempty, convex i.i.d. DRS's on B, where each is given by $G_i = R_iH$, where $\{R_1, R_2, \ldots\}$ form an i.i.d. sequence of $\mathcal{Z}_+$ valued r.v.'s, which is independent of Ψ, and each R_i is distributed according to a pmf $f_R(r)$, which is compactly supported on $\{0,1,\ldots,\bar{R}\}$. Define

$$X = \bigcup_{i=1,2,\ldots} G_i\oplus\{y_i\}$$

where $\Psi = \{y_1, y_2, \ldots\}$. Then, X will be called a **discrete radial Boolean random set (DRBRS)** and will be denoted by (λ_s, H, f_R) DRBRS. The points $\{y_1, y_2, \ldots\}$ will be called *the germs*, and the DRS's $\{G_1, G_2, \ldots\}$ will be called *the primary grains* of the DRBRS X. The function λ_s will be called the *intensity function (or simply the intensity)* of both the DRBRS and the underlying Bernoulli lattice process.

[4] Again, we assume that $\Pr(X^c\cup N^c\subseteq K^c) > 0$, and this implies that $T_{X^c}(K) < 1$, and $T_{N^c}(K) < 1$. In addition, similar remarks hold here regarding the best choice of K^c.

[5] Here, we follow the original definitions of Serra [26]. In his work, the symbol $\oplus$ stands for Minkowski set addition, and the symbol $\ominus$ stands for Minkowski set subtraction.

[6] In digital topology [18], [26], [13], [12], the *convex hull* of a bounded set $H\subset\mathbf{Z}^2$ is defined as the intersection of the convex hull of H in the topology of $\mathcal{R}^2$, with $\mathbf{Z}^2$. A bounded set $H\subset\mathbf{Z}^2$ is *convex* if it is identical to its convex hull.

Remark: For brevity, we assume that for the purposes of this section the result of a $\oplus$ operation is automatically restricted to B. In addition, ^{c} stands for complement with respect to B. A sample realization of a DRBRS is given in Fig. 7.

In order to compute the capacity functional of a (λ_s, H, f_R) DRBRS, let us define

$$d^H(z, K) = \min_{k \in K} ||z - k||_H$$

where

$$||z - k||_H = \min\{n \geq 0 | (\{z\} \oplus nH) \cap \{k\} \neq \emptyset\}.$$

Observe that for $z \in K$, $d^H(z, K) = 0$ since H contains the origin. With this notation in place, and employing some geometric arguments, it can be shown [28] that

$$T_X(K) = 1 - \prod_{z \in K \oplus \check{R}H^s} \left[(1 - \lambda_s(z)) + \lambda_s(z) F_R(d^H(z, K) - 1)\right]$$

where

$$F_R(m) = \sum_{l=0}^{m} f_R(l)$$

and $F_R(-1) = 0$ by convention.

If we assume that both the signal X and the noise N can be modeled as DRBRS's (possibly with different structuring elements), then we can simply plug this expression into the formulas of propositions 1–3 and obtain the optimal filter as a function of the signal and noise parameter.

A simple simulation experiment is presented in Figs. 1 and 2. Fig. 1 depicts a realization of a DRBRS model ("the signal") of constant intensity and deterministic primary grain corrupted by i.i.d. union noise (an independent realization of a Bernoulli lattice process of constant intensity). Fig. 2 depicts the restored image, which was obtained by applying the optimal adaptive mask filter of proposition 2 to the noisy observation (K was taken to be the observation itself). To the trained eye, the restored image appears to be the morphological opening of the observed image using the primary grain of the signal as structuring element. This is, indeed, the case. Let $1 - q_X$, $1 - q_N$ be the intensities of the DRBRS and the i.i.d. union noise, respectively. It can be shown [28] that if $q_N \geq 2 - q_X^{-1}$, then, modulo edge effects, the optimal adaptive mask filter of proposition 2 is exactly a morphological opening using the primary grain of the signal as structuring element, that is, if the noise intensity is below a given threshold (which solely depends on the intensity of the signal), then the optimal adaptive mask filter is simply a morphological opening. This identification is important for two reasons. First, we will soon see (cf., Section V-A) that under our suppositions for the signal and the noise, the morphological opening is the maximum likelihood estimator of the signal based on the noisy observation. Second, the morphological opening is generally considered a good choice for the given filtering problem (e.g. see [15], [8], [7]).

This identification can be generalized as follows. Consider the case of two DRBRS's X, N of constant intensities and deterministic primary grains $1 - q_X$, $1 - q_N$, and H_X, H_N, respectively. It can be shown [28] that if

$$2 - q_N^{|H_N^s|} \leq q_X^{-1} (\Longleftrightarrow q_N^{|H_N^s|} \geq 2 - q_X^{-1})$$

then, modulo edge effects and taking K to be the observation itself, the optimal adaptive mask filter of proposition 2 reduces to the morphological opening of the observation by H_X, which is the (deterministic) primary grain of the signal DRBRS X. This is no longer guaranteed to be the ML estimator of X on the basis of the observation; however, it is widely believed to be a good estimator (e.g. see [15], [8], [7]. For example, if $|H_N| < |H_X|$ then opening by H_X will eliminate all instances of isolated noisy patterns.

An example is given in Figs. 3 and 4. Fig. 3 depicts a realization of the observable DRS $X \cup N$, where H_X, H_N were taken to be a discrete hexagon of size 12 and a discrete square of size 10, respectively. Fig. 4 depicts the opening of the DRS realization of Fig. 3 using H_X as structuring element.

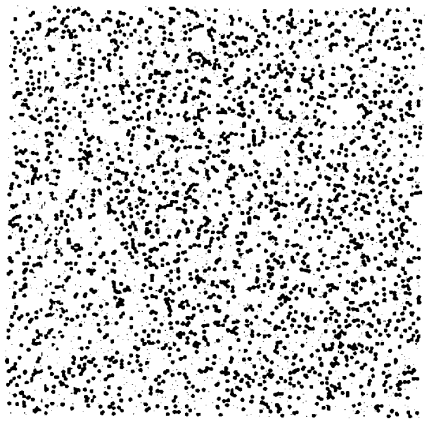


Fig. 1. Realization of a DRBRS corrupted by i.i.d. union noise.

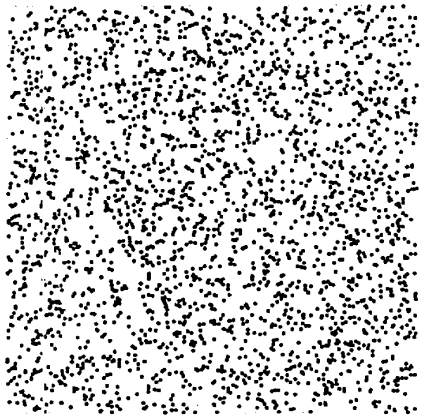


Fig. 2. Restored image obtained by using the optimal adaptive mask filter.

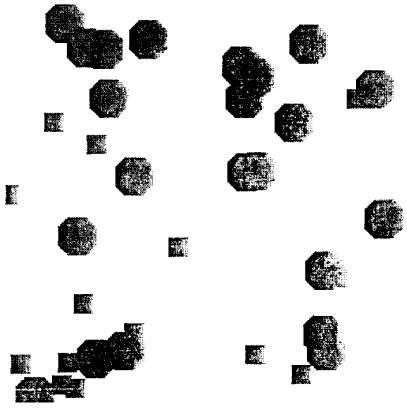


Fig. 3. Realization of the union of two DRBRS's.

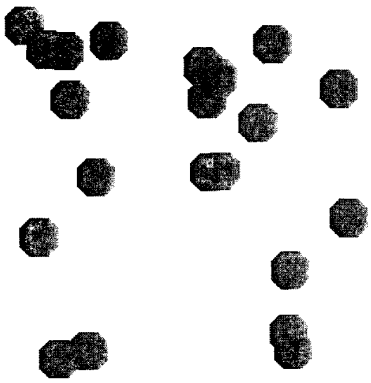


Fig. 4. Restored image, obtained byusing the optimal adaptive mask filter.

This somewhat surprising identification of the optimal adaptive mask filter of proposition 2 with the morphological opening filter is rather interesting. We have started with the objective of optimizing the statistical behavior of a mask filter structure and ended up with a morphological filter, which is the intuitively "obvious" choice from the viewpoint of syntactical optimization. This reflects the ability of the statistical optimization procedure to pick up the morphological structure of the signal and the noise, and, in effect, take both statistical and syntactical properties into consideration. This is the first example of such a joint optimization. We will see more of it as we move on. Note that this identification provides some independent corroborating evidence of the usefulness of optimal adaptive mask filters.

V. Optimal Morphological Filters

Complex morphological filters can be constructed by composing more elementary operators. For example, the family of alternating sequential filters (ASF's) is constructed by alternating openings and closings with structuring elements of increasing size.[7] One good reason for the widespread use of morphological filters is their excellent shape-preservation (syntactic) properties. Important characterizations (e.g., root signal structure) are well developed and easily understood [34], and this has helped build valuable intuition in the image processing community. Consequently, the empirical design of these filters has been greatly facilitated, and the resulting filters perform surprisingly well in a variety of noisy environments. However, with few exceptions [25], very little has been done in terms of "generic" DRS-theoretic optimization of morphological filters.

[7] See [27] for a recent survey of morphological filtering.

A. Some Results on Constrained Optimality or Why Morphology is Popular

Morphological filters are very flexible, mainly because of the freedom to choose the structuring element(s) to meet specified criteria. Among other things, morphological filters have been widely used to filter out certain kinds of impulsive noise, such as the so-called salt-and-pepper noise, in both binary and gray-scale images [25], [7], [10], [9], [5], [6], [34]. For example, it is widely believed that opening is suitable under a union noise model, whereas closing is suitable under an intersection noise model. ASF's are deemed appropriate under a combined union/intersection noise model. Indeed, these filters are used extensively, and they deliver adequate filtering in a variety of noisy environments. The natural question, then, is whether we can provide some sort of theoretical justification for their use. As it turns out, these filters are indeed optimal under a reasonable worst-case scenario. In particular, if we assume that the signal X is sufficiently smooth and the noise is i.i.d., then these operators provide the maximum a posteriori (MAP) estimate of X on the basis of the observation Y. For the rest of this subsection, we assume that structuring elements contain the origin. We have the following results.

Theorem 2: Let $O_W(B)$ denote the collection of all W open subsets of B. Assume that the signal DRS X on B induces the following probability mass function on $\Sigma(B)$:

$$P_X(X = K) = \begin{cases} \frac{1}{|O_W(B)|}, & \text{if } K \in O_W(B) \\ 0, & \text{otherwise} \end{cases}$$

where $||$ stands for set cardinality. Furthermore, assume that the observable DRS is $Y = X \cup N$, where N is a homogeneous Bernoulli lattice process of intensity $r \in [0, 1)$ (i.e., each point $z \in B$ is included in N with probability r independently of all other points), which is independent of X. Then, $Y \circ W$ is the unique MAP estimate of X on the basis of Y, regardless of the specific value of r.

Proof: Let $\widehat{X}_{\text{MAP}}(Y)$ denote the MAP estimate of X on the basis of Y. Then, by definition

$$\widehat{X}_{\text{MAP}}(Y) = \arg \max_{K \in \Sigma(B)} \{\Pr(X = K|Y)\}$$

Using Bayes' rule

$$\begin{aligned}\widehat{X}_{\mathrm{MAP}}(Y) &= \arg\max_{K\in\Sigma(B)}\{\Pr(Y|X=K)P_X(X=K)\}\\ &= \arg\max_{K\in O_W(B)}\left\{\Pr(Y|X=K)\frac{1}{|O_W(B)|}\right\}\\ &= \arg\max_{K\in O_W(B)}\{\Pr(Y|X=K)\}\\ &= \arg\max_{K\in O_W(B),K\subseteq Y}\{\Pr(Y|X=K)\}\\ &= \arg\max_{K\in O_W(B),K\subseteq Y}\left\{r^{|Y|-|K|}(1-r)^{|B|-|Y|}\right\}\\ &= \arg\max_{K\in O_W(B),K\subseteq Y}\left\{r^{-|K|}\right\}\\ &= \arg\max_{K\in O_W(B),K\subseteq Y}\{|K|\}.\end{aligned}$$

Therefore, $\widehat{X}_{\mathrm{MAP}}(Y)$ is the largest W-open subset of Y, which is by definition the opening of Y by W, i.e.

$$\widehat{X}_{\mathrm{MAP}}(Y) = Y \circ W$$

and the proof is complete. □

A little reflection on the above result is in order. First, observe that the proof *crucially* depends on $|B|$ being finite. Indeed, theorem 2, as well as the three theorems that follow, do not make sense when the lattice extends to infinity. Thus, a uniformly bounded discrete random set approach offers a fresh statistical perspective of morphological filtering, one that is not apparent within other formulations. The suppositions of the theorem indeed correspond to a worst-case statistical scenario: If all that is known about the signal is that it is almost surely (a.s.) smooth (open) with respect to W, then it is reasonable to model this knowledge using a uniform distribution over the set of all W-open subsets of B to reflect the fact that the signal exhibits no other (known) probabilistic structure. In addition, i.i.d. noise is the worst kind of noise in the sense of maximizing the Shannon entropy of the noise DRS N. Both these suppositions are plausible in practice, and this explains why the opening filter is successful under a union noise model. It is worth noting that the MAP estimate does not depend on the noise level r. Similarly, we have the following theorem.

Theorem 3: Let $C_W(B)$ denote the collection of all W-closed subsets of B. Assume that the signal DRS X on B induces the following probability mass function on $\Sigma(B)$

$$P_X(X=K) = \begin{cases}\frac{1}{|C_W(B)|}, & \text{if } K\in C_W(B)\\ 0, & \text{otherwise}\end{cases}$$

Furthermore, assume that the observable DRS is $Y = X \cap N$, where N is a homogeneous Bernoulli lattice process of intensity $r \in [0,1]$, which is independent of X. Then, $Y \bullet W$ is the unique MAP estimate of X on the basis of Y regardless of the specific value of r.

Proof: By definition

$$\begin{aligned}\widehat{X}_{\mathrm{MAP}}(Y) &= \arg\max_{K\in\Sigma(B)}\{\Pr(X=K|Y)\}\\ &= \arg\max_{K\in\Sigma(B)}\{\Pr(Y|X=K)P_X(X=K)\}\\ &= \arg\max_{K\in C_W(B)}\left\{\Pr(Y|X=K)\frac{1}{|C_W(B)|}\right\}\\ &= \arg\max_{K\in C_W(B)}\{\Pr(Y|X=K)\}\\ &= \arg\max_{K\in C_W(B),K\supseteq Y}\{\Pr(Y|X=K)\}\\ &= \arg\max_{K\in C_W(B),K\supseteq Y}\left\{r^{|Y|}(1-r)^{|K|-|Y|}\right\}\\ &= \arg\max_{K\in C_W(B),K\supseteq Y}\left\{(1-r)^{|K|}\right\}\\ &= \arg\min_{K\in C_W(B),K\supseteq Y}\{|K|\}.\end{aligned}$$

Therefore, $\widehat{X}_{\mathrm{MAP}}(Y)$ is the smallest W-closed superset of Y, which is, by definition, the closing of Y by W, i.e.

$$\widehat{X}_{\mathrm{MAP}}(Y) = Y \bullet W$$

and the proof is complete. □

The following two theorems are straightforward extensions of the above theorems. We state them here without proof.

Theorem 4: Let $O_{W_1,\ldots W_M}(B)$ denote the collection of all subsets K of B that can be written as

$$K = \bigcup_{i=1,\ldots,M} K_i, K_i \in O_{W_i}(B), i = 1,\ldots,M.$$

Assume that the signal DRS X on B induces the following probability mass function on $\Sigma(B)$:

$$P_X(X=K) = \begin{cases}\frac{1}{|O_{W_1,\ldots,W_M}(B)|}, & \text{if } K\in O_{W_1,\ldots,W_M}(B)\\ 0, & \text{otherwise}\end{cases}$$

Furthermore, assume that the observable DRS is $Y = X \cup N$, where N is a homogeneous Bernoulli lattice process of intensity $r \in [0,1)$, which is independent of X. Then,

$$\widehat{X}_{\mathrm{MAP}}(Y) = \bigcup_{i=1,\ldots,M} Y \circ W_i.$$

Theorem 5: Let $C_{W_1,\ldots W_M}(B)$ denote the collection of all subsets K of B that can be written as

$$K = \bigcap_{i=1,\ldots,M} K_i, K_i \in C_{W_i}(B), i = 1,\ldots,M.$$

Assume that the signal DRS X on B induces the following probability mass function on $\Sigma(B)$:

$$P_X(X=K) = \begin{cases}\frac{1}{|C_{W_1,\ldots,W_M}(B)|}, & \text{if } K\in C_{W_1,\ldots,W_M}(B)\\ 0, & \text{otherwise}\end{cases}$$

Furthermore, assume that the observable DRS is $Y = X \cap N$, where N is a homogeneous Bernoulli lattice process of intensity $r \in [0,1)$, which is independent of X. Then

$$\widehat{X}_{\mathrm{MAP}}(Y) = \bigcap_{i=1,\ldots,M} Y \bullet W_i.$$

A natural question that arises is what happens if we loose the uniform probability structure over the collection of smooth

realizations. The answer is that the MAP estimate will typically be intractable. However, we can still claim that the proposed estimate in any of the above theorems is the maximum likelihood (ML) estimate of X on the basis of Y. For example, a (λ_s, H, f_R)-DRBRS X, with $f_R(R = 0) = 0$, and primary grains that are properly contained in B, induces a pmf that satisfies $P_X(X = K) = 0$, iff $K \in \Sigma(B) \backslash O_H(B)$, but $P_X(X = K)$ is not uniform over $O_H(B)$. Nevertheless, we can still claim that, under the remaining assumptions of theorem 2, $Y \circ H$ is the ML estimate of X on the basis of Y.

In general, if we assume that X satisfies some arbitrary (not necessarily morphological) smoothness conditions, i.e., $X \in \mathcal{S}$, which is a class of smooth subsets of B, and that X is uniformly distributed over $\mathcal{S}$, then under an i.i.d. symmetric (binary symmetric channel (BSC)) noise model of pixel inversion probability $r < 0.5$, it is easy to see that

$$\hat{X}_{\mathrm{MAP}}(Y) = \arg \min_{K \in \mathcal{S}} d(Y, K)$$

where $d(Y, K)$ is the area of the symmetric set difference distance between Y and K. In other words, $\hat{X}_{\mathrm{MAP}}(Y)$ is the "projection" of the data Y onto $\mathcal{S}$. However, it is not clear how to compute this projection under general smoothness conditions. Furthermore, quite often, the noise is not i.i.d., and the signal is nonsmooth or only approximately smooth. The lack of a rigorous DRS-theoretic optimization approach for this general case has been evident in the literature. Our program is to develop such an approach. Specifically, for each degradation model, we will construct a suitable class of morphological operators, argue about its merits, and derive results that explicitly characterize the optimal choice of structuring element(s) in terms of the fundamental functionals of random set theory, namely, the generating functional of the signal and the generating functional of the noise.

B. Optimal Increasing, Shift-Invariant Filters with a Basis Constraint

A surprising result, which was originally due to Matheron [22] and subsequently improved on and used by Maragos [20], Dougherty *et al.*, and Giardina [7]–[10], is that a very large class of (linear and nonlinear) shift-invariant operations can be decomposed into a union of erosions by suitable structuring elements.

Let $E = \mathbf{Z}^2$, and let $\Sigma(E)$ denote the power set of E. Let $\Psi : \Sigma(E) \mapsto \Sigma(E)$. Recall that Ψ is *increasing* **iff** $X_1 \subseteq X_2 \Rightarrow \Psi(X_1) \subseteq \Psi(X_2), \forall X_1 \in \Sigma(E), X_2 \in \Sigma(E)$. We now reproduce some key theorems, which were taken from [20] and [22].

Theorem 6 [22]: For any *shift-invariant and increasing* mapping $\Psi : \Sigma(E) \mapsto \Sigma(E)$ and for all $X \in \Sigma(E)$

$$\Psi(X) = \bigcup_{W \in Ker(\Psi)} X \ominus W^s$$

where *the kernel of* Ψ, $Ker(\Psi)$ is defined as

$$Ker(\Psi) \triangleq \{W \in \Sigma(E) | \bar{0} \in \Psi(W)\}.$$

Theorem 7 [20]: For any *shift-invariant and increasing* mapping $\Psi : \Sigma(E) \mapsto \Sigma(E)$ and for all $X \in \Sigma(E)$

$$\Psi(X) = \bigcup_{W \in Bas(\Psi)} X \ominus W^s$$

where the *erosion basis* of Ψ, $Bas(\Psi)$, is defined as

$$Bas(\Psi) \triangleq \{W \in Ker(\Psi) | W' \in Ker(\Psi) \text{ and } W' \subseteq W \Rightarrow W' = W\}.$$

As a result of the latter theorem, the number of structuring elements that are needed for the decomposition is greatly reduced. Dougherty *et al.* [8]–[10] have made extensive use of this result to reduce the complexity associated with the design and implementation of optimal mean-square morphological filters. By duality, there exists an equivalent decomposition of any shift-invariant and increasing mapping as an intersection of dilations [5] over a *dilation basis.*

Strictly speaking, these theorems cannot be used with bounded domains. However, modulo some modifications that account for edge effects, they can be utilized. Then, the question of finding the optimal shift-invariant and increasing filter reduces to the problem of optimal basis design. This reduction is a significant one; the former problem is highly unstructured, whereas the latter admits a natural hierarchical decomposition in terms of a basis size constraint. In other words, we can consider a sequence of problems characterized by an increasing basis size. The performance of the optimal constrained filter is necessarily a nondecreasing function of the size of the basis. The upper bound on the size of the basis is usually determined by design and implementation considerations. However, under a basis size constraint, we are faced with an additional problem: Should we choose the expansion in terms of an erosion basis or in terms of a dilation basis? We will argue for the following point: *Under an intersection noise model, contrary to our intuition, we should think of the optimal filter as a union of erosions, whereas under a union noise model, we should think of the optimal filter as an intersection of dilations.* In both cases, we can work out theoretical formulas for the cost function.

C. Optimizing a Single Structuring Element

In the case of union noise, the simplest nontrivial expansion in terms of a dilation basis involves two structuring elements, one of which is constrained to be the origin. This is because we want the overall operation to be antiextensive, i.e., the output must be contained in the input. Dilation by the origin simply yields the input itself. Therefore, the simplest nontrivial class of constrained dilation basis filters for union noise can be written as follows:

$$\begin{aligned} f(Y) = f_W(Y) &= (Y \oplus W^s) \cap Y \\ &= [(X \cup N) \oplus W^s] \cap (X \cup N), \\ &\quad \text{for some structuring element, } W. \end{aligned}$$

Similarly, in the case of intersection noise, the simplest nontrivial expansion in terms of an erosion basis involves two structuring elements, one of which is constrained to be the

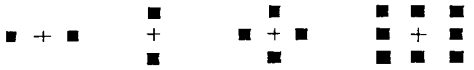


Fig. 5. Some structuring elements that can be used in a "gap-filling" mode.

(a) (b) (c)

Fig. 6. (a) Original image; (b) intersection of the image in (a) with a Bernoulli random field; (c) restored image.

origin (because we want the overall operation to be extensive, i.e., the output must contain the input). Again, since erosion by the origin yields the input itself, the simplest nontrivial class of constrained erosion basis filters for intersection noise can be written as follows:

$$f(Y) = f^W(Y) = (Y \ominus W^s) \cup Y$$
$$= [(X \cap N) \ominus W^s] \cup (X \cap N),$$
$$\text{for some structuring element, } W.$$

Some motivation is necessary at this point. Let us first consider the case of intersection noise. Intuitively, since the noise removes points from the signal, we should use some sort of "fill-in" operation to cancel the effect of noise. By definition

$$Y \ominus W^s = \{z | W_z \subseteq Y\}.$$

If the structuring element W is appropriately chosen (in particular, it must not contain the origin), then the erosion operation is a fill-in operation, i.e., it fills gaps in the "body" of the observation. However, it also introduces new gaps, which is an undesired side effect. Nevertheless, we can easily get rid of these "spurious" new gaps by simply taking the union of the resulting eroded set with the input set (i.e., the observation Y) itself. Some structuring elements that can be used in this mode are depicted in Fig. 5 (a cross indicates the location of the origin). An example is given in Fig. 6. Fig. 6(a) depicts a test image, whereas Fig. 6(b) depicts a degraded version of the test image, which is obtained by intersecting it with the set of points that make up a realization of a homogeneous Bernoulli random field of intensity 0.9. Fig. 6(c) depicts the estimate $\hat{X} = [(X \cap N) \ominus W^s] \cup (X \cap N)$, where X is the original test image depicted in Fig. 6(a), N is the set of points of the Bernoulli field, $X \cap N$ is the observation depicted in Fig. 6(b), and W is the leftmost of the structuring elements that appear in Fig. 5. For this example, the structuring element was not optimized.

In loose terms, if a structuring element does not contain *a neighborhood* of the origin, then it can be used in a gap-filling mode. The larger this neighborhood is, the wider the gaps that can be (partially) filled by an erosion with the given structuring element can be.

Similarly, by duality, if the structuring element is appropriately chosen (again, it must not contain the origin), dilation can remove points from the observation, and therefore, it can be appropriate under a union noise model. After performing a dilation with a suitably chosen structuring element, we take the intersection of the resulting set with the input (observation) set to eliminate points that have been introduced by the dilation operation.

This mode of use of the two basic morphological operations may seem strange at first since, for example (and partially because of its name), most people think of erosion as a shrink-type operation. However, one should keep in mind that this is only true if the erosion structuring element contains the origin. In fact, most people would consider using the operations in a reverse fashion: dilation for the case of intersection noise[8] and erosion for the case of union noise. The reason for our "unconventional" approach is that this way, we can take advantage of certain distributivity properties and obtain closed-form characterizations of the optimal filters.

Let us first consider intersection noise. Here,[9]

$$g(X, N) = X \cap N$$

and

$$\hat{X} = f(Y) = f^W(Y)$$
$$= (Y \ominus W^s) \cup Y = [(X \cap N) \ominus W^s] \cup (X \cap N),$$
$$\text{for some structuring element, } W \in \mathcal{W}$$

where $\mathcal{W}$ is the collection of structuring elements over which we intend to optimize. We need to make a small modification to our fidelity criterion in order to account for incomplete data close to the border of B. Towards this end, define

$$B \backslash \partial B = B \cap \left(\bigcap_{W \in \mathcal{W}} B \ominus W^s \right)$$

where $B \backslash \partial B$ is exactly the set of points $z \in B$ with the property that $W_z \subseteq B, \forall W \in \mathcal{W}$. Then, we only consider the total expected error restricted to $B \backslash \partial B$. We also assume that estimates of X are only valid within $B \backslash \partial B$. For brevity, we use the same symbol to denote a DRS and its restriction to $B \backslash \partial B$. The meaning is clear from context. We have the following proposition.

Proposition 4: Under the assumption of mutual independence of the signal and noise DRS's X, N, the value of the expected error $E(e) = Ed(X, \hat{X})$ incurred when X is estimated by $\hat{X} = [(X \cap N) \ominus W^s] \cup (X \cap N)$ is given by

$$E(e) = \sum_{z \in B \backslash \partial B} \{Q_{X^c}(\{z\})(1 - Q_{N^c}(\{z\}))$$
$$+ Q_{N^c}(W_z)(Q_{X^c}(W_z) - Q_{X^c}(\{z\} \cup W_z))$$
$$+ Q_{X^c}(\{z\} \cup W_z)(Q_{N^c}(\{z\} \cup W_z) - Q_{N^c}(W_z))\}.$$

[8] See [16] for an account of such an approach, when the intersection mask N is a deterministic, regularly spaced grid, which undersamples the observation.

[9] This operation can be viewed as *random sampling* the DRS X. In this context, our results characterize the optimal (within a class) morphological reconstruction filter for DRS's that have undergone random sampling.

Proof: **See the Appendix.**

The structuring element W should be chosen to minimize this expression. Observe that the total expected error is equal to the sum of the probabilities of individual pixel errors. If we make the natural[10] assumption that both X and N are obtained by sampling *stationary* random sets [22], then all the functionals in the above sum are independent of the location $\{z\}$, and we obtain the following result.

Corollary 1: Under the condition of mutual independence of the signal and noise DRS's X, N, assuming that X, N are obtained by sampling stationary random sets and that X is estimated by $\widehat{X} = [(X \cap N) \ominus W^s] \cup (X \cap N)$, the optimal choice of the structuring element W is the one that minimizes the *probability of pixel error*

$$\begin{aligned} P_{\text{pixel error}}(W) &= Q_{X^c}(\{\bar{0}\})(1 - Q_{N^c}(\{\bar{0}\})) \\ &\quad + Q_{N^c}(W)(Q_{X^c}(W) - Q_{X^c}(\{\bar{0}\} \cup W)) \\ &\quad - Q_{X^c}(\{\bar{0}\} \cup W) \\ &\qquad \times (Q_{N^c}(W) - Q_{N^c}(\{\bar{0}\} \cup W)). \end{aligned}$$

Let us examine the individual terms of this sum. The first term $Q_{X^c}(\{\bar{0}\})(1 - Q_{N^c}(\{\bar{0}\}))$ of the probability of pixel error $P_{\text{pixel error}}(W)$ is exactly the probability of pixel error between the signal X and the observation $X \cap N$ (this can be seen by setting $W = \{\bar{0}\}$, which corresponds to no filtering of the observation). This first term is independent of W, and therefore, it is not under our control. The remaining two terms of the sum are both nonnegative functions of W (it can be easily shown that the generating functional of an arbitrary DRS is constrained to be decreasing). When considered as a function of W, this sum clearly brings out the interplay between "signal power" and "noise power" and how it determines the structuring element that achieves the optimal tradeoff between eliminating gaps introduced by noise and retaining gaps that are present in the signal itself.

Some notes on the applicability of this result are in order. If the generating functionals $Q_{X^c}(\cdot)$, $Q_{N^c}(\cdot)$ (or, equivalently, the capacity functionals $T_{X^c}(\cdot)$, $T_{N^c}(\cdot)$) are given, then optimization of W over a relatively small collection of allowable $W's$ is straightforward. In general, for large collections of candidate structuring elements, some sort of suboptimal search must be pursued to avoid a potentially difficult exhaustive search. See [19] for an "expert" structuring element library design approach. We shall return to this point later on. At any rate, even if the generating functionals are not available (which is the case in most applications), all the quantities that are relevant to our optimization problem can be efficiently and accurately estimated from running (sample) averages by virtue of stationarity and the law of large numbers. For example, $Q_{X^c}(W)$ can be estimated by "sliding" the structuring element W across a realization of X^c and counting the number of times that the two have an empty intersection and similarly for the others.

Let us now turn to union noise. Here

$$g(X, N) = X \cup N$$

and $\widehat{X} = f(Y) = f_W(Y) = (Y \oplus W^s) \cap Y = [(X \cup N) \oplus W^s] \cap (X \cup N)$, *for some structuring element* $W \in \mathcal{W}$. As expected, we can once more resort to duality. In particular, since

$$\begin{aligned} (\widehat{X})^c &= ([(X \cup N) \oplus W^s] \cap (X \cup N))^c \\ &= [(X \cup N) \oplus W^s]^c \cup (X \cup N)^c \\ &= [(X \cup N)^c \ominus W^s] \cup (X^c \cap N^c) \\ &= [(X^c \cap N^c) \ominus W^s] \cup (X^c \cap N^c) \end{aligned}$$

and

$$d(X_1, X_2) = d(X_1^c, X_2^c)$$

we can easily reduce this case to the case of the previous subsection by replacing X, N by their complements X^c, N^c. Thus we have the following result.

Proposition 5: Under the assumption of mutual independence of the signal and noise DRS's X, N, the value of the expected error $E(e) = E[\mathrm{d}(X, \widehat{X})]$ incurred when X is estimated by $\widehat{X} = [(X \cup N) \oplus W^s] \cap (X \cup N)$ is given by

$$\begin{aligned} E(e) = \sum_{z \in B \setminus \partial B} &\{Q_X(\{z\})(1 - Q_N(\{z\})) \\ &+ Q_N(W_z)(Q_X(W_z) - Q_X(\{z\} \cup W_z)) \\ &+ Q_X(\{z\} \cup W_z)(Q_N(\{z\} \cup W_z) - Q_N(W_z))\}. \end{aligned}$$

Again, if we make the assumption that both X and N are obtained by sampling stationary random sets, then we obtain the following result.

Corollary 2: Under the condition of mutual independence of the signal and noise DRS's X, N, assuming that X, N are obtained by sampling stationary random sets and that X is estimated by $\widehat{X} = [(X \cup N) \oplus W^s] \cap (X \cup N)$, the optimal choice of the structuring element W is the one that minimizes the *probability of pixel error*

$$\begin{aligned} P_{\text{pixel error}}(W) &= Q_X(\{\bar{0}\})(1 - Q_N(\{\bar{0}\})) \\ &\quad + Q_N(W)(Q_X(W) - Q_X(\{\bar{0}\} \cup W)) \\ &\quad - Q_X(\{\bar{0}\} \cup W) \\ &\qquad \times (Q_N(W) - Q_N(\{\bar{0}\} \cup W)). \end{aligned}$$

As in the case of intersection noise, similar remarks hold here regarding the interpretation of the individual terms of the sum. Again, if the generating functionals $Q_X(\cdot), Q_N(\cdot)$ are given, then optimization over a small collection of candidate structuring elements is straightforward. If these functionals are not available, their values can be estimated from running averages as before.

Let us now show how one can reduce the complexity of the search for the optimal structuring element by assuming that the signal DRS X is smooth (i.e., morphologically open) with respect to some structuring element. We will need the following.

Definition 5: A DRS X is H-open iff[11]

$$P_X(X = K) = P_{X \circ H}(X \circ H = K), \forall K \in \Sigma(B).$$

[10] This is because we are using a *shift-invariant* filtering operation.

[11] Observe that this definition asserts that X is H open iff $X \circ H = X$ in

Lemma 2: X is H-open iff $Q_X(K) = Q_{X\oplus H^s}(K \oplus H^s), \forall K \in \Sigma(B)$.

Proof: **See the Appendix.**

Therefore, now let us assume that the signal DRS X is H-open, where H is convex and contains the origin. Let $\mathcal{W}$ denote the collection of candidate structuring elements over which we intend to optimize. Consider the second term of the sum for the probability of pixel error. Using the above lemma,

$$Q_X(W) - Q_X(\{\bar{0}\} \cup W) = Q_{X\ominus H^s}(W \oplus H^s) - Q_{X\ominus H^s}((\{\bar{0}\} \cup W) \oplus H^s).$$

By distributivity of dilation over union

$$Q_{X\ominus H^s}((\{\bar{0}\} \cup W) \oplus H^s) = Q_{X\ominus H^s}((\{\bar{0}\} \oplus H^s) \cup (W \oplus H^s)) = Q_{X\ominus H^s}(H^s \cup (W \oplus H^s)).$$

Thus, under the condition

$$H^s \subseteq W \oplus H^s, \quad \forall W \in \mathcal{W}$$

the second term of the sum for the probability of pixel error is zero. In loose terms, this condition amounts to requiring H to be "large enough" relative to the structuring elements in $\mathcal{W}$. Since the signal is usually associated with the more prominent patterns in the image, this requirement is not very restrictive. For example, if $\mathcal{W}$ is the collection of structuring elements depicted in Fig. 5 and H is a square of side 3 pixels centered at the origin, then the above condition is satisfied. Therefore, the optimal $W \in \mathcal{W}$ should maximize the third term of the sum, namely

$$G(W) \triangleq Q_X(\{\bar{0}\} \cup W)(Q_N(W) - Q_N(\{\bar{0}\} \cup W))$$

Now[12]

$$(Q_N(W) - Q_N(\{\bar{0}\} \cup W)) = \Gamma_{N,1}(W; \{\bar{0}\}) = P_N(N \cap W = \emptyset, N \cap \{\bar{0}\} \neq \emptyset)$$

which is clearly decreasing in W. Furthermore, $Q_X(\{\bar{0}\} \cup W)$ is decreasing in W. Thus, $G(W)$ is decreasing in W, i.e.

$$W_1 \subseteq W_2 \Rightarrow G(W_2) \leq G(W_1).$$

Hence, since we seek to maximize $G(W)$, we can eliminate from consideration all structuring elements in $\mathcal{W}$ that properly contain other structuring elements in $\mathcal{W}$, i.e., it suffices to optimize over the subcollection

$$\widetilde{\mathcal{W}} = \{W \in \mathcal{W} \mid W' \in \mathcal{W} \text{ and } W' \subseteq W \Rightarrow W' = W\}.$$

Thus, we have proven the following.

Corollary 3: Under the condition of mutual independence of the signal and noise DRS's X, N, assuming that X, N are obtained by sampling stationary random sets, X is H-open, where H is convex, containing the origin, and such that $H^s \subseteq W \oplus H^s, \forall W \in \mathcal{W}$ and that X is estimated by $\widehat{X} = [(X \cup N) \oplus W^s] \cap (X \cup N)$, the optimal structuring element is

$$[W^* = \underset{W\in\mathcal{W}}{arg\ min} P_{pixel\ error}(W) = \underset{W\in\tilde{\mathcal{W}}}{arg\ max} G(W).$$

This elimination can translate to a significant reduction in search complexity. For example, if $\mathcal{W}$ is the collection of structuring elements depicted in Fig. 5, and H is a square of side 3 pixels that is centered at the origin, then it suffices to optimize over the two left-most structuring elements. By duality, a similar reduction can be achieved under an intersection noise model if we assume that X^c is H-open, i.e., that X is H-closed.

D. Multiple Structuring Elements

In certain situations, particularly when the noise level is high, a single erosion followed by a union, even if optimal, may not suffice to properly reconstruct the signal. In this case, it is beneficial to consider larger bases, i.e., filters with multiple structuring elements. The structuring elements must be *jointly* optimized to eliminate a wider class of error patterns. Using exactly the same algebraic methods as in the proof of proposition 4, and with some patience, we can obtain similar optimality results for the case of multiple structuring elements. For example, we state the following proposition (see [28] for a proof).

Proposition 6: Under the assumption of mutual independence of the signal and noise DRS's X, N, the value of the expected error $E(e) = E[\mathrm{d}(X, \widehat{X})]$ incurred when X is estimated by

$$\widehat{X} = \left[(X \cap N) \ominus (W^1)^s\right] \cup \left[(X \cap N) \ominus (W^2)^s\right] \cup (X \cap N)$$

is given by

$$\begin{aligned} E(e) = \sum_{z \in B \setminus \partial B} \{ & Q_{X^c}(\{z\})(1 - Q_{N^c}(\{z\})) \\ & + Q_{X^c}(W_z^1) Q_{N^c}(W_z^1) + Q_{X^c}(W_z^2) Q_{N^c}(W_z^2) \\ & + Q_{X^c}(\{z\} \cup W_z^1) Q_{N^c}(\{z\} \cup W_z^1) \\ & \quad + Q_{X^c}(\{z\} \cup W_z^2) \times Q_{N^c}(\{z\} \cup W_z^2) \\ & - 2Q_{X^c}(\{z\} \cup W_z^1) Q_{N^c}(W_z^1) \\ & \quad - 2Q_{X^c}(\{z\} \cup W_z^2) Q_{N^c}(W_z^2) \\ & - Q_{X^c}(W_z^1 \cup W_z^2) Q_{N^c}(W_z^1 \cup W_z^2) \\ & - Q_{X^c}(\{z\} \cup W_z^1 \cup W_z^2) Q_{N^c}(\{z\} \cup W_z^1 \cup W_z^2) \\ & + 2Q_{X^c}(\{z\} \cup W_z^1 \cup W_z^2) Q_{N^c}(W_z^1 \cup W_z^2) \}. \end{aligned}$$

Again, by assuming that X, N are obtained by sampling stationary random sets, we can obtain a characterization of the optimal pair of structuring elements in terms of the probability of pixel error. Under appropriate smoothness conditions, we can reduce the complexity of the search for the optimal pair in a manner similar to the one of the previous subsection.

the sense of distributions. However, this implies that

$$P_X(X \circ H \neq X) = \sum_{K \circ H \neq K} P_X(X = K) = \sum_{K \circ H \neq K} P_{X \circ H}(X \circ H = K) = 0$$

i.e., $P_X(X \circ H = X) = 1$, which implies $X \circ H = X, P_X - a.s.$ Thus, we do not make any distinction.

[12] For the definition of the functional Γ, see the proof of proposition 4 in the Appendix.

TABLE I
ESTIMATED VALUES OF THE GENERATING FUNCTIONAL $Q_{X^c}(\cdot)$

	$W = \{\bar{0}\}$	$W = W_1$	$W = W_2$	$W = W_3$	$W = W_4$
$Q_{X^c}(W)$	0.494698	0.438927	0.438970	0.426667	0.383403
$Q_{X^c}(\{\bar{0}\} \cup W)$		0.437678	0.437682	0.426082	0.383403

TABLE II
ESTIMATED VALUES OF THE GENERATINGFUNCTIONAL $Q_{N^c}(\cdot)$

	$W = \{\bar{0}\}$	$W = W_1$	$W = W_2$	$W = W_3$	$W = W_4$
$Q_{N^c}(W)$	0.900431	0.810915	0.810519	0.657578	0.432145
$Q_{N^c}(\{\bar{0}\} \cup W)$		0.730154	0.729550	0.591795	0.389004

TABLE III
ESTIMATED VALUES OF THE PROBABILITY OF PIXEL ERROR

	$W = \{\bar{0}\}$	$W = W_1$	$W = W_2$	$W = W_3$	$W = W_4$
$P_{pixel\ error}(W)$	0.0493	0.01491	0.01490	0.0217	0.0328

The details are straightforward but cumbersome. Obviously, by duality, similar results can be obtained for the case of union noise as well as for more than two structuring elements.

E. Experimental Results

In order to corroborate our theoretical results, we have designed a series of simulation experiments. One such experiment is described here in detail. The results of another experiment involving a real-life image are also presented. These experiments are solely intended to serve as "proof of concept." No claims are made regarding the relative merit of our approach as measured against other approaches in the literature. A thorough comparative study of different filter structures is analytically difficult,[13] and it therefore requires extensive simulation, which is beyond our present scope. Our purpose here is to demonstrate that our theoretical results actually make sense in practice.

Let us make the assumptions of Corollary 1. For the purposes of simulation, we need models for the signal and noise. We assume that the signal X is a DRBRS of constant intensity and that the noise N is given by the set of points of a Bernoulli random field of constant intensity $p = 0.9$. These models are only used to generate realizations of the signal, the noise, and the observation. The entire simulation is data driven, and all relevant probabilities are estimated using running averages. This approach is honest and close to real-world problems.

Given the scope of these experiments, it seems appropriate, for pedagogical purposes, to limit ourselves to relatively few structuring elements in order to avoid unnecessary complication and help the reader gain intuition. It is much easier to convince ourselves about the relative merits of different structuring elements when we are working with just a handfull of them, rather than being faced with a long list of candidate structuring elements. Let us therefore consider the collection depicted in Fig. 5 and label the structuring elements $W_1, \ldots, W_4$ from left to right.

[13] Cf. the section that immediately follows.

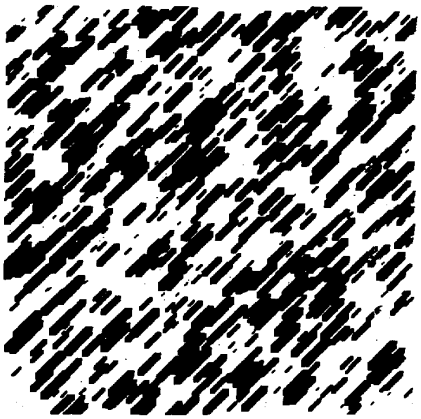


Fig. 7. Realization of the signal DRS X.

Fig. 7 depicts a realization of the signal DRS X, whereas Fig. 8 depicts a realization of the noise DRS N. These are solely used to estimate the relevant probabilities. The results for the signal and the noise are tabulated in Tables I and II, respectively. The results for the estimated probability of pixel error are tabulated in Table III. These have been computed using Tables I and II and the formula of corollary 1. In Table III, the left-most entry is the estimated probability of pixel error between the signal X and the observation $X \cap N$, i.e., when no filtering takes place (this corresponds to $W = \{\bar{0}\}$). It is given here for comparison purposes. Clearly, the optimal structuring element is W_2, with W_1 running a close second (this is justified by the symmetry in the data). The worst structuring element is W_4.

Fig. 9 depicts another (independent) realization of X, whereas Fig. 10 depicts a realization of the observation $Y = X \cap N$ obtained by intersecting the DRS realization of Fig. 9 with an independent realization of N. Fig. 11 depicts the restored image $\hat{X} = (Y \ominus W_2^s) \cup Y$, where Y is the DRS realization of Fig. 10. This is the best possible restoration within the given family of structuring elements. For comparison purposes, Fig. 12 depicts the restored image, $\hat{X} = (Y \ominus W_4^s) \cup$

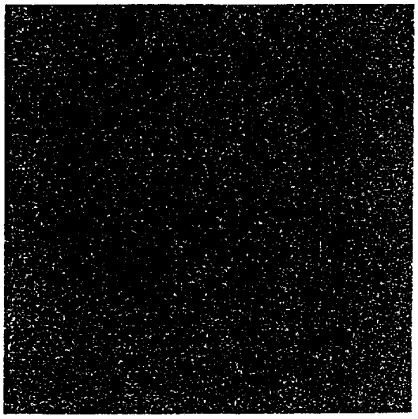


Fig. 8. Realization of the noise DRS N.

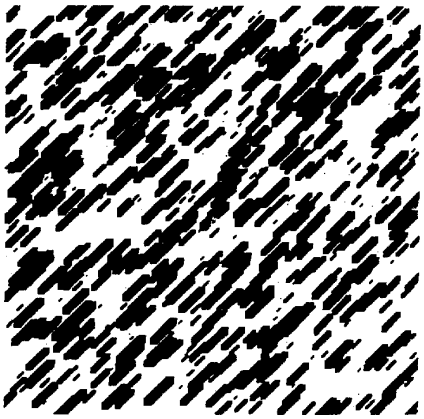


Fig. 9. Another (independent) realization of the signal DRS X.

Y, where Y is the DRS realization of Fig. 10. This is the worst (nontrivial) restoration within the given family of structuring elements. Close inspection of these figures reveals several interesting phenomena. In particular, even though W_2 does a better job than W_4 in filling up gaps introduced by noise, it also bridges together signal components which were originally disconnected. This is evident in the upper right-hand part of the figures. Nevertheless, this source of error is counterbalanced by the relative effectiveness of W_2 in terms of noise elimination. As a result, the overall quality of restoration achieved by W_2 is still visibly better. However, under a low-noise scenario, this situation can be reversed, i.e., retaining the connectivity structure of the signal will become more important, and eventually, it will supersede noise elimination as the dominant factor. In this case, W_4 will provide superior performance.

These simulation results are encouraging; they clearly support the theory and satisfy our intuition. Furthermore, considering the fact that the optimal filter essentially consists of two set translations and two set unions (two translations and one union are needed to implement the erosion with W_2), the quality of restoration seems good. Even better results can be achieved using multiple structuring elements.

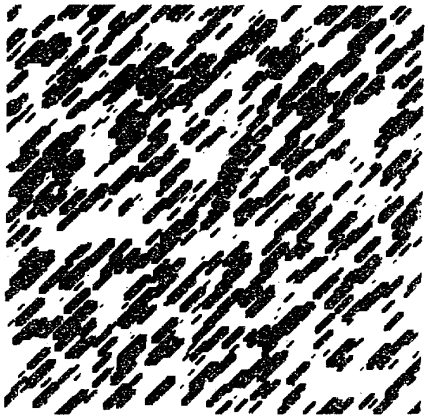


Fig. 10. Result of intersecting the DRS realization of Fig. 9 with another (independent) realization of the noise DRS N.

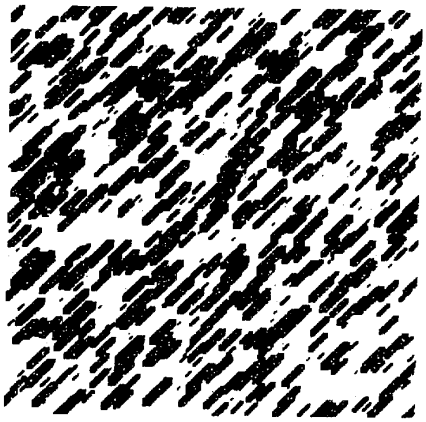


Fig. 11. Restored image obtained by filtering the DRS realization of Fig. 10 using structuring element W_2 (the best one).

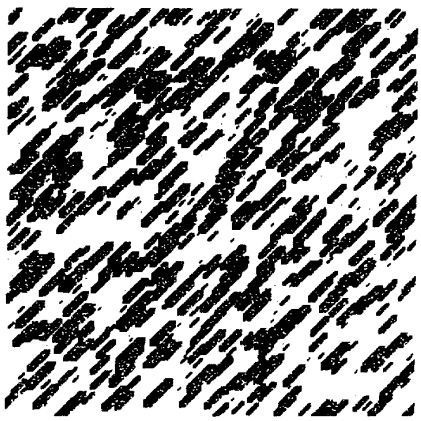


Fig. 12. Restored image obtained by filtering the DRS realization of Fig. 10 using structuring element W_4 (the worst one).

In real life, image statistics are often spatially varying. In most cases, it is possible to model such images as piecewise statistically invariant. This approach is taken quite often when

Fig. 13. Binary Lena picture.

Fig. 14. Lena picture, degraded by a combination of burst and memoryless transmission errors.

Markov random field (MRF) models are used. In this setting, one can either segment the image in disjoint regions with approximately invariant statistics *prior* to the filtering step or rely on the filtering algorithm to perform simultaneous signal estimation *and* segmentation. Either way, this is not a trivial task, and these algorithms are generally not amenable to in-depth statistical analysis. Therefore, it is of interest to test our algorithms in images that violate our assumptions. For this purpose, consider Fig. 13. It depicts a binary version of the well-known "Lena" picture. This picture can be modeled as piecewise statistically invariant. However, let us bypass the segmentation step and blindly apply our algorithm. Fig. 14 depicts a version of Lena that has been degraded by a combination of burst and memoryless transmission errors. Specifically, we assume that the image is scanned row-wise, and individual bits are transmitted over a channel that is memoryless most of the time but occasionally switches to a burst error channel. We assume that the noise only affects the white part of the image, i.e., a union noise model. The signal and noise statistics were estimated from the original picture and an independent realization of the noise. Then, based on the "dual" of proposition 6, as it applies to the case of sampling stationary random sets, the optimal two-fold dilation filter was sought within the class of filters with structuring elements in the collection of Fig. 5. The optimal pair of structuring elements was found to be (W_2, W_4). The restored image is depicted in Fig. 15. Given that we have violated the stationarity assumption, the overall result is surprisingly good.

Fig. 15. Restored Lena picture.

Conclusion and Further Research

In this paper, we have described two optimal digital binary image filtering strategies. Mask filtering is a natural approach to the problem of digital binary image restoration under a union/intersection degradation model. We have discussed both optimal fixed mask filtering and optimal adaptive mask filtering. Although adaptive mask filtering is superior, it essentially requires knowledge of the capacity functionals of the signal and noise. This is the case when both the signal X and the noise N can be modeled as DRBRS's. On the other hand, fixed mask filtering only requires knowledge of first-order statistics (pixel hitting probabilities), which can be easily and accurately estimated from training data. Therefore, it provides a simple and robust alternative when the signal and noise processes are not known in detail.

In the second part of this work, we have demonstrated that certain popular morphological filtering schemes are indeed optimal under some fairly plausible assumptions. We have also described a general optimal morphological binary image filtering approach, which is more appropriate when the signal and noise DRS's exhibit a statistical behavior that is spatially invariant. We have demonstrated that by choosing the right expansion of the optimal filter, namely, as a union of erosions (intersection of dilations), under an intersection (union) noise model, we can obtain universal optimal filtering results, which do not rely on strong assumptions concerning the nature of the signal and noise, and the mode of their spatial interaction. In particular, they are valid when the signal and noise patterns are spatially overlapping. This situation contrasts with the optimality results of Haralick *et al.*[15], which are based on the assumption that the signal and noise patterns are "noninterfering," and the results of Schonfeld and Goutsias

[25], which rely on strong separability of the noise patterns. In contrast with the aforementioned model-based approaches, we have chosen to avoid restricting the class of input signals under consideration. Obviously, a model-based approach is superior when the underlying assumptions are justified in practice. However, if this is not the case, then our approach may prove safer.

An important open research problem involves the following question: How much better can we do using $n+1$ structuring elements as compared to using n structuring elements? In other words, how fast does the minimum expected error converge? This is directly related to the rate of convergence of the basis expansion of increasing, shift-invariant filters in terms of morphological erosions/dilations. As of now, the answer to this question remains largely unknown due to fundamental analytical difficulties. Certainly, with $n+1$ structuring elements, we can do no worse than with n structuring elements. It currently seems that the best way to choose n is by trial and error. Note that computational complexity considerations would normally dictate an upper bound on n. Thus, the tradeoff is in terms of improvement in expected error versus increase in (run-time and design) computational complexity. In related experimental approaches [19], this compromise leads to a relatively small number of structuring elements.

It would also be interesting to do direct comparisons with other filtering approaches. Theorems 6, 7 (Matheron, Maragos *et al.*, Dougherty *et al.*) provide the basis for such a comparison. Theorem 7 states that every increasing, shift-invariant filter can be expressed as a sup-sum of erosions over a suitable basis set of structuring elements. Thus, this class contains linear shift-invariant averagers, order statistics (e.g., median), stack filters, etc. However, the length of the required expansion for each such filter is unknown, and it can be infinite. The effect of blocking the length of the basis is yet unknown. The following question is central to this problem: How large a class of increasing, shift-invariant filters can we span with n structuring elements? Again, things reduce to questions of rate of convergence. We intend to investigate these questions in future work.

Finally, It is important to keep in mind that our results can be extended to finite-gray-level digital images of compact support and sup/inf noise via threshold decomposition of functions and/or by treating functions as sets via their umbrae. In this case, $B \subset \mathbf{Z}^3$. Although this extension does not pose any additional theoretical problems, it warrants proper attention to certain complexity issues involved. In particular, the probability structure induced on the measurable space $(\Sigma(B), \Sigma(\Sigma(B)))$, $B \subset \mathbf{Z}^3$ by treating random finite-gray-level digital images of compact support via their umbrae is not arbitrary due to the constraints imposed by the fact that function umbrae are not arbitrary subsets of B. These constraints must be (implicitly or explicitly) incorporated into the probability measures, and this complicates statistical modeling. In addition, from the viewpoint of computational complexity, it is important to understand under which conditions function processing via threshold decomposition and/or function umbrae computationally efficient are relative to standard function processing techniques. These issues are currently under investigation.

APPENDIX
COLLECTED PROOFS OF LEMMAS AND PROPOSITIONS

Proof of Lemma 1— Uniqueness: Assume that the external decomposition formula holds. Look at the right-hand side of the inversion formula.

$$\begin{aligned}\sum_{C\subseteq S}(-1)^{|C|}v(S^c\cup C) &= \sum_{C\subseteq S}(-1)^{|C|}\sum_{D\subseteq S\cap C^c}u(D)\\ &= \sum_{C\subseteq S}(-1)^{|C|}\sum_{D\subseteq S\setminus C}u(D) = \sum_{C\subseteq S}\sum_{D\subseteq S\setminus C}(-1)^{|C|}u(D)\\ &= \sum_{D\subseteq S}\sum_{C\subseteq S\setminus D}(-1)^{|C|}u(D)\\ &= \sum_{D\subseteq S}u(D)\sum_{C\subseteq S\setminus D}(-1)^{|C|} = u(S).\end{aligned}$$

Since

$$\sum_{C\subseteq S}(-1)^{|C|} = \begin{cases}0, & S\neq\emptyset\\ 1, & S=\emptyset\end{cases}$$

Existence: Assume that the inversion formula holds, and look at the right-hand side of the external decomposition formula.

$$\begin{aligned}\sum_{S\subseteq A^c}u(S) &= \sum_{S\subseteq A^c}\sum_{C\subseteq S}(-1)^{|C|}v(S^c\cup C)\\ &= \sum_{S\subseteq A^c}\sum_{C\subseteq S}(-1)^{|C|}v((S\setminus C)^c)\\ &= \sum_{D\subseteq A^c}\sum_{C\subseteq A^c\setminus D}(-1)^{|C|}v(D^c)\\ &= \sum_{D\subseteq A^c}v(D^c)\sum_{C\subseteq A^c\setminus D}(-1)^{|C|} = v((A^c)^c) = v(A)\end{aligned}$$

as for the uniqueness part. □

Proof of Proposition 1: Without loss of generality, we may assume that $W_1 \subseteq W_2$ since it makes no sense removing points from the observation only to reinstate them at the next filtering step. After some manipulation

$$\begin{aligned}E(e) = \mathrm{E}[\mathrm{d}(X,\hat{X})] = E|X\cap(N_1^c\cup W_2^c)\cap[(N_2^c\cap W_1^c)\cup W_2^c]|\\ + E|(X^c\cap N_2\cap W_2)\cup(X^c\cap W_1)|.\end{aligned}$$

A crucial observation here is that

$$\begin{aligned}E|X| &= E\sum_{z\in B}1(z\in X) = \sum_{z\in B}E1(z\in X)\\ &= \sum_{z\in B}\Pr(z\in X) = \sum_{z\in B}\Pr(X\cap\{z\}\neq\emptyset)\\ &= \sum_{z\in B}T_X(\{z\}).\end{aligned}$$

Consider the first term of the expected error

$$
\begin{aligned}
&E|X \cap (N_1^c \cup W_2^c) \cap [(N_2^c \cap W_1^c) \cup W_2^c]| \\
&\quad = \sum_{z \in B} \Pr(z \in X \cap (N_1^c \cup W_2^c) \cap [(N_2^c \cap W_1^c) \cup W_2^c]) \\
&\text{(by indep. of } X, N_1, N_2) \\
&\quad = \sum_{z \in B} \Pr(z \in X) \Pr(z \in N_1^c \cup W_2^c) \\
&\qquad \times \Pr(z \in (N_2^c \cap W_1^c) \cup W_2^c) \\
&\quad = \sum_{z \in B} T_X(\{z\}) \\
&\qquad \times [1(z \in W_2^c) + 1(z \in W_2)(1 - T_{N_1}(\{z\}))] \\
&\qquad \times [1(z \in W_2^c) + 1(z \in W_2) 1(z \in W_1^c)(1 - T_{N_2}(\{z\}))].
\end{aligned}
$$

Next, consider the second term of the expected error

$$
\begin{aligned}
&E|(X^c \cap N_2 \cap W_2) \cup (X^c \cap W_1)| \\
&\quad = \sum_{z \in B} \Pr(z \in (X^c \cap N_2 \cap W_2) \cup (X^c \cap W_1)) \\
&\quad = \sum_{z \in B} \Pr(z \in X^c \cap [(N_2 \cap W_2) \cup W_1]) \\
&\quad = \sum_{z \in B} \Pr(z \in X^c, z \in (N_2 \cap W_2) \cup W_1) \\
&\text{(by independence of } X, N_2) \\
&\quad = \sum_{z \in B} \Pr(z \in X^c) \Pr(z \in (N_2 \cap W_2) \cup W_1) \\
&\quad = \sum_{z \in B} (1 - T_X(\{z\})) \\
&\quad \times [1(z \in W_1) + 1(z \in W_1^c) 1(z \in W_2) T_{N_2}(\{z\})].
\end{aligned}
$$

Therefore, the overall expression for the expected cost becomes

$$
\begin{aligned}
&E(e) = \mathrm{E}[\mathrm{d}(X, \hat{X})] \\
&\quad = \sum_{z \in B} \{T_X(\{z\})[1(z \in W_2^c) + 1(z \in W_2)(1 - T_{N_1}(\{z\}))] \\
&[1(z \in W_2^c) + 1(z \in W_2) 1(z \in W_1^c)(1 - T_{N_2}(\{z\}))] \\
&+ (1 - T_X(\{z\}))[1(z \in W_1) + 1(z \in W_1^c) 1(z \in W_2) T_{N_2}(\{z\})]\}.
\end{aligned}
$$

Consider the term in curly braces. As we have mentioned before, $W_1 \subseteq W_2$. Therefore, for each $z \in B$, we have the following three choices:

$$
\begin{aligned}
&\text{i) } z \in W_1^c, z \in W_2^c, \text{ or ii) } z \in W_1^c, z \in W_2, \text{ or} \\
&\text{iii) } z \in W_1, z \in W_2
\end{aligned}
$$

In case i), the term in curly braces is equal to $T_X(\{z\})$, in case ii), it is equal to $T_1(\{z\})$, and in case iii), it is equal to $T_2(\{z\})$. The result follows. □

Proof of Proposition 2: The total conditional cost is[14]

$$E|X \cap W^c| + E|N \cap W \cap X^c|$$

[14]Here, for brevity, we let E denote conditional expectation conditioned on the event $X \cup N \subseteq K$.

Now

$$
\begin{aligned}
E|X \cap W^c| &= E \sum_{z \in B} 1(z \in X \cap W^c) \\
&= \sum_{z \in B} E1(z \in X \cap W^c) \\
&= \sum_{z \in B} \Pr(z \in X \cap W^c \mid (X \cup N) \cap K^c = \emptyset) \\
&= \sum_{z \in K} \Pr(z \in X \cap W^c \mid (X \cup N) \cap K^c = \emptyset) \\
&= \sum_{z \in K} \frac{\Pr(z \in X \cap W^c, (X \cup N) \cap K^c = \emptyset)}{\Pr((X \cup N) \cap K^c = \emptyset)}.
\end{aligned}
$$

Observe that

$$
\begin{aligned}
&\Pr((X \cup N) \cap K^c = \emptyset) = \Pr(X \cap K^c = \emptyset, N \cap K^c = \emptyset) \\
&\text{(by independence of } X, N) = (1 - T_X(K^c))(1 - T_N(K^c))
\end{aligned}
$$

and that

$$
\begin{aligned}
&\Pr(z \in X \cap W^c, (X \cup N) \cap K^c = \emptyset) \\
&= \Pr(X \cap W^c \cap \{z\} \neq \emptyset, X \cap K^c = \emptyset, N \cap K^c = \emptyset) \\
&\text{(by independence of } X, N) \\
&= \Pr(X \cap (W^c \cap \{z\}) \neq \emptyset, X \cap K^c = \emptyset) \Pr(N \cap K^c = \emptyset) \\
&\qquad = (T_X(K^c \cup (W^c \cap \{z\})) - T_X(K^c))(1 - T_N(K^c)).
\end{aligned}
$$

Therefore, the first term of the expected cost becomes

$$\sum_{z \in K} \frac{T_X(K^c \cup (W^c \cap \{z\})) - T_X(K^c)}{1 - T_X(K^c)}.$$

For the second term of the expected cost

$$
\begin{aligned}
E|N \cap W \cap X^c| &= \sum_{z \in B} E1(z \in N \cap W \cap X^c) \\
&= \sum_{z \in B} \Pr(z \in N \cap W \cap X^c \mid (X \cup N) \cap K^c = \emptyset) \\
&= \sum_{z \in K} \Pr(z \in N \cap W \cap X^c \mid (X \cup N) \cap K^c = \emptyset) \\
&= \sum_{z \in K} \frac{\Pr(z \in N \cap W \cap X^c, (X \cup N) \cap K^c = \emptyset)}{\Pr((X \cup N) \cap K^c = \emptyset)}.
\end{aligned}
$$

We have seen that the denominator is equal to

$$(1 - T_X(K^c))(1 - T_N(K^c))$$

whereas the nominator

$$\begin{aligned}
&\Pr(z \in N \cap W \cap X^c, (X \cup N) \cap K^c = \emptyset) \\
&\quad = \Pr((\{z\} \cap W) \cap N \cap X^c \neq \emptyset, (X \cup N) \cap K^c = \emptyset) \\
&\quad = \Pr((\{z\} \cap W) \in N, (\{z\} \cap W) \in X^c, X \cap K^c \\
&\quad = \emptyset, N \cap K^c = \emptyset) \\
&\quad \text{(by indep. of } X, N) \\
&\quad = \Pr(X \cap K^c = \emptyset, (\{z\} \cap W) \in X^c) \\
&\quad \times \Pr(N \cap K^c = \emptyset, (\{z\} \cap W) \in N) \\
&\quad = \Pr(X \cap K^c = \emptyset, X \cap (\{z\} \cap W) = \emptyset) \\
&\quad \times \Pr(N \cap K^c = \emptyset, N \cap (\{z\} \cap W) \neq \emptyset) \\
&\quad = \Pr(X \cap (K^c \cup (\{z\} \cap W)) = \emptyset) \\
&\quad \times \Pr(N \cap K^c = \emptyset, N \cap (\{z\} \cap W) \neq \emptyset) \\
&\quad = [1 - T_X(K^c \cup (\{z\} \cap W))] \\
&\quad\ [T_N(K^c \cup (\{z\} \cap W)) - T_N(K^c)].
\end{aligned}$$

Therefore, the second term of the expected cost becomes the first equation at the bottom of the page, and the overall expected cost becomes the second equation at the bottom of the page, from which the claimed results follow by inspection. □

Proof of Proposition 4: Observe that by distributivity of erosion over intersection

$$(X \cap N) \ominus W^s = (X \ominus W^s) \cap (N \ominus W^s).$$

This property is crucial for the proof. During the course of the proof, we will need the following elementary result. Define the functional

$$\begin{aligned}
&\Gamma_{X,n}(K_0; K_1, \ldots, K_n) \\
&\quad \triangleq \Pr(X \cap K_0 = \emptyset, X \cap K_1 \neq \emptyset, \ldots, X \cap K_n \neq \emptyset).
\end{aligned}$$

By definition, $\Gamma_{X,0}(K) = Q_X(K)$. Using Bayes' rule, one can easily show that this functional satisfies the following recursion, which is known as the *inclusion-exclusion principle*.

$$\begin{aligned}
\Gamma_{X,n}(K_0; K_1, \ldots, K_n) &= \Gamma_{X,n-1}(K_0; K_1, \ldots, K_{n-1}) \\
&\quad - \Gamma_{X,n-1}(K_0 \cup K_n; K_1, \ldots, K_{n-1})
\end{aligned}$$

We are now ready to proceed with the proof of the proposition. The total expected error is

$$E(e) = E|X \backslash \widehat{X}| + E|\widehat{X} \backslash X|.$$

Let us first consider the second term.

$$\begin{aligned}
E|\widehat{X} \backslash X| &= E|\widehat{X} \cap X^c| \\
&= E|([(X \cap N) \ominus W^s] \cup (X \cap N)) \cap X^c| \\
&= E|([(X \cap N) \ominus W^s] \cap X^c) \cup (X \cap N \cap X^c)| \\
&= E|[(X \cap N) \ominus W^s] \cap X^c|.
\end{aligned}$$

Now, since

$$(X \cap N) \ominus W^s = (X \ominus W^s) \cap (N \ominus W^s)$$

the last expression is equal to

$$\begin{aligned}
&E|(X \ominus W^s) \cap (N \ominus W^s) \cap X^c| \\
&\quad = E \sum_{z \in B \backslash \partial B} 1(z \in (X \ominus W^s) \cap (N \ominus W^s) \cap X^c) \\
&\quad = \sum_{z \in B \backslash \partial B} E1(z \in (X \ominus W^s) \cap (N \ominus W^s) \cap X^c) \\
&\quad = \sum_{z \in B \backslash \partial B} \Pr(z \in (X \ominus W^s) \cap (N \ominus W^s) \cap X^c) \\
&\quad = \sum_{z \in B \backslash \partial B} \Pr(z \in (X \ominus W^s) \cap X^c), z \in N \ominus W^s) \\
&\quad \text{(by independence of } X, N) \\
&\quad = \sum_{z \in B \backslash \partial B} \Pr(z \in (X \ominus W^s) \cap X^c) \Pr(z \in N \ominus W^s) \\
&\quad = \sum_{z \in B \backslash \partial B} \Pr(W_z \subseteq X, z \in X^c) \Pr(W_z \subseteq N) \\
&\quad = \sum_{z \in B \backslash \partial B} \Pr(X^c \cap W_z = \emptyset, X^c \cap \{z\} \neq \emptyset) \\
&\qquad \Pr(N^c \cap W_z = \emptyset).
\end{aligned}$$

$$\sum_{z \in K} \frac{[1 - T_X(K^c \cup (\{z\} \cap W))][T_N(K^c \cup (\{z\} \cap W)) - T_N(K^c)]}{(1 - T_X(K^c))(1 - T_N(K^c))}$$

$$\begin{aligned}
E(e) &= \frac{1}{(1 - T_X(K^c))(1 - T_N(K^c))} \cdot \\
&\sum_{z \in K} \{[1 - T_X(K^c \cup (\{z\} \cap W))][T_N(K^c \cup (\{z\} \cap W)) - T_N(K^c)] + [T_X(K^c \cup (\{z\} \cap W^c)) - T_X(K^c)][1 - T_N(K^c)]\}
\end{aligned}$$

The first term of the total expected error

$$
\begin{aligned}
E|X\backslash\widehat{X}| &= E|X\cap(\widehat{X})^c| \\
&= E|X\cap([(X\cap N)\ominus W^s]\cup(X\cap N))^c| \\
&= E|X\cap[(X\cap N)\ominus W^s]^c\cap(X\cap N)^c| \\
&= E|[X\cap(X\cap N)^c]\cap[(X\cap N)\ominus W^s]^c| \\
&= E|[X\cap(X^c\cup N^c)]\cap[(X\cap N)\ominus W^s]^c| \\
&= E|[(X\cap X^c)\cup(X\cap N^c)]\cap[(X\cap N)\ominus W^s]^c| \\
&= E|X\cap N^c\cap[(X\cap N)\ominus W^s]^c| \\
&= E|X\cap N^c\cap[(X\ominus W^s)\cap(N\ominus W^s)]^c| \\
&= E|X\cap N^c\cap[(X\ominus W^s)^c\cup(N\ominus W^s)^c]| \\
&= E|(X\cap N^c\cap(X\ominus W^s)^c)\cup(X\cap N^c\cap(N\ominus W^s)^c)| \\
&= E|X\cap N^c\cap(X\ominus W^s)^c| + E|X\cap N^c\cap(N\ominus W^s)^c| \\
&\quad - E|X\cap N^c\cap(X\ominus W^s)^c\cap(N\ominus W^s)^c|
\end{aligned}
$$

Now

$$
\begin{aligned}
E|X\cap N^c\cap(X\ominus W^s)^c| &= E|(X\cap(X\ominus W^s)^c)\cap N^c| \\
&= \sum_{z\in B\backslash\partial B}\Pr(z\in X,\neg(W_z\subseteq X))\Pr(z\in N^c) \\
&= \sum_{z\in B\backslash\partial B}\Pr(X^c\cap\{z\}=\emptyset, X^c\cap W_z\neq\emptyset) \\
&\quad \Pr(N^c\cap\{z\}\neq\emptyset)
\end{aligned}
$$

where $\neg$ denotes logical negation. In addition

$$
\begin{aligned}
&E|X\cap N^c\cap(N\ominus W^s)^c| \\
&= \sum_{z\in B\backslash\partial B}\Pr(z\in X)\Pr(z\in N^c,\neg(W_z\subseteq N)) \\
&= \sum_{z\in B\backslash\partial B}\Pr(X^c\cap\{z\}=\emptyset)\Pr(N^c\cap\{z\}\neq\emptyset, N^c\cap W_z\neq\emptyset)
\end{aligned}
$$

and

$$
\begin{aligned}
&E|X\cap N^c\cap(X\ominus W^s)^c\cap(N\ominus W^s)^c| \\
&= E|(X\cap(X\ominus W^s)^c)\cap(N^c\cap(N\ominus W^s)^c)| \\
&= \sum_{z\in B\backslash\partial B}\Pr(z\in X,\neg(W_z\subseteq X))\Pr(z\in N^c,\neg(W_z\subseteq N)) \\
&= \sum_{z\in B\backslash\partial B}\Pr(X^c\cap\{z\}=\emptyset, X^c\cap W_z\neq\emptyset) \\
&\quad\times\Pr(N^c\cap\{z\}\neq\emptyset, N^c\cap W_z\neq\emptyset).
\end{aligned}
$$

Therefore, putting everything together, we have the expression at the bottom of this page, from which, after some manipulations, we obtain

$$
\begin{aligned}
E(e) = \sum_{z\in B\backslash\partial B}&\{Q_{X^c}(\{z\})(1-Q_{N^c}(\{z\})) \\
&+ Q_{N^c}(W_z)(Q_{X^c}(W_z)-Q_{X^c}(\{z\}\cup W_z)) \\
&+ Q_{X^c}(\{z\}\cup W_z)(Q_{N^c}(\{z\}\cup W_z)-Q_{N^c}(W_z))\}
\end{aligned}
$$

and the result is established. □

Proof of Lemma 2: First, observe that for any DRS X and $\forall K\in\Sigma(B)$

$$
\begin{aligned}
Q_{X\oplus H}(K) &= P_X((X\oplus H)\cap K=\emptyset) \\
&= P_X(X\cap(K\oplus H^s)=\emptyset) = Q_X(K\oplus H^s).
\end{aligned}
$$

Thus, assuming X is H open

$$
\begin{aligned}
Q_{X\ominus H^s}(K\oplus H^s) &= Q_{(X\ominus H^s)\oplus H}(K) = Q_{X\circ H}(K) \\
&= \sum_{K'\subseteq K^c} P_{X\circ H}(X\circ H=K') \\
&= \sum_{K'\subseteq K^c} P_X(X=K') = Q_X(K).
\end{aligned}
$$

Conversely, assume that $Q_X(K)=Q_{X\ominus H^s}(K\oplus H^s),\forall K\in\Sigma(B)$. Then, since $Q_{X\ominus H^s}(K\oplus H^s)=Q_{(X\ominus H^s)\oplus H}(K)=$

$$
\begin{aligned}
E(e) &= E|X\backslash\widehat{X}| + E|\widehat{X}\backslash X| \\
&= \sum_{z\in B\backslash\partial B}\{\Pr(X^c\cap\{z\}=\emptyset, X^c\cap W_z\neq\emptyset)\Pr(N^c\cap\{z\}\neq\emptyset) \\
&\quad + \Pr(X^c\cap\{z\}=\emptyset)\Pr(N^c\cap\{z\}\neq\emptyset, N^c\cap W_z\neq\emptyset) \\
&\quad - \Pr(X^c\cap\{z\}=\emptyset, X^c\cap W_z\neq\emptyset)\Pr(N^c\cap\{z\}\neq\emptyset, N^c\cap W_z\neq\emptyset) \\
&\quad + \Pr(X^c\cap W_z=\emptyset, X^c\cap\{z\}\neq\emptyset)\Pr(N^c\cap W_z=\emptyset)\} \\
&= \sum_{z\in B\backslash\partial B}\{\Gamma_{X^c,1}(\{z\};W_z)(1-Q_{N^c}(\{z\})) \\
&\quad + Q_{X^c}(\{z\})\Gamma_{N^c,2}(\emptyset;\{z\},W_z) \\
&\quad - \Gamma_{X^c,1}(\{z\};W_z)\Gamma_{N^c,2}(\emptyset;\{z\},W_z) + \Gamma_{X^c,1}(W_z;\{z\})Q_{N^c}(W_z)\} \\
&= \sum_{z\in B\backslash\partial B}\{(Q_{X^c}(\{z\})-Q_{X^c}(\{z\}\cup W_z))(1-Q_{N^c}(\{z\})) \\
&\quad + Q_{X^c}(\{z\})(1-Q_{N^c}(\{z\})-Q_{N^c}(W_z)+Q_{N^c}(\{z\}\cup W_z)) \\
&\quad - (Q_{X^c}(\{z\})-Q_{X^c}(\{z\}\cup W_z)) \\
&\quad\times(1-Q_{N^c}(\{z\})-Q_{N^c}(W_z)+Q_{N^c}(\{z\}\cup W_z)) \\
&\quad + (Q_{X^c}(W_z)-Q_{X^c}(\{z\}\cup W_z))Q_{N^c}(W_z)\}
\end{aligned}
$$

$Q_{X\circ H}(K)$, it follows that $Q_X(K) = Q_{X\circ H}(K), \forall K \in \Sigma(B)$. However,

$$\begin{aligned} P_{X\circ H}(X \circ H = K) &= \sum_{K' \subseteq K} (-1)^{|K'|} Q_{X\circ H}(K^c \cup K') \\ &= \sum_{K' \subseteq K} (-1)^{|K'|} Q_X(K^c \cup K') \\ &= P_X(X = K), \forall K \in \Sigma(B) \end{aligned}$$

and the proof is complete. □

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their interesting comments, which helped improve the quality of this manuscript.

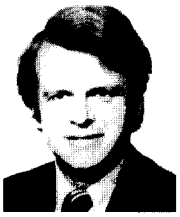
REFERENCES

[1] M. Aigner, *Combinatorial Theory*. New York: Springer-Verlag, 1979.
[2] Z. Artstein and R. A. Vitale, "A strong law of large numbers for random compact sets," *Annals Probabil.*, vol. 3, no. 5, pp. 879–882, 1975.
[3] G. Choquet, "Theory of capacities." *Ann. Inst. Fourier*, vol. 5, pp. 131–295, 1953.
[4] N. Cressie, "A strong limit theorem for random sets," *Supplement to Advances Applied Probabil.* vol. 10, pp. 36–46, 1978.
[5] E. Dougherty, "Optimal mean square N-observation digital morphological filters—I. Optimal binary filters," *Comput. Vision Graphics Image Processing: Image Understanding*, vol. 55, no. 1, pp. 36–54, Jan. 1992.
[6] ——, "Optimal mean square N-observation digital morphological filters—II. Optimal gray-scale filters," *Comput. Vision Graphics Image Processing: Image Understanding*, vol. 55, no. 1, pp. 55–72, Jan. 1992.
[7] E. Dougherty and C. Giardina, *Morphological Methods in Image and Signal Processing*. Englewood Cliffs, NJ: Prentice Hall, 1988.
[8] E. R. Dougherty *et al.*, "Morphological pattern-spectra-based tau-opening optimization," *Proc. SPIE*, vol. 1606, Nov. 1991.
[9] E. R. Dougherty and R. P. Loce, "Constrained optimal digital morphological filters," *in Proc. 25th Ann. Conf. Inform. Sci. Syst.*, (Baltimore, MD), Mar. 1991.
[10] E. R. Dougherty, A. Mathew, and V. Swarnakar, "A conditional-expectation-based implementation of the optimal mean-square binary morphological filter," *Proc. SPIE*, vol. 1451, 1991.
[11] J. Serra, Ed., *Image Analysis and Mathematical Morphology, Vol. 2*. San Diego: Academic, 1988.
[12] J. Goutsias, "Morphological analysis of discrete random shapes," *J. Math. Imaging Vision*, vol. 2, pp. 193–215, 1992.
[13] J. Goutsias and C. Wen, "Modeling discrete random shapes: A random set theory approach," *Tech. Rep. JHU/ECE* 90–13, Johns Hopkins Univ., 1990.
[14] ——, "Discrete random set models for shape synthesis as analysis," *Proc. SPIE*, vol. 1606, pp. 174–185, 1991.
[15] R. M. Haralick, E. R. Dougherty, and P. L. Katz, "Model-based morphology." *Proc. SPIE* vol. 1472, Apr. 1991.
[16] R. M. Haralick, X. Zhuang, C. Lin, and J. S. J. Lee. "The digital morpological sampling theorem," *IEEE Trans. Acoust. Speech Signal Processing*, vol. 37, no. 12, pp. 2067–2089, 1989.
[17] D. G. Kendall, "Foundations of a theory of random sets," in *Stochastic Geometry* (E. F. Harding and D. G. Kendall, Eds.). London: Wiley, 1974, pp. 322–376.
[18] T. Y. Kong and A. Rosenfeld, "Digital topology: Introduction and survey," *Comput. Vision Graphics Image Processing*, vol. 48, pp. 357–393, 1989.
[19] R. P. Loce and E. R. Dougherty, "Using structuring element libraries to design suboptimal morphological filters," in *Proc. Conf. Image Algebra Morphological Image Processing II, SPIE* (San Diego, CA), July 1991, pp. 233–246, vol. 1568.
[20] P. Maragos, "A representation theory for morphological image and signal processing," *IEEE Trans. Patt. Anal. Machine Intell.*, vol. 11, no. 6, pp. 586–599, June 1989.
[21] G. Matheron, *Elements pour une theorie des Milieux Poreux*. Masson, 1967.
[22] ——, *Random Sets and Integral Geometry*. New York: Wiley, 1975.
[23] H. V. Poor, *An Introduction to Signal Detection and Estimation*. New York: Springer-Verlag, 1988.
[24] B. D. Ripley, "Locally finite random sets: Foundations for point process theory," *Ann. Probabil.*, vol. 4, pp. 983–994, 1976.
[25] D. Schonfeld and J. Goutsias, "Optimal morphological pattern restoration from noisy binary images," *IEEE Trans. Patt. Anal. Machine Intell.*, vol. 13, no. 1, pp. 14–29, Jan. 1991.
[26] J. Serra, Image Analysis and Mathematical Morphology.. New York: Academic, 1982.
[27] J. Serra and L. Vincent, "An overview of morphological filtering," *Circuits Syst. Signal Processing*, vol. 11, no. 1, pp. 47–108, Jan. 1992.
[28] N. D. Sidiropoulos, "Statistical inference, filtering, and modeling of discrete random sets." Ph.D. thesis, Univ. Maryland, June 1992.
[29] N. D. Sidiropoulos, J. Baras, and C. Berenstein, "The structure of divisible discrete random sets and their randomized superpositions," Tech. Rep. TR 91–54, Syst. Res. Cent., Univ. of Maryland, Apr. 1991.
[30] N. D. Sidiropoulos, J. S. Baras, and C. A. Berenstein, "Discrete random sets: An inverse problem, plus tools for the statistical inference of the discrete Boolean model," in *Proc. Conf. Image Algebra Morphological Image Processing III, SPIE* (San Diego), July 1992, vol. 1769.
[31] ——, "Optimal mask filtering of discrete random sets under a union/intersection noise model," in *Proc. 26th Ann. Conf. Inform. Sci. Syst.* (Princeton, NJ), Mar. 1992.
[32] ——, "Optimal morphological filters for discrete random sets under a union or intersection noise model," in *Proc. Conf. Visual Commun. Image Processing, SPIE* (Boston), Nov. 1992, vol. 1818.
[33] ——, "An algebraic analysis of the generating functional for discrete random sets and statistical inference for intensity in the discrete Boolean random set model," *J. Math. Imaging Vision*, 1993.
[34] J. Song and E. J. Delp, "A study of the generalized morphological filter," *Circuits Syst. Signal Processing*, vol. 11, no. 1, pp. 229–252, Jan. 1992.
[35] D. Stoyan, W. S. Kendall, and J. Mecke, *Stochastic Geometry and its Applications*. Berlin: Wiley, 1987.
[36] R. A. Vitale, "Random convex hulls: Floating bodies and expectations," *J. Approx. Theory*, to appear.
[37] ——, "Some developments in the theory of random sets." *Bull. Int. Stat. Inst.*, vol. 50, pp. 863–871, 1983.
[38] ——, "Symmetric statistics and random shape," in *Proc. First World Cong. Bernoulli Soc.* (Tashkent, USSR), Sept. 1986, invited talk.
[39] ——, "Expected convex hulls, order statistics and Banach space probabilities," *Acta Applicandae Mathematicae*, vol. 9, pp. 97–102, 1987.
[40] ——, "An alternative formulation of mean value for random geometric Figure," *J. Microscopy*, vol. 151, no. 3, pp. 197–204, Sept. 1988.
[41] ——, "The translative expectation of a random set," *J. Math. Anal. Applications*, vol. 160, no. 2, pp. 556–562, Sept. 1991.

Nicholas D. Sidiropoulos (M'92) received the Diploma in electrical engineering from the Aristotelian University of Thessaloniki, Greece, in 1988 and the M.S. and Ph.D. degrees in electrical engineering from the University of Maryland at College Park in 1990 and 1992, respectively.

From 1988 to 1992, he was a Fulbright Fellow and a research assistant at the Center for Systems Research of the University of Maryland. He is currently on military duty, serving as scientific assistant and lecturer at the Hellenic Air Force Academy. His current research interests are in the areas of random set theory, image and signal processing, and statistical communications.

Dr. Sidiropoulos is a member of the Technical Chamber of Greece.



John S. Baras (F'84) was born in Piraeus, Greece, on March 13, 1948. He received the B.S. degree in electrical engineering with highest distinction from the National Technical University of Athens, Greece, in 1970. He received the M.S. and Ph.D. degrees in applied mathematics from Harvard University, Cambridge, MA, in 1971 and 1973, respectively.

Since 1973, he has been with the Department of Electrical Engineering, University of Maryland at College Park, where he is currently Professor and member of the Faculty of Applied Mathematics. From 1985 to 1991, he was the Founding Director of the Systems Research Center, which is now the Institute for Systems Research. In February of 1990, he was appointed to the Martin Marietta Chair in Systems Engineering. Since 1991, he has been the Co-Director of the Center for Satellite and Hybrid Communication Networks, a NASA Center for the Commercial Development of Space, which he co-founded. He has held visiting research scholar positions with Stanford, the Massachusetts Institute of Technology, Harvard University, the Institute National de reserche en Informatique et en Automatique, and the University of California, Berkeley. He has numerous publications in control and communication systems and is the co-editor of *Recent Progress in Stochastic Calculus* (Springer-Verlag, 1990). His current research interests include stochastic systems, signal processing and understanding with emphasis on speech and image signals, real-time architectures, symbolic computation, intelligent control systems, robust nonlinear control, distributed parameter systems, and hybrid communicationnetwork simulation and management.

Dr. Baras has won several awards including a 1978 Naval Research Laboratory Research Publication Award, the 1980 Outstanding Paper Award of the IEEE Control Sytems Society, and the 1983 Alan Berman Research Publication Award from the Naval Research Laboratory. He has served with the IEEE Engineering Research and Development Committee, the Aerospace Industries Association Advisory Committee on Advanced Sensors, the IEEE Fellow Evaluation Committee, and the IEEE Control Systems Society Board of Governors (1991–1993). He is currently serving on the editorial boards of *Mathematics of Control, Signals, and Systems; Systems and Control: Foundations and Applications; IMA Journal of Mathematical Control and Information; Systems Automation—Research and Applications*, and he is Managing Editor of the series *Progress in Automation and Information Systems* (Springer-Verlag). He is a member of Sigma Xi, the American Mathematical Society, and the Society for Industrial and Applied Mathematics.

Carlos A. Berenstein received the Licenciado en Matematicas degree from the University of Buenos Aires, Argentina, in 1966 and the Ph.D. degree in mathematics from the Courant Institute, New York University, New York, NY, in 1970.

Since 1970, he has taught at Harvard University and Brandeis University and is currently a Professor of Mathematics and a member of the Institue for Systems Research at the University of Maryland, College Park. He has also been a visiting professor at several foreign universities, including the Universities of Paris (Bordeaux) and the Scuola Normale Superiore (Pisa, Italy). He has written several monographs and more than 100 research articles on Fourier analysis, function theory in one and several complex variables, convolution equations, difference-differential equations, boudary value problems, integral geometry, exponential polynomials and functional analysis, and digital geometry. His current interests include applications of the above mathematical theories to tomography, filtering theory, image processing, and complexity theory.