92-17

PROCEEDINGS

SPIE—The International Society for Optical Engineering

Visual Communications and Image Processing '92

Petros Maragos
Chair/Editor

18–20 November 1992
Boston, Massachusetts

SPIE PROCEEDINGS SERIES

Volume 1818
Part One of Three Parts

Optimal Morphological Filters for Discrete Random Sets under a Union or Intersection Noise Model *

N.D. Sidiropoulos J.S. Baras C.A. Berenstein

Systems Research Center
University of Maryland
College Park, MD 20742

Abstract

We consider the problem of optimal binary image restoration under a union or intersection noise model. Union noise is well suited to model random clutter (obscuration), whereas intersection noise is a good model for random sampling. Our approach is random set-theoretic, i.e. digital images are viewed as realizations of a uniformly bounded discrete random set. First we provide statistical proofs of some "folk theorems" of Morphological filtering. In particular, we prove that, under some reasonable worst-case statistical scenarios, Morphological openings, closings, unions of openings, and intersections of closings, can be viewed as MAP estimators of the signal based on the noisy observation. Then we propose a "generic" procedure for the design of optimal Morphological filters for independent union or intersection noise.

1 Introduction

An important problem in digital image processing and analysis is the development of optimal filtering procedures which attempt to restore a binary image ("signal") from its degraded version [17, 5]. 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 [9]. This degradation model is known as the union noise model. Other models exist, such as the intersection noise model, and the combined 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 [9], occlusion, etc.

This research has been largely motivated by the works of Haralick, Dougherty, and Katz [9], and Schonfeld and Goutsias [17]. Their approach is model-based, in that they assume specific probabilistic/geometrical

*Keywords: Optimal Filtering, Mathematical Morphology, Discrete Random Sets, Union or Intersection Noise model

 0-8194-1018-7/92/$4.00

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, Dougherty, and Katz assume that the signal and noise patterns are "non-interfering" 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, Dougherty, and Katz 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 which determines the optimal filter within a restricted family of Morphological Opening Filters [14, 18, 4]. 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) [14, 18, 4], 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 which 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. [5, 7, 6].

Our work differs from the reviewed literature in many ways. The most important is that it is based on *statistical*, rather than *topological*, assumptions on the signal and the noise. We start with some background on discrete random sets.

2 Discrete random set fundamentals

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. We have the following fundamental theorem. See [19], or [20, 21] for the proof.

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

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

The *uniqueness* part of this theorem is originally due to Choquet [1], and it has been independently introduced in the context of continuous-domain random set theory by Kendall [10] and Matheron [13, 14]. Related results can also be found in Ripley [16]. However, the measure reconstruction formulas are essentially only applicable within a uniformly bounded discrete random set setting.

3 Mathematical Morphology

The theory of *Mathematical Morphology* has been developed mainly by Serra [18, 8], Matheron [14], and their collaborators, during the 70's and early 80's. It is founded on two elementary set operators, namely set dilation/erosion, which are defined in terms of a *structuring element*, i.e. a "small", primitive set of points. Let

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

The dilation of a set $Y \subset \mathbf{Z}^2$ by a structuring element W is defined as[1]

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

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

$$Y \ominus W^s \triangleq \{z \in \mathbf{Z}^2 \mid W_z \subseteq Y\}$$

The opening, $Y \circ W$, of a set $Y \subset \mathbf{Z}^2$ by a structuring element W, is defined as

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

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

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

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.

[1] Here we follow the original definitions of Serra [18]. In his work the symbol $\oplus$ stands for Minkowski set addition, and the symbol $\ominus$ stands for Minkowski set subtraction.

4 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 [17, 4, 7, 6, 2, 3, 22]. For example, it is widely believed that opening is suitable under a union noise model, while 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 reasonable worst-case scenarios. In particular, if we assume that the signal is sufficiently smooth, and the noise is i.i.d., then these filters provide the Maximum A Posteriori (MAP) estimate of the signal, X, on the basis of the observation Y. 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 $\Sigma(B)$:*

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

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

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

Using Bayes' rule,

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

$$= arg\ max_{K \in O_W(B)} \left\{ Pr(Y \mid X = K) \frac{1}{|O_W(B)|} \right\} = arg\ max_{K \in O_W(B)} \{Pr(Y \mid X = K)\}$$

$$= arg\ max_{K \in O_W(B),\ K \subseteq Y} \{Pr(Y \mid X = K)\} = arg\ max_{K \in O_W(B),\ K \subseteq Y} \left\{ r^{|Y|-|K|} (1-r)^{|B|-|Y|} \right\}$$

$$= arg\ max_{K \in O_W(B),\ K \subseteq Y} \left\{ r^{-|K|} \right\} = arg\ max_{K \in O_W(B),\ K \subseteq Y} \{|K|\}$$

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

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

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 which 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. Also, 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. By duality, we have a similar theorem for MAP optimality of closing under i.i.d. intersection noise [19]. We state the following generalizations without proof.

Theorem 3 *Let $O_{W_1,\cdots W_M}(B)$ denote the collection of all subsets K of B which can be written as*

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

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

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

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

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

Theorem 4 *Let $C_{W_1,\cdots W_M}(B)$ denote the collection of all subsets K of B which can be written as*

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

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

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

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

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

In practice, it may not be realistic to assume that the noise is i.i.d., and/or that the signal is uniformly distributed over a collection of Morphologically smooth sets. Our programme is to develop a "generic" optimal design procedure to address this problem. For this, we need to move away from our current notion of optimality. Optimality is usually defined by means of minimizing a cost function. Up to this point, we have focused on MAP optimality, which corresponds to minimizing the total probability of error. In what follows, we reformulate the optimal filtering problem in a more general setting, and consider an alternative cost function.

5 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 $g : \Sigma(B) \times \Sigma(B) \longmapsto \Sigma(B)$ 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 $d : \Sigma(B) \times \Sigma(B) \longmapsto \mathbf{Z}_+$ be a distance metric between subsets of B. In this context, the optimal filtering problem is to find a mapping $f : \Sigma(B) \longmapsto \Sigma(B)$ such that the expected cost (expected error)

$$E(e) \triangleq Ed(X, \widehat{X}), \quad \widehat{X} = f(Y) = f(g(X, N))$$

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 $f : \Sigma(B) \longmapsto \Sigma(B)$ is just too big. It is common practice to *impose* structure on the search space, i.e. constrain f to lie in $\mathcal{F}$, 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)

$$d(X, \widehat{X}) = |(X \backslash \widehat{X}) \cup (\widehat{X} \backslash X)| = |(X \backslash \widehat{X})| + |(\widehat{X} \backslash X)| = |(X \cup \widehat{X}) \backslash (X \cap \widehat{X})| = |(X \cup \widehat{X})| - |(X \cap \widehat{X})|$$

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* [15] when $X, \widehat{X}$ 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 $\{0,1\}^{|B|}$.

In terms of the degradation, we assume that N is independent of X, and that the mapping g is given by

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

or,

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

6 Optimal increasing, shift-invariant filters with a basis constraint

A surprising result, originally due to Matheron [14], and subsequently improved upon, and used by Maragos [12], and Dougherty et al., and Giardina [4, 5, 7, 6], is that a very large class of (linear and non-linear) shift-invariant operators 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, taken from [14, 12].

Theorem 5 *[14] 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*

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

Theorem 6 *[12] 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*

$$Bas(\Psi) \triangleq \left\{W \in Ker(\Psi) \mid W' \in Ker(\Psi) \text{ and } W' \subseteq W \Rightarrow W' = W\right\}$$

As a result of the latter theorem, the number of structuring elements that are needed for the decomposition is greatly reduced. Dougherty et al. [5, 7, 6], 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 [2] over a *dilation basis.* Using these decompositions, the problem of finding the optimal shift-invariant and increasing filter reduces to a problem of optimal basis design, which admits a natural hierarchical parameterization in terms of a basis size constraint. The upper bound on the size of the basis is usually determined by design and implementation complexity 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.

7 Optimizing a single structuring element

In the case of union noise, the simplest non-trivial 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 anti-extensive, i.e. the output must be contained in the input. Dilation by the origin simply yields the input itself. Therefore, the simplest non-trivial class of constrained dilation basis filters for union noise can be written as follows:

$$\hat{X} = f(Y) = f_W(Y) = (Y \oplus W^s) \cap Y = [(X \cup N) \oplus W^s] \cap (X \cup N), \textit{ for some structuring element, } W$$

Similarly, in the case of intersection noise, the simplest non-trivial expansion in terms of an erosion basis involves two structuring elements, one of which is constrained to be the 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 non-trivial class of constrained erosion basis filters for intersection noise can be written as follows:

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

Let us first consider intersection noise. Let $\mathcal{W}$ denote 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

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

$B \backslash \partial B$ 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 1 *Under the assumption of mutual independence of the signal and noise DRS's, X, N, the value of the expected error, $E(e) = Ed(X, \widehat{X})$, incurred when X is estimated by $\widehat{X} = [(X \cap N) \ominus W^s] \cup (X \cap N)$, is given by*

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

$$+ Q_{N^c}(W_z)(Q_{X^c}(W_z) - Q_{X^c}(\{z\} \cup W_z)) + Q_{X^c}(\{z\} \cup W_z)(Q_{N^c}(\{z\} \cup W_z) - Q_{N^c}(W_z))\}$$

Proof: See [19] □

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[2] assumption that both X, and N, are obtained by sampling *stationary* random sets [14], 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 which minimizes the* probability of pixel error

$$P_{pixel\ error}(W) = Q_{X^c}(\{\bar{0}\})(1 - Q_{N^c}(\{\bar{0}\}))$$

$$+ Q_{N^c}(W)(Q_{X^c}(W) - Q_{X^c}(\{\bar{0}\} \cup W)) - Q_{X^c}(\{\bar{0}\} \cup W)(Q_{N^c}(W) - Q_{N^c}(\{\bar{0}\} \cup W))$$

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

[2] Since we are using a *shift-invariant* filtering operation.

elements, some sort of suboptimal search must be pursued, to avoid a potentially difficult exhaustive search. See [11] 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 which 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. By appealing to duality, we obtain the following result.

Proposition 2 *Under the assumption of mutual independence of the signal and noise DRS's, X, N, the value of the expected error, $E(e) = Ed(X, \widehat{X})$, incurred when X is estimated by $\widehat{X} = [(X \cup N) \oplus W^s] \cap (X \cup N)$, is given by*

$$E(e) = \sum_{z \in B \setminus \partial B} \{Q_X(\{z\})(1 - Q_N(\{z\}))$$

$$+Q_N(W_z)(Q_X(W_z) - Q_X(\{z\} \cup W_z)) + Q_X(\{z\} \cup W_z)(Q_N(\{z\} \cup W_z) - Q_N(W_z))\}$$

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 which minimizes the* probability of pixel error

$$P_{pixel\ error}(W) = Q_X(\{\bar{0}\})(1 - Q_N(\{\bar{0}\}))$$

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

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 4 *A DRS X is H-open iff*[3]

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

Lemma 1 *X is H-open iff $Q_X(K) = Q_{X \ominus H^s}(K \oplus H^s)$, $\forall K \in \Sigma(B)$.*

---

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

Proof: See [19] □

So now let us assume that the signal DRS, X, is H-open, where H is convex and contains the origin. Then we can prove the following [19].

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*

$$W^* = arg\ min_{W \in \mathcal{W}} P_{pixel\ error}(W) = arg\ max_{W \in \widetilde{\mathcal{W}}} G(W)$$

where

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

and

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

This elimination can translate to a significant reduction in search complexity. 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.

8 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. This joint optimization does not pose additional analytical problems. For example, we have the following proposition (see [19] for a proof).

Proposition 3 *Under the assumption of mutual independence of the signal and noise DRS's, X, N, the value of the expected error, $E(e) = Ed(X, \widehat{X})$, incurred when X is estimated by*

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

is given by

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

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 section. 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.

9 Conclusions

We have pursued a DRS-theoretic approach to the problem of digital image restoration under a union or intersection noise model. A generic procedure for the statistical optimization of Morphological filters subject to a basis size constraint has been proposed and investigated. New statistical insight into some "folk theorems" of Morphological filtering has been gained as a direct consequence of our approach.

10 Acknowledgments

This research was partially supported by NSF grant NSFD CDR 8803012, through the Engineering Research Center program. N.D. Sidiropoulos and J.S. Baras are also with the Department of Electrical Engineering. C.A. Berenstein is also with the Department of Mathematics.

References

[1] G. Choquet. Theory of capacities. *Ann. Institute Fourier*, 5:131–295, 1953.

[2] E. Dougherty. Optimal Mean Square N-Observation Digital Morphological Filters - I. Optimal Binary Filters. *Computer Vision, Graphics, and Image Processing: Image Understanding*, 55(1):36–54, January 1992.

[3] E. Dougherty. Optimal Mean Square N-Observation Digital Morphological Filters - II. Optimal Gray - Scale Filters. *Computer Vision, Graphics, and Image Processing: Image Understanding*, 55(1):55–72, January 1992.

[4] E. Dougherty and C. Giardina. *Morphological Methods in Image and Signal Processing*. Prentice-Hall, Englewood Cliff, 1988.

[5] E.R. Dougherty, R.M. Haralick, Y. Chen, B. Li, C. Agerskov, U. Jacobi, and P. H. Sloth. Morphological pattern-spectra-based tau-opening optimization. In *Proc. SPIE Vol. 1606. Boston, Massachusetts*. Society for Optical Engineering, November 1991.

[6] E.R. Dougherty and R.P. Loce. Constrained optimal digital morphological filters. In *Proc. of the 25th Annual Conference on Information Sciences and Systems, The Johns Hopkins University, Baltimore, Maryland*, March 1991.

[7] E.R. Dougherty, A. Mathew, and V. Swarnakar. A conditional-expectation-based implementation of the optimal mean-square binary morphological filter. In *Proc. SPIE Vol. 1451, San Jose, California.* Society for Optical Engineering, February 1991.

[8] J. Serra Ed. *Image Analysis and Mathematical Morphology, vol. 2, Theoretical Advances.* Academic, San Diego, 1988.

[9] R.M. Haralick, E.R. Dougherty, and P.L. Katz. Model-based morphology. In *Proc. SPIE Vol. 1472, Orlando, Florida.* Society for Optical Engineering, April 1991.

[10] D.G. Kendall. Foundations of a theory of random sets. In E.F. Harding and D.G. Kendall, editors, *Stochastic Geometry,* pages 322–376. John Wiley, London, England, 1974.

[11] R.P. Loce and E.R. Dougherty. Using Structuring Element Libraries to Design Suboptimal Morphological Filters. In P.D. Gader and E.R. Dougherty, editors, *Proc. of Conference on Image Algebra and Morphological Image Processing II, SPIE vol. 1568, San Diego, CA*, pages 233–246. Society for Optical Engineering, July 1991.

[12] P. Maragos. A Representation Theory for Morphological Image and Signal Processing. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 11(6):586–599, June 1989.

[13] G. Matheron. *Elements pour une theorie des Milieux Poreux.* Masson, 1967.

[14] G. Matheron. *Random Sets and Integral Geometry.* Wiley, New York, 1975.

[15] H. V. Poor. *An introduction to Signal Detection and Estimation.* Springer-Verlag, New York, 1988.

[16] B.D. Ripley. Locally finite random sets: foundations for point process theory. *Ann. Probab.*, 4:983–994, 1976.

[17] D. Schonfeld and J. Goutsias. Optimal morphological pattern restoration from noisy binary images. *IEEE Transactions on Pattern Anal. Mach. Intell.*, 13(1):14–29, Jan. 1991.

[18] J. Serra. *Image Analysis and Mathematical Morphology.* Academic, New York, 1982.

[19] N.D. Sidiropoulos. *Statistical Inference, Filtering, and Modeling of Discrete Random Sets.* PhD thesis, University of Maryland, June 1992.

[20] 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 P.D. Gader, E.R. Dougherty, and J. Serra, editors, *Proc. of Conference on Image Algebra and Morphological Image Processing III, SPIE vol. 1769, San Diego, CA.* Society for Optical Engineering, July 1992.

[21] N.D. Sidiropoulos, J.S. Baras, and C.A. Berenstein. Optimal Mask Filtering of Discrete Random Sets under a Union / Intersection Noise Model. In *Proc. 26th Annual Conference on Information Sciences and Systems*, Princeton University, Princeton, N.J., March 1992.

[22] J. Song and E.J. Delp. A Study of the Generalized Morphological Filter. *Circuits, Systems and Signal Processing*, 11(1):229–252, January 1992.