TECHNICAL RESEARCH REPORT

SYSTEMS RESEARCH CENTER

Supported by the National Science Foundation Engineering Research Center Program (NSFD CD 8803012), Industry and the University

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

by N.D. Sidiropoulos, J.S. Baras and C.A. Berenstein

TR 92-7r1

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

N.D. Sidiropoulos‡ J.S. Baras§ C.A. Berenstein¶

Systems Research Center
University of Maryland
College Park, MD 20742

Abstract

We model digital binary image data as realizations of a bounded discrete random set, a mathematical object which can be directly defined on a finite lattice. We consider the problem of estimating realizations of discrete random sets distorted by a degradation process which can be described by a union/intersection model. First we present an important structural result concerning the probabilistic specification of discrete random sets defined on a finite lattice. Then we formulate the optimal filtering problem for the case of discrete random sets. Two distinct filtering approaches are pursued. For images which feature strong spatial statistical variations we propose a simple family of spatially varying filters, which we call *mask filters*, and, for each degradation model, derive explicit formulas for the optimal Mask filter. We also consider adaptive mask filters, which can be effective in a more general setting. For images which exhibit a stationary statistical behavior, we consider the class of Morphological filters. First we provide some theoretical justification for the popularity of certain Morphological filtering schemes. In particular, we show that if the signal is smooth, then these schemes are optimal (in the sense of providing the MAP estimate of the signal) under a reasonable worst-case statistical scenario. Then we show that, by using an appropriate (under a given degradation model) expansion of the optimal filter, we can obtain universal characterizations of optimality which do not rely on strong assumptions regarding the spatial interaction of geometrical primitives of the signal and the noise. This approach corresponds to a somewhat counter-intuitive use of fundamental Morphological operators; however it is exactly this mode of use that enables us to arrive at characterizations of optimality in terms of the fundamental functionals of random set theory, namely the generating functionals of the signal and the noise.

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

†Research partially supported by NSF grant NSFD CDR 8803012, through the Engineering Research Center s Program

‡Also with the Department of Electrical Engineer ing

§Also with the Department of Electrical Engineerin g

¶Also with the Department of Mathematics

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

N.D. Sidiropoulos‡ J.S. Baras§ C.A. Berenstein¶

Systems Research Center
University of Maryland
College Park, MD 20742

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 noisy version [22, 6]. Here, the noise process usually models the combined effect of two distinct types of degradation, 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 [12]. 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. These models are well justified in practice, because, usually, binary images are obtained by thresholding gray-level images. If the threshold value is set sufficiently low, then the resulting degraded binary images will be well described by a union noise model. Alternatively, if the threshold is set sufficiently high, then the intersection noise model will be appropriate. In between these extreme choices, a union/intersection noise model will be most appropriate. The assumption of independence is crucial for the theoretical analysis of optimal filters, and it is plausible in many practical situations.

This research has been largely motivated by the works of Haralick/Dougherty/Katz [12], and Schonfeld/Goutsias [22]. 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/Dougherty/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/Goutsias make a stronger assumption concerning the separability of noise patterns. These assumptions are reasonable if the image is sparse,

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

†Research partially supported by NSF grant NSFD CDR 8803012, through the Engineering Research Center s Program

‡Also with the Department of Electrical Engineer ing

§Also with the Department of Electrical Engineerin g

¶Also with the Department of Mathematics

i.e. the signal and noise patterns are most likely to remain uncluttered. Haralick/Dougherty/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 [19, 23, 5]. In their work, the signal and noise patterns are all assumed to be of the same basic shape, and only their size varies. Schonfeld/Goutsias consider Morphological Alternating Sequential Filters (ASF's) [19, 23, 5], 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. [6, 8, 7].

This work focuses on a different viewpoint. As it turns out, by restricting our attention to suitable classes of filtering operations, and Discrete Random Sets (defined below), we can obtain truly 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 "non-interfering". 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.

The rest of this paper is organized as follows. Section 2 summarizes some Discrete Random Set fundamentals. Section 3 formalizes the optimal filtering problem. Sections 4 and 5 contain the core of this work. Section 4 contains results on optimal mask filtering, while section 5 contains results on optimal Morphological filtering. A simulation experiment is also included. Finally, section 6 offers some conclusions.

2 Discrete Random Set Fundamentals

Definition 1 *Let B be a bounded subset of $\mathcal{Z}^2$. 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* **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*

$$T_X(K) = P_X(X \cap K \neq \emptyset)$$

is known as the **capacity functional** *of the DRS X.*

Definition 3 *The functional*

$$Q_X(K) = P_X(X \cap K = \emptyset) = 1 - T_X(K)$$

is known as the **generating functional** *of the DRS X.*

The following Lemma will be useful. Its proof can be found in the appendix. See [1] for basic Mobius inversion.

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

$$v(A) = \sum_{S \subseteq A^c} u(S) \qquad \text{“external decomposition”}$$

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.

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 reconstruction formula for the functional $P_X(X = K)$ in terms of the functional Q_X is a direct consequence of Lemma 1 and the fact that Q_X can be expressed in terms of P_X as

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

□

The *uniqueness* part of this theorem is originally due to Choquet [2], and it has been independently introduced in the context of continuous-domain random set theory by Kendall [14] and Matheron [18, 19]. Related results can also be found in Ripley [21]. However, the measure reconstruction formulas are essentially only applicable within a discrete, bounded setting. In the continuous case, the uniqueness result relies heavily on Kolmogorov's extension theorem, which is non-constructive. See [24, 10, 11] for some other interesting results on DRS's.

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.

3 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 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 $f : \Sigma(B) \mapsto \Sigma(B)$ is a chaotic search space! 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)

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

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* [20] 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 $\{0,1\}^{|B|}$, i.e., with some abuse of notation,

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

where on the left hand side symbols are interpreted as sets, while 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) [20] applies. It states that a sufficient condition for the existence of a $f^* \in \mathcal{F}$ such that

$$E\left[(X - f^*(Y))^T (X - f^*(Y))\right] \leq E\left[(X - f(Y))^T (X - f(Y))\right], \quad \forall f \in \mathcal{F}$$

is that

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

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. For certain choices of $\mathcal{F}$ it is easy to show that the necessity part of the OP does not hold. At any rate, it is often easier to write down an expression for $Ed(X,\widehat{X})$, and optimize over $\mathcal{F}$ by brute force.

This choice of distance has many advantages [22], not the least of which is that it enables the derivation of explicit optimality conditions. Even though, technically speaking, $d(X,\widehat{X})$ can be considered as a quadratic distance measure when we view $X,\widehat{X}$ as vectors in $\{0,1\}^{|B|}$, from a set-theoretic point of view $d(X,\widehat{X})$ 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*. From a quadratic estimation-theoretic point of view, a proper *formal* definition of the expectation of a DRS X, would be as follows.

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

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

$$arg\ min_{W\in\Sigma(B)} E\left[d(X,W)\right]^2 =$$

$$arg\ min_{W\in\Sigma(B)} \left\{ |W|^2 + 2|W| \left(\sum_{z\in W^c} Pr(z\in X) - \sum_{z\in W} Pr(z\in X) \right) - 4 \sum_{z\in W^c} \sum_{y\in W} Pr(z\in X\ ,\ y\in X) \right\}$$

If we assume that $Pr(z\in X) = p$, $\forall z\in B$, and $Pr(z\in X\ ,\ y\in X) = Pr(z\in X)Pr(y\in X) = p^2$, $\forall\ z,y\ s.t.\ z\neq y$, 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*, formally defined as

$$MX \triangleq arg\ min_{W\in\Sigma(B)} Ed(X,W)$$

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, which, for each point $z\in B$, is either true, or false. Define the *indicator function*

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

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 that

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

is the unique minimum cardinality solution to the minimization problem

$$min_{W\in\Sigma(B)} Ed(X,W)$$

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

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

or,

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

Alternatively, we may allow g to be the following measurable mapping from $\Sigma(B) \times \Sigma(B) \times \Sigma(B)$ to $\Sigma(B)$

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

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

4 Optimal Mask Filters

In the case of union noise, a simple yet intuitive class of filters is

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

Similarly, in the case of intersection noise, we can consider the following class of filters

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

Finally, in the case of union/intersection noise, we can consider

$$f(Y) = f_{W_2}^{W_1}(Y) = (Y \cap W_2) \cup W_1 = ((((X \cap N_1) \cup N_2) \cap W_2) \cup W_1), \ \textit{for some } W_1, W_2, \ \textit{both in } \Sigma(B)$$

We call these filters *mask filters*. In the simplest case the mask W is fixed; in the adaptive case W is allowed to depend on the observation. We will consider both cases. Observe that, under the given degradation models, unconstrained adaptive mask filtering is the most general filtering structure that we can consider. We will show that explicit optimization is possible under some restrictions on the adaptation strategy. On the other hand, simple fixed mask filtering is appropriate when the signal and noise exhibit a highly nonstationary behavior. In this case, traditional shift-invariant neighborhood filtering operators fail to provide adequate filtering, and a simple but optimal spatially varying approach can produce better results. We will further discuss this point later on.

4.1 Optimal fixed-mask filtering

4.1.1 The case of union noise

Here,

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

and

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

As before, *supp* $1(C(z))$ stands for the subset of B where $C(z)$ is true, and $T_X(\cdot)$ stands for the capacity functional of the DRS X. We have the following result.

Proposition 1 *Under the expected symmetric set difference measure, the optimal fixed intersection mask, W, for filtering out independent union noise is given by*

$$W = supp\ 1\left(T_N(\{z\})\,[1 - T_X(\{z\})] \leq T_X(\{z\})\right)$$

The corresponding minimum expected error achieved by such an optimal choice of W is given by

$$E(e^*) = \sum_{z \in B} min\left(T_X(\{z\}), T_N(\{z\})\,[1 - T_X(\{z\})]\right)$$

Proof: In this setting, it is easy to see that $E(e)$ is given by

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

The crucial observation here is that

$$E|X| = E\sum_{z \in B} 1(z \in X) = \sum_{z \in B} E\ 1(z \in X)$$

$$= \sum_{z \in B} Pr(z \in X) = \sum_{z \in B} Pr(X \cap \{z\} \neq \emptyset)$$

$$= \sum_{z \in B} T_X(\{z\})$$

Therefore,

$$E|X \cap W^c| = \sum_{z \in B} T_{X \cap W^c}(\{z\})$$

$$= \sum_{z \in B} T_X(\{z\} \cap W^c) = \sum_{z \in B} 1(z \in W^c) T_X(\{z\})$$

and, similarly

$$E|N \cap W \cap X^c| = \sum_{z \in B} T_{N \cap W \cap X^c}(\{z\})$$

$$= \sum_{z \in B} T_{N \cap X^c}(\{z\} \cap W) = \sum_{z \in B} 1(z \in W) T_{N \cap X^c}(\{z\})$$

Now

$$T_{N \cap X^c}(\{z\}) = Pr(z \in N \cap X^c) = Pr(z \in N, z \in X^c)$$
$$= Pr(z \in N) Pr(z \in X^c) = T_N(\{z\})\,[1 - T_X(\{z\})]$$

The last identity is due to the independence of X, N. Hence,

$$E|N \cap W \cap X^c| = \sum_{z \in B} 1(z \in W) T_N(\{z\})\,[1 - T_X(\{z\})]$$

and

$$E(e) = \sum_{z \in B} [1(z \in W^c) T_X(\{z\}) + \ 1(z \in W) T_N(\{z\})\,[1 - T_X(\{z\})]]$$

From which, by inspection, we obtain the claimed optimality results. □

It is interesting to compare the above optimality condition with that of standard Wiener filtering. What it says is that if the effective union noise "power" at z is less than or equal to the signal "power" at z, then filtering should retain the input value at point z, otherwise it should reject it. This is intuitively appealing, and highly reminiscent of a form of binary Wiener filtering. Also notice that all that is required for the design of the optimal W is just the first-order statistics of the signal and the noise, i.e. $T_X(\{z\})$ and $T_N(\{z\})$ for all $z \in B$. These can be efficiently and accurately estimated from training samples of X and N respectively.

If the first-order statistics of the signal and the noise are spatially invariant, then, obviously, the optimal intersection mask is either B ("all pass"), or $\emptyset$ ("reject all"). This case is clearly not interesting. 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. Let us illustrate this point by using a (rather simplistic) artificial example. Consider figure (1). It depicts a realization of a DRS which features a prominent periodic vertical line structure (throughout the sequence of figures in this work sets are graphically depicted by the subset of black points in an image). Figure (2) depicts a degraded version of the same image, obtained by taking the union of the DRS realization of figure (1) (the "signal") with an independent realization of another DRS (the "noise"). Figure (3) depicts the restored image, obtained by intersecting the DRS realization of figure (2), with the "optimal" intersection mask, computed by using the optimality condition of Proposition 1, and substituting *estimates* of the pixel hitting probabilities in place of the true probabilities. These estimates have been obtained by means of simple counting of pixel hitting events over a collection of learning samples of the signal and the noise. In this setting, shift-invariant filtering would not work well. Also, shape-sensitive filtering would not do, because the signal and the noise consist of replicas of the same elementary pattern, namely a square of side 5 pixels. A potentially big gain in quality of restoration rests exactly with proper exploitation of the non-stationary nature of the signal.

4.1.2 The case of intersection noise

Here,

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

and

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

This is the "dual" of the case of union noise. One can simply take the complement of all the sets and operations involved, and apply the results of the previous section. This is clear, because

$$d(X, \widehat{X}) = d(X^c, (\widehat{X})^c)$$

and, thus, minimizing $Ed(X, \widehat{X})$ is the same as minimizing $Ed(X^c, (\widehat{X})^c)$, and

$$((X \cap N) \cup W)^c = (X^c \cup N^c) \cap W^c$$

Therefore, employing proposition 1, the optimal f^W corresponds to the following choice of W^c

$$W^c = supp\ 1\left(T_{N^c}(\{z\})\left[1 - T_{X^c}(\{z\})\right] \leq T_{X^c}(\{z\})\right)$$

and the corresponding minimum expected error achieved by such an optimal choice of W^c is given by

$$E(e^*) = \sum_{z \in B} min\left(T_{X^c}(\{z\}), T_{N^c}(\{z\})\left[1 - T_{X^c}(\{z\})\right]\right)$$

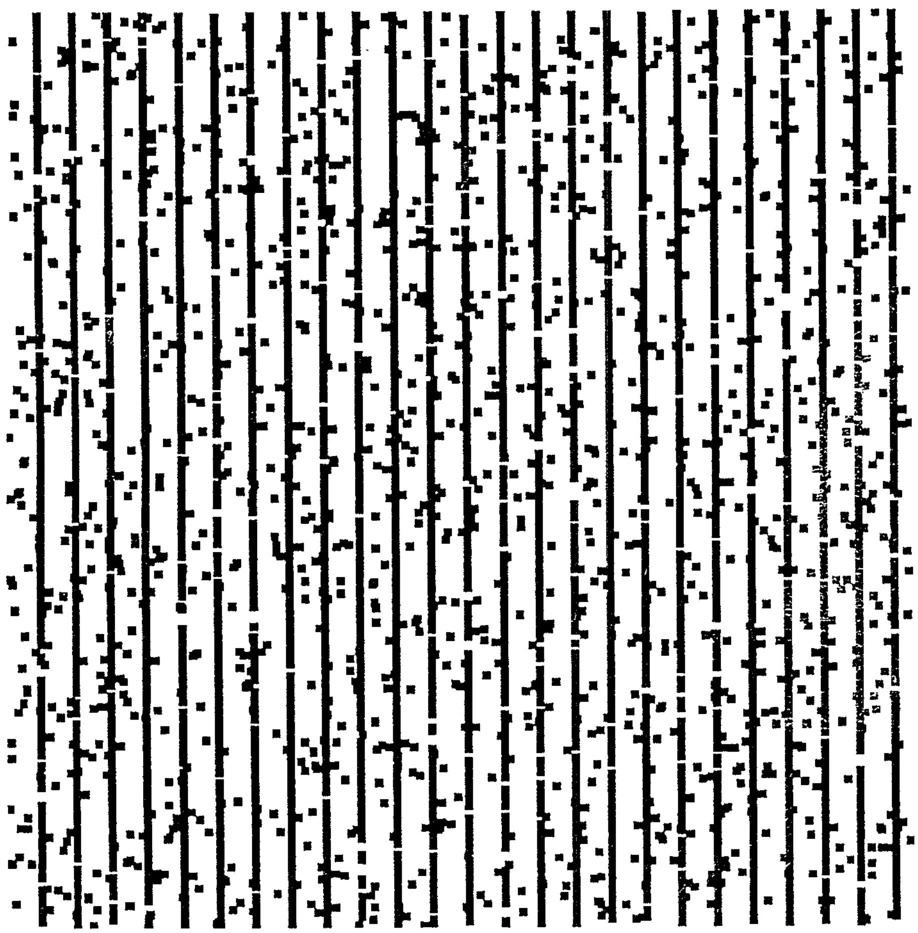

Figure 1: A realization of a non-stationary DRS, X.

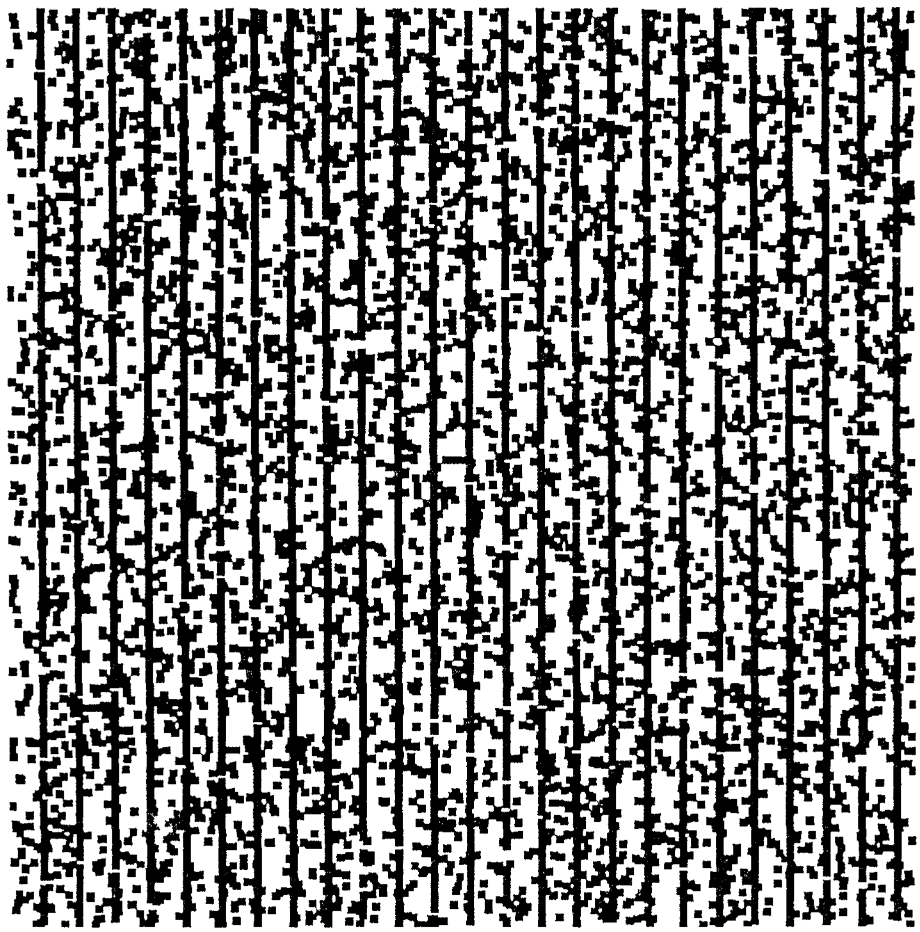

Figure 2: The result of taking the union of the DRS realization of figure (1) with an independent realization of the noise DRS, N.

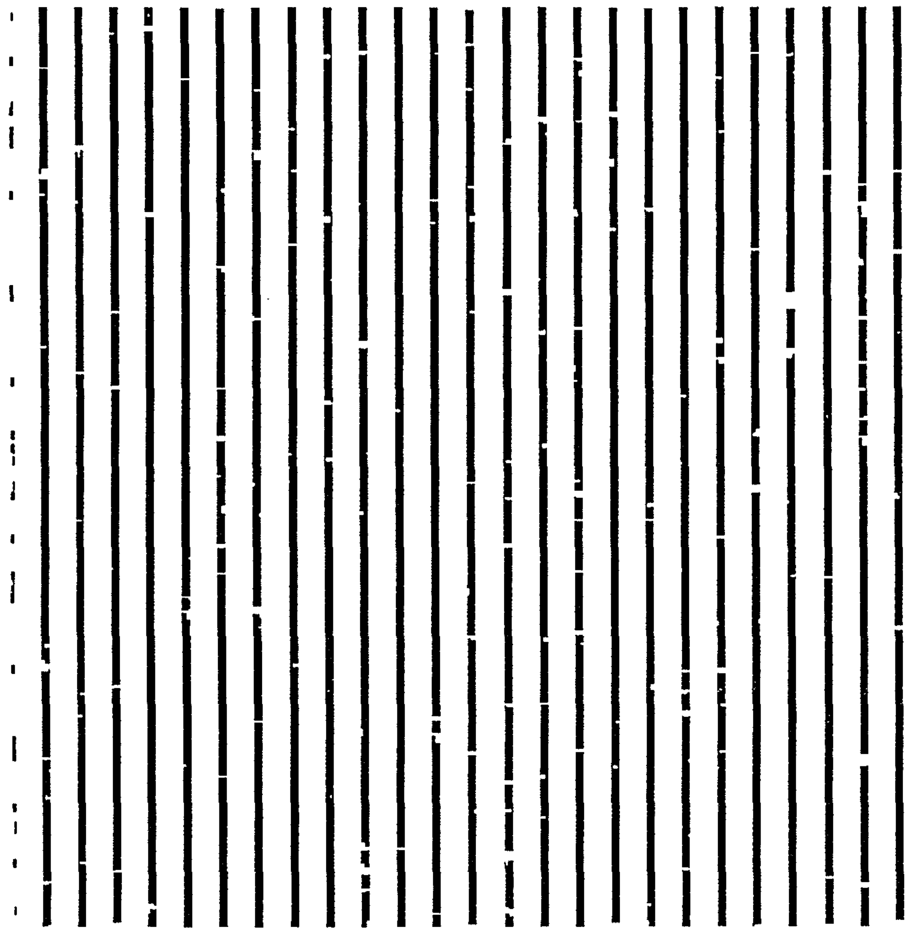

Figure 3: Restored image, obtained by filtering the DRS realization of figure (2) using the "optimal" intersection mask.

Noting that, for singletons $\{z\}$

$$T_{X^c}(\{z\}) = 1 - T_X(\{z\})$$

we obtain the following result.

Proposition 2 *Under the expected symmetric set difference measure, the optimal fixed "fill" mask, W, for independent intersection noise is given by*

$$W = supp\ 1\left([1 - T_N(\{z\})]\, T_X(\{z\}) > 1 - T_X(\{z\})\right)$$

The corresponding minimum expected error achieved by such an optimal choice of W is given by

$$E(e^*) = \sum_{z \in B} min\left(1 - T_X(\{z\}), [1 - T_N(\{z\})]\, T_X(\{z\})\right)$$

Observe that, once more, the result is intuitively appealing.

4.1.3 The composite problem: union/intersection noise

Here,

$$g(X, N_1, N_2) = (X \cap N_1) \cup N_2$$

where X, N_1, N_2 are assumed to be mutually independent, and

$$f(Y) = f_{W_2}^{W_1}(Y) = (Y \cap W_2) \cup W_1 = ((((X \cap N_1) \cup N_2) \cap W_2) \cup W_1), \ \textit{for some } W_1, W_2, \textit{ both in } \Sigma(B)$$

We have the following proposition.

Proposition 3 *Under the expected symmetric set difference measure, the optimal fixed pair of masks, (W_1, W_2), is given by*

$$W_2 = supp\ 1\left(T_X(\{z\}) > max(T_1(\{z\}), T_2(\{z\}))\right)$$

$$W_1 = supp\ 1\left(T_2(\{z\}) \leq min(T_X(\{z\}), T_1(\{z\}))\right)$$

whereas, the associated minimum expected 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

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

and

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

Proof: 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,

$$E(e) = Ed(X, \widehat{X}) = E\left|X \cap (N_1^c \cup W_2^c) \cap [(N_2^c \cap W_1^c) \cup W_2^c]\right| + E\left|(X^c \cap N_2 \cap W_2) \cup (X^c \cap W_1)\right|$$

Consider the first term

$$E\,|X\cap(N_1^c\cup W_2^c)\cap[(N_2^c\cap W_1^c)\cup W_2^c]| = \sum_{z\in B} Pr(z\in X\cap(N_1^c\cup W_2^c)\cap[(N_2^c\cap W_1^c)\cup W_2^c])$$

$$(by\ indep.\ of\ X, N_1, N_2)\ = \sum_{z\in B} Pr(z\in X)\,Pr(z\in N_1^c\cup W_2^c)\,Pr(z\in(N_2^c\cap W_1^c)\cup W_2^c)$$

$$= \sum_{z\in B} T_X(\{z\})\,[1(z\in W_2^c)+1(z\in W_2)\,(1-T_{N_1}(\{z\}))]\,[1(z\in W_2^c)+1(z\in W_2)1(z\in W_1^c)\,(1-T_{N_2}(\{z\}))]$$

Next, consider the second term of the expected error

$$E\,|(X^c\cap N_2\cap W_2)\cup(X^c\cap W_1)| = \sum_{z\in B} Pr(z\in(X^c\cap N_2\cap W_2)\cup(X^c\cap W_1))$$

$$= \sum_{z\in B} Pr(z\in X^c\cap[(N_2\cap W_2)\cup W_1]) = \sum_{z\in B} Pr(z\in X^c, z\in(N_2\cap W_2)\cup W_1)$$

$$(by\ independence\ of\ X, N_2)\ = \sum_{z\in B} Pr(z\in X^c)\,Pr(z\in(N_2\cap W_2)\cup W_1)$$

$$= \sum_{z\in B}(1-T_X(\{z\}))\,[1(z\in W_1)+1(z\in W_1^c)1(z\in W_2)T_{N_2}(\{z\})]$$

Therefore, the overall expression for the expected cost becomes

$$E(e) = Ed(X,\hat{X}) = \sum_{z\in B}\{T_X(\{z\})\,[1(z\in W_2^c)+1(z\in W_2)\,(1-T_{N_1}(\{z\}))]\cdot$$
$$[1(z\in W_2^c)+1(z\in W_2)1(z\in W_1^c)\,(1-T_{N_2}(\{z\}))]+$$
$$(1-T_X(\{z\}))\,[1(z\in W_1)+1(z\in W_1^c)1(z\in W_2)T_{N_2}(\{z\})]\}$$

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

$$(i)\ z\in W_1^c,\ z\in W_2^c,\ or\ (ii)\ z\in W_1^c,\ z\in W_2,\ or\ (iii)\ z\in W_1,\ z\in W_2$$

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. □

4.2 Optimal adaptive mask filtering

A drawback of the optimal filters which have been derived up to this point is that they are non-adaptive; no matter what the observation is, the filter is fixed. One would like to improve upon these filters by allowing for an adaptation of the mask using information extracted from the given input. The trade-off is an increase in design/implementation complexity. For simplicity, we only consider union noise *or* intersection noise.

4.2.1 The case of union noise

Assume that we are presented with a specific input, K, i.e. we are given that $Y = X \cup N = K$. One adaptation strategy is to incorporate this information into the cost function. This is done by considering the *conditional* expectation of $d(X, \widehat{X})$, conditioned on the given information. However, this does not lead to a closed-form solution for the optimal filter. Instead, we can condition on part of the available information. If we condition on the event $X \cup N \subseteq K$, i.e. $(X \cup N) \cap K^c = \emptyset$, then we can work out closed-form expressions for the optimal filter and the associated minimum error. In what follows E denotes conditional expectation, conditioned on $(X \cup N) \cap K^c = \emptyset$.

Proposition 4 *Given that $X \cup N \subseteq K$, the optimal intersection mask, W, for filtering out the noise component, N, is given by*

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

The corresponding minimum cost achieved by such an optimal choice of W is given by

$$E(e^*) = \frac{1}{(1 - T_X(K^c))(1 - T_N(K^c))}\cdot$$

$$\sum_{z \in K} min\left\{[1 - T_X(K^c \cup \{z\})]\,[T_N(K^c \cup \{z\}) - T_N(K^c)]\,,[T_X(K^c \cup \{z\}) - T_X(K^c)]\,[1 - T_N(K^c)]\right\}$$

Observe that for $K^c = \emptyset$ (i.e. $K = B$, no information available about the input) the formulas above reduce to the ones for the non-adaptive case, as they should (note that $T_Z(\emptyset) = 0$, for all DRS's Z).

Proof: The cost again breaks down to

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

Now,

$$E|X \cap W^c| = E\sum_{z \in B} 1\,(z \in X \cap W^c)$$

$$= \sum_{z \in B} E1\,(z \in X \cap W^c) = \sum_{z \in B} Pr\,(z \in X \cap W^c \mid (X \cup N) \cap K^c = \emptyset)$$

$$= \sum_{z \in K} Pr\,(z \in X \cap W^c \mid (X \cup N) \cap K^c = \emptyset)$$

$$= \sum_{z \in K} \frac{Pr\,(z \in X \cap W^c, (X \cup N) \cap K^c = \emptyset)}{Pr\,((X \cup N) \cap K^c = \emptyset)}$$

Observe that

$$Pr\,((X \cup N) \cap K^c = \emptyset) = Pr\,(X \cap K^c = \emptyset, N \cap K^c = \emptyset)$$

$$(by\ independence\ of\ X, N)\ = (1 - T_X(K^c))(1 - T_N(K^c))$$

also

$$Pr\,(z \in X \cap W^c, (X \cup N) \cap K^c = \emptyset) = Pr\,(X \cap W^c \cap \{z\} \neq \emptyset, X \cap K^c = \emptyset, N \cap K^c = \emptyset)$$

$$(by\ independence\ of\ X,N)\ = Pr(X \cap (W^c \cap \{z\}) \neq \emptyset, X \cap K^c = \emptyset)\, Pr(N \cap K^c = \emptyset)$$
$$= (T_X(K^c \cup (W^c \cap \{z\})) - T_X(K^c))\,(1 - T_N(K^c))$$

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

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

We have seen that the denominator is equal to

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

Whereas the nominator

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

Therefore, the second term of the expected cost becomes

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

and the overall expected cost becomes

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

From which the claimed results follow by inspection. □

4.2.2 The case of intersection noise

If we condition 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, by duality.

Proposition 5 *Given that $X^c \cup N^c \subseteq K^c$ the optimal union ("fill") mask, W, for filtering out the intersection noise component, N, is specified by*

$$W^c = K^c \cap supp\, 1\left([1 - T_{X^c}(K \cup \{z\})]\,[T_{N^c}(K \cup \{z\}) - T_{N^c}(K)] \leq [T_{X^c}(K \cup \{z\}) - T_{X^c}(K)]\,[1 - T_{N^c}(K)]\right)$$

The corresponding minimum cost achieved by such an optimal choice of W is given by

$$E(e^*) = \frac{1}{(1 - T_{X^c}(K))(1 - T_{N^c}(K))}\cdot$$

$$\sum_{z \in K^c} min\left\{[1 - T_{X^c}(K \cup \{z\})]\,[T_{N^c}(K \cup \{z\}) - T_{N^c}(K)]\,,[T_{X^c}(K \cup \{z\}) - T_{X^c}(K)]\,[1 - T_{N^c}(K)]\right\}$$

Again, for $K^c = B$ (i.e. $K = \emptyset$, no information available about the input) the formulas above reduce to the ones for the non-adaptive case, as they should.

Remark: For $K \neq$ *singleton*

$$T_{X^c}(K) \neq 1 - T_X(K)$$

4.3 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, and its analytical tractability. This model has received considerable attention in the literature (see [25] 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.

Let H be a convex[1] subset of B', $|B'| << |B|$, which contains the origin. In the terminology of Mathematical Morphology[2], H is a convex *structuring element*. In the discrete case the notion of size can be formalized via the operation of *Minkowski set addition*:

$$A \oplus C = \bigcup_{z \in C} A_z$$

where A_z is the translate of A by the vector z. Define

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

[1]In digital topology [15, 23, 10], the *convex hull* of a bounded set, $H \subset \mathcal{Z}^2$, is defined as the intersection of the convex hull of H in the topology of $\mathcal{R}^2$, with $\mathcal{Z}^2$. A bounded set, $H \subset \mathcal{Z}^2$, is *convex* if it is identical to its convex hull.

[2]Refer to [23] for a thorough exposure to the principles of Mathematical Morphology, and the beginning of section 5 for definitions of set dilation/erosion, along with some brief introductory comments.

Definition 4 *Let Ψ be a generalized Bernoulli lattice process 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 $\{G_1, G_2, \cdots\}$ be a set of nonempty, convex i.i.d. DRS's on B', each given by $G_i = R_i H$, where $\{R_1, R_2, \cdots\}$ 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 $\Psi = \{y_1, y_2, \cdots\}$. Then X will be called a **Discrete Radial Boolean Random Set (DRBRS)** *, with parameters (λ_s, H, f_R), and will be denoted by (λ_s, H, f_R)-DRBRS. The points $\{y_1, y_2, \cdots\}$ will be called* the germs*, and the DRS's $\{G_1, G_2, \cdots\}$ will be called* the primary grains *of the DRBRS X.*

Remark: For brevity, we assume that for the purposes of this subsection the result of a $\oplus$ operation is automatically restricted to B. Also, ^{c} stands for complement with respect to B. A sample realization of a DRBRS is given in figure 6.

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

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

Observe that for $z \in K$, $d^H(z, K) = 0$, since H contains the origin. We remark that $d^H(z, K)$, as defined above, is a digital uniform step metric, which is a generalization of the digital Housdorff metric. With this notation in place, and employing some geometric arguments, it can be shown that

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

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 1-5, and obtain the optimal filter as a function of the signal and noise parameters.

5 Optimal Morphological Filters

The theory of Mathematical Morphology has been developed mainly by Serra [23, 9], Matheron [19], and their collaborators, during the 70's and early 80'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 non-linear. These mappings are defined in terms of a *structuring element*, a "small" primitive shape (set of points) which interacts with the input image to transform it, and, in the process, extract useful information about its geometrical and topological structure. Let

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

The dilation of a set $Y \subset \mathcal{Z}^2$ by a structuring element W is defined as[3]

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

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

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

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 $\mathcal{Z}^2$. Two fundamental composite Morphological operators are opening and closing. The opening, $Y \circ W$, of a set $Y \subset \mathcal{Z}^2$ by a structuring element W, is defined as

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

Similarly, the closing, $Y \bullet W$, of a set $Y \subset \mathcal{Z}^2$ 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.

5.1 Some results on constrained optimality, or, why Morphology is popular.

Complex Morphological operators can be constructed by composing more elementary mappings. For example, the family of Alternating Sequential Filters (ASF's) is constructed by alternating openings and closings with structuring elements of increasing size. Morphological operators are very flexible, mainly because of the freedom to choose the structuring element(s), to meet specified criteria. Among other things, Morphological operators 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 [22, 5, 8, 7, 3, 4]. 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

[3]Here we follow the original definitions of Serra [23]. In his work the symbol $\oplus$ stands for Minkowski set addition, and the symbol $\ominus$ stands for Minkowski set subtraction.

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 $\Sigma(B)$:*

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

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 (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 MAP estimate of X on the basis of Y.

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. 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. 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 $\Sigma(B)$:*

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

Furthermore, assume that the observable DRS is $Y = X \cap N$, where N is a homogeneous Bernoulli lattice process of intensity r, which is independent of X. Then $Y \bullet W$ is the MAP estimate of X on the basis of Y.

Proof: By definition,

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

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

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

$$= arg\ max_{K \in C_W(B),\ K \supseteq Y} \left\{ (1-r)^{|K|} \right\} = arg\ min_{K \in C_W(B),\ K \supseteq Y} \{|K|\}$$

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

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

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 $O_{W_1, \cdots W_M}(B)$ denote the collection of all subsets K of B which can be expressed 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 $\Sigma(B)$:

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

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

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

Theorem 5 *Let $C_{W_1,\cdots W_M}(B)$ denote the collection of all subsets K of B which can be expressed 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 $\Sigma(B)$:

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

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

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

In general, if we assume that X satisfies some arbitrary (not necessarily Morphological) smoothness conditions, i.e. $X \in \mathcal{S}$, 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

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

where $d(Y, K)$ is the area of the symmetric set difference distance between Y and K. In other words, $\widehat{X}_{MAP}(Y)$ 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 programme 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 which 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.

5.2 Optimizing a single structuring element

In the case of union noise we constrain the mapping f to be

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

Similarly, in the case of intersection noise we constrain the mapping f to be

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

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,

$$Y \ominus W^s = \{z \in B \mid W_z \subseteq Y\}$$

Figure 4: Some structuring elements that can be used in a "gap-filling" mode.

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 figure (4) (a cross indicates the location of the origin). An example is given in figure (5). Figure (5a) depicts a test image, while figure (5b) depicts a degraded version of the test image, obtained by intersecting it with the set of points which make up a realization of a Bernoulli random field, of intensity 0.9. Figure (5c) depicts the estimate, $\widehat{X} = [(X \cap N) \ominus W^s] \cup (X \cap N)$ where X is the original test image depicted in (5a), N is the set of points of the Bernoulli field, $X \cap N$ is the observation depicted in (5b), and W is the leftmost of the structuring elements which appear in figure (4). 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, the wider the gaps that can be (partially) filled by an erosion with the given structuring element.

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[4], 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

[4] See [13] for an account of such an approach, when the intersection mask, N, is a deterministic, regularly spaced grid, which undersamples the observation.

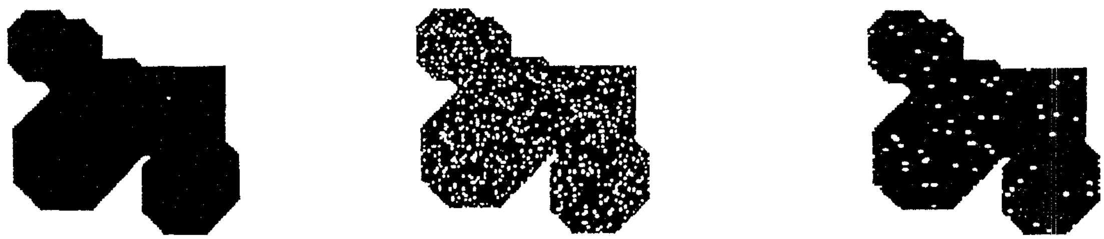

Figure 5: (a) Original image, (b) Intersection of the image in (a) with a Bernoulli random field, (c) Restored image.

the optimal filters.

5.2.1 The case of intersection noise

Here[5],

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

and

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

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

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

$B \backslash \partial B$ 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.

[5]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.

Proposition 6 *Under the assumption of mutual independence of the signal and noise DRS's, X, N, the value of the expected error, $E(e) = Ed(X, \widehat{X})$, incurred when X is estimated by $\widehat{X} = [(X \cap N) \ominus W^s] \cup (X \cap N)$, is given by*

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

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

Proof: 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

$$\Gamma_{X,n}(K_0; K_1, \cdots, K_n) \triangleq Pr(X \cap K_0 = \emptyset \ , \ X \cap K_1 \neq \emptyset, \cdots, X \cap K_n \neq \emptyset)$$

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

$$\Gamma_{X,n}(K_0; K_1, \cdots, K_n) = \Gamma_{X,n-1}(K_0; K_1, \cdots, K_{n-1}) - \Gamma_{X,n-1}(K_0 \cup K_n; K_1, \cdots, K_{n-1})$$

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

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

Let us first consider the second term.

$$E|\widehat{X} \setminus X| = E|\widehat{X} \cap X^c| = E\,|([(X \cap N) \ominus W^s] \cup (X \cap N)) \cap X^c|$$

$$= E\,|([(X \cap N) \ominus W^s] \cap X^c) \cup (X \cap N \cap X^c)| = E\,|[(X \cap N) \ominus W^s] \cap X^c|$$

Now, since

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

The last expression is equal to

$$E|(X \ominus W^s) \cap (N \ominus W^s) \cap X^c|$$

$$= E \sum_{z \in B \setminus \partial B} 1(z \in (X \ominus W^s) \cap (N \ominus W^s) \cap X^c) = \sum_{z \in B \setminus \partial B} E\ 1(z \in (X \ominus W^s) \cap (N \ominus W^s) \cap X^c)$$

$$= \sum_{z \in B \setminus \partial B} Pr(z \in (X \ominus W^s) \cap (N \ominus W^s) \cap X^c) = \sum_{z \in B \setminus \partial B} Pr(z \in (X \ominus W^s) \cap X^c) \ , \ z \in N \ominus W^s)$$

$$(\textit{by independence of } X, N) = \sum_{z \in B \setminus \partial B} Pr(z \in (X \ominus W^s) \cap X^c) Pr(z \in N \ominus W^s)$$

$$= \sum_{z \in B \setminus \partial B} Pr(W_z \subseteq X \ , \ z \in X^c) Pr(W_z \subseteq N) = \sum_{z \in B \setminus \partial B} Pr(X^c \cap W_z = \emptyset \ , \ X^c \cap \{z\} \neq \emptyset) Pr(N^c \cap W_z = \emptyset)$$

The first term of the total expected error

$$E|X\backslash\widehat{X}| = E|X\cap(\widehat{X})^c| = E\,|X\cap([(X\cap N)\ominus W^s]\cup(X\cap N))^c|$$

$$= E\,|X\cap[(X\cap N)\ominus W^s]^c\cap(X\cap N)^c| = E\,|[X\cap(X\cap N)^c]\cap[(X\cap N)\ominus W^s]^c|$$

$$= E\,|[X\cap(X^c\cup N^c)]\cap[(X\cap N)\ominus W^s]^c| = E\,|[(X\cap X^c)\cup(X\cap N^c)]\cap[(X\cap N)\ominus W^s]^c|$$

$$= E\,|X\cap N^c\cap[(X\cap N)\ominus W^s]^c| = E\,|X\cap N^c\cap[(X\ominus W^s)\cap(N\ominus W^s)]^c|$$

$$= E\,|X\cap N^c\cap[(X\ominus W^s)^c\cup(N\ominus W^s)^c]| = E\,|(X\cap N^c\cap(X\ominus W^s)^c)\cup(X\cap N^c\cap(N\ominus W^s)^c)|$$

$$= E\,|X\cap N^c\cap(X\ominus W^s)^c| + E\,|X\cap N^c\cap(N\ominus W^s)^c| - E\,|X\cap N^c\cap(X\ominus W^s)^c\cap(N\ominus W^s)^c|$$

Now,

$$E\,|X\cap N^c\cap(X\ominus W^s)^c| = E\,|(X\cap(X\ominus W^s)^c)\cap N^c|$$

$$= \sum_{z\in B\backslash\partial B} Pr(z\in X\ ,\ \neg(W_z\subseteq X))Pr(z\in N^c) = \sum_{z\in B\backslash\partial B} Pr(X^c\cap\{z\}=\emptyset\ ,\ X^c\cap W_z\neq\emptyset)Pr(N^c\cap\{z\}\neq\emptyset)$$

where $\neg$ denotes logical negation. Also,

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

$$= \sum_{z\in B\backslash\partial B} Pr(X^c\cap\{z\}=\emptyset)Pr(N^c\cap\{z\}\neq\emptyset\ ,\ N^c\cap W_z\neq\emptyset)$$

And

$$E\,|X\cap N^c\cap(X\ominus W^s)^c\cap(N\ominus W^s)^c| = E\,|(X\cap(X\ominus W^s)^c)\cap(N^c\cap(N\ominus W^s)^c)|$$

$$= \sum_{z\in B\backslash\partial B} Pr(z\in X\ ,\ \neg(W_z\subseteq X))Pr(z\in N^c\ ,\ \neg(W_z\subseteq N))$$

$$= \sum_{z\in B\backslash\partial B} Pr(X^c\cap\{z\}=\emptyset\ ,\ X^c\cap W_z\neq\emptyset)Pr(N^c\cap\{z\}\neq\emptyset\ ,\ N^c\cap W_z\neq\emptyset)$$

Therefore, putting everything together,

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

$$= \sum_{z\in B\backslash\partial B}\{Pr(X^c\cap\{z\}=\emptyset\ ,\ X^c\cap W_z\neq\emptyset)Pr(N^c\cap\{z\}\neq\emptyset)$$

$$+Pr(X^c\cap\{z\}=\emptyset)Pr(N^c\cap\{z\}\neq\emptyset\ ,\ N^c\cap W_z\neq\emptyset)$$

$$-Pr(X^c\cap\{z\}=\emptyset\ ,\ X^c\cap W_z\neq\emptyset)Pr(N^c\cap\{z\}\neq\emptyset\ ,\ N^c\cap W_z\neq\emptyset)$$

$$+Pr(X^c\cap W_z=\emptyset\ ,\ X^c\cap\{z\}\neq\emptyset)Pr(N^c\cap W_z=\emptyset)\}$$

$$= \sum_{z \in B \setminus \partial B} \{\Gamma_{X^c,1}(\{z\}; W_z)(1 - Q_{N^c}(\{z\}))$$

$$+ Q_{X^c}(\{z\})\Gamma_{N^c,2}(\emptyset; \{z\}, W_z)$$

$$- \Gamma_{X^c,1}(\{z\}; W_z)\Gamma_{N^c,2}(\emptyset; \{z\}, W_z)$$

$$+ \Gamma_{X^c,1}(W_z; \{z\})Q_{N^c}(W_z)\}$$

$$= \sum_{z \in B \setminus \partial B} \{(Q_{X^c}(\{z\}) - Q_{X^c}(\{z\} \cup W_z))(1 - Q_{N^c}(\{z\}))$$

$$+ Q_{X^c}(\{z\})(1 - Q_{N^c}(\{z\}) - Q_{N^c}(W_z) + Q_{N^c}(\{z\} \cup W_z))$$

$$- (Q_{X^c}(\{z\}) - Q_{X^c}(\{z\} \cup W_z))(1 - Q_{N^c}(\{z\}) - Q_{N^c}(W_z) + Q_{N^c}(\{z\} \cup W_z))$$

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

From which, after some manipulations, we obtain

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

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

□

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[6] assumption that both X, and N, are obtained by sampling *stationary* random closed sets (RACS) [19], 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 RACS, 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 remarks are in place. The first term, $Q_{X^c}(\{\bar{0}\})(1 - Q_{N^c}(\{\bar{0}\}))$, of the probability of pixel error, $P_{pixel\ error}(W)$, 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 non-negative functions of W (it can be easily shown that the generating functional of an arbitrary DRS is constrained to be non-increasing). 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

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

achieves the optimal trade-off 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 [16] 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.

5.2.2 The case of union noise

Here,

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

and

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

We can proceed directly, as in the case of intersection noise, to derive the optimality results from scratch. However, we can also obtain the same results by appealing to duality. In particular, since

$$(\widehat{X})^c = ([(X \cup N) \oplus W^s] \cap (X \cup N))^c = [(X \cup N) \oplus W^s]^c \cup (X \cup N)^c$$

$$= [(X \cup N)^c \ominus W^s] \cup (X^c \cap N^c) = [(X^c \cap N^c) \ominus W^s] \cup (X^c \cap N^c)$$

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 with their complements, X^c, N^c. Thus we have the following result.

Proposition 7 *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 RACS, 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 RACS, 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 making some further assumptions concerning the signal and noise models. We will demonstrate this by example. Suppose that X, N, are both DRBRS's, of constant intensity, $p_X = 1 - q_X$, $p_N = 1 - q_N$, and with deterministic primary grains, H_X, H_N, respectively, which are both convex and contain the origin. Then it can be shown that

$$Q_X(K) = q_X^{|K \oplus H_X^s|}, \quad Q_N(K) = q_N^{|K \oplus H_N^s|}$$

Let $\mathcal{W}$ denote the collection of candidate structuring elements over which we intend to optimize. It can be shown that under the condition

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

the second term of the sum for the probability of pixel error is zero. In loose terms, this condition amounts to requiring H_X 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 figure 4, and H_X is a square of side 3 pixels which is 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

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

which can be written as

$$G(W) = q_X^{|W \oplus H_X^s|} q_N^{|W \oplus H_N^s|} \left(1 - q_N^{|H_N^s| - |H_N^s \cap (W \oplus H_N^s)|}\right)$$

From this last expression, it is easy to see that

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

Thus, 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

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

This elimination can translate to a significant decrease of search complexity. For example, if $\mathcal{W}$ is the collection of structuring elements depicted in figure 4, and H_X is a square of side 3 pixels which is centered at the origin, then it suffices to optimize over the two leftmost structuring elements.

The same result can be worked out when X, N, are both DRBRS's of constant intensity and primary grains of random size. If both p_X and p_N are sufficiently small (i.e. both q_X and q_N are sufficiently close to 1), then it can be shown that

$$Q_X(K) \cong q_X^{E|K \oplus R_X H_X^s|}, \quad Q_N(K) \cong q_N^{E|K \oplus R_N H_N^s|}$$

where the expectation is taken with respect to the pmf of R_X, R_N, respectively. In the following we will assume that these approximations are valid, and that

$$Pr(R_X = 0) = Pr(R_N = 0) = 0$$

By virtue of the fact that $\oplus$ is a commutative, associative and increasing operator, i.e.

$$A \oplus C = C \oplus A$$

$$A \oplus (C \oplus D) = (A \oplus C) \oplus D$$

$$A \subseteq C \Rightarrow A \oplus D \subseteq C \oplus D$$

it can be shown that if

$$H = H_1 \oplus H_2 \oplus \cdots \oplus H_n$$

and *one* of the H_i's satisfies

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

then so does H. In particular, it can be shown that if H_X satisfies

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

then so does rH_X, $r = 1, 2, \ldots$, which in turn implies that the second term of the sum of the probability of pixel error is zero. Therefore, the optimal $W \in \mathcal{W}$ should maximize the third term of the sum, which can be written as

$$G(W) = q_X^{E|W \oplus R_X H_X^s|} q_N^{E|W \oplus R_N H_N^s|} \left(1 - q_N^{E|R_N H_N^s| - E|R_N H_N^s \cap (W \oplus R_N H_N^s)|}\right)$$

from which it can again be seen that

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

Thus, assuming that both q_X and q_N are sufficiently close to 1, it is only necessary to optimize over the subcollection

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

5.3 Extensions to optimal filters with multiple structuring elements, and relation to basis decomposition.

In this section, we discuss important extensions to some of the optimality results which have been obtained so far. For simplicity, we restrict our attention to the case of intersection noise, although, by duality, similar extensions hold for the case of union noise.

5.3.1 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 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 6, and with some patience, we can obtain similar optimality results for the case of multiple structuring elements. For example, we state the following, without proof.

Proposition 8 *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*

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

is given by

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

Corollary 3 *Under the assumption of mutual independence of the signal and noise DRS's, X, N, assuming that X, N, are obtained by sampling stationary RACS, and that X is estimated by*

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

the optimal pair of structuring elements, W^1, W^2, is the one which minimizes the probability of pixel error

$$\begin{aligned} P_{pixel\ error}(W^1, W^2) = {} & Q_{X^c}(\{\bar{0}\})\,(1 - Q_{N^c}(\{\bar{0}\})) \\ &+ Q_{X^c}(W^1) Q_{N^c}(W^1) + Q_{X^c}(W^2) Q_{N^c}(W^2) \\ &+ Q_{X^c}(\{\bar{0}\} \cup W^1) Q_{N^c}(\{\bar{0}\} \cup W^1) + Q_{X^c}(\{\bar{0}\} \cup W^2) Q_{N^c}(\{\bar{0}\} \cup W^2) \\ &- 2Q_{X^c}(\{\bar{0}\} \cup W^1) Q_{N^c}(W^1) - 2Q_{X^c}(\{\bar{0}\} \cup W^2) Q_{N^c}(W^2) \\ &- Q_{X^c}(W^1 \cup W^2) Q_{N^c}(W^1 \cup W^2) - Q_{X^c}(\{\bar{0}\} \cup W^1 \cup W^2) Q_{N^c}(\{\bar{0}\} \cup W^1 \cup W^2) \\ &+ 2Q_{X^c}(\{\bar{0}\} \cup W^1 \cup W^2) Q_{N^c}(W^1 \cup W^2) \end{aligned}$$

Obviously, similar results can be obtained for more than two structuring elements.

5.3.2 Optimal increasing, shift-invariant filters with a basis constraint.

A surprising result, originally due to Matheron [19], and subsequently improved upon, and used by Maragos [17], and Dougherty et al., and Giardina [5, 6, 8, 7], is that a very large class of (linear and non-linear) shift-invariant operations can be decomposed into a union of erosions by suitable structuring elements.

Let $E = \mathcal{Z}^2$, 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 [19, 17].

Theorem 6 *[19] 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 7 *[17] 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 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. [6, 8, 7], 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 [3].

Strictly speaking, these theorems cannot be used with bounded domains. However, modulo some edge effects, which are proportional to the size of the largest basis element, they can be used.

In light of these theorems, let us review our results. For example, consider

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

which can also be written as

$$\widehat{X} = [(X \cap N) \ominus (W^1)^s] \cup [(X \cap N) \ominus (W^2)^s] \cup [(X \cap N) \ominus \{\bar{0}\}]$$

Therefore, under the usual assumptions for the signal and the degradation, Corollary 3 provides the means of finding the optimal filter, within the class of shift-invariant and increasing filters, whose basis consists of at most three structuring elements, one of which is constrained to be $\{\bar{0}\}$. This last restriction is natural[7]

[7]Because it assures that the overall filtering operation is *extensive*, i.e. its result contains the input. This property is obviously desirable under an intersection noise model.

for the case of intersection noise. With patience, similar results can be developed for larger basis sets. Relative to the *reduced basis search* and the *switching algorithm* of Dougherty et al. [8, 7], our work provides tractable *theoretical* formulas for the cost function, thus establishing a more rigorous statistical framework for the application of optimization algorithms. Our point is that, *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. The size of the expansion is usually determined from design and implementation considerations.

5.4 Experimental Results

In order to corroborate our theoretical results, we have designed a series of simulation experiments. One such experiment is described in this section.

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.

We also have to chose a collection of structuring elements, over which we will optimize. We consider the collection depicted in figure (4), and label the structuring elements $W_1, \cdots, W_4$, from left to right.

Figure (6) depicts a realization of the signal DRS, X, while figure (7) 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 (1) and (2), respectively. The results for the estimated probability of pixel error are tabulated in table (3). These have been computed using tables (1),(2), and the formula of Corollary 1. In table (3), the leftmost 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.

Figure (8) depicts another (independent) realization of X, while figure (9) depicts a realization of the observation, $Y = X \cap N$, obtained by intersecting the DRS realization of figure (8) with an independent realization of N. Figure (10) depicts the restored image, $\widehat{X} = (Y \ominus W_2^s) \cup Y$, where Y is the DRS realization of figure (9). This is the best possible restoration within the given family of structuring elements. For comparison purposes, figure (11) depicts the restored image, $\widetilde{X} = (Y \ominus W_4^s) \cup Y$, where Y is the DRS realization of figure (9). This is the worst (non-trivial) restoration within the given family of structuring elements.

These 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.

6 Conclusions

In this paper we have described two optimal digital binary image filtering strategies. Mask filtering is a simple, yet intuitive, 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 clearly superior when compared to fixed-mask filtering, 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. Generally speaking, fixed-mask filtering is suitable when the signal and noise DRS's are highly non-stationary, in which case optimal filters turn out being spatially varying, and traditional shift-invariant filters are very hard to optimize, and out of context. Adaptive mask filtering can be effective in both the stationary and nonstationary case.

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 stationary behavior. 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/Dougherty/Katz [12], which are based on the assumption that the signal and noise patterns are "non-interfering", and the results of Schonfeld/Goutsias [22], 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.

7 Appendix - Proof of Lemma 1

Uniqueness: Assume that the external decomposition formula holds. Look at the right hand side of the inversion formula.

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

$$\sum_{C \subseteq S} (-1)^{|C|} \sum_{D \subseteq S \setminus C} u(D) = \sum_{C \subseteq S} \sum_{D \subseteq S \setminus C} (-1)^{|C|} u(D) =$$

$$\sum_{D \subseteq S} \sum_{C \subseteq S \setminus D} (-1)^{|C|} u(D) = \sum_{D \subseteq S} u(D) \sum_{C \subseteq S \setminus D} (-1)^{|C|} = u(S)$$

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.

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

$$\sum_{S\subseteq A^c}\sum_{C\subseteq S}(-1)^{|C|}v((S\backslash C)^c) = \sum_{D\subseteq A^c}\sum_{C\subseteq A^c\backslash D}(-1)^{|C|}v(D^c) =$$

$$\sum_{D\subseteq A^c} v(D^c)\sum_{C\subseteq A^c\backslash D}(-1)^{|C|} = v((A^c)^c) = v(A)$$

As for the uniqueness part. □

References

[1] M. Aigner. *Combinatorial Theory.* Springer-Verlag, New York, 1979.

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

[3] 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.

[4] 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.

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

[6] 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.

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

[8] 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.

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

[10] John Goutsias and Chuanju Wen. Modeling Discrete Random Shapes: A Random Set Theory Approach. Technical Report JHU/ECE 90-13, The Johns Hopkins University, 1990. Submitted for publication: IEEE Trans. PAMI.

[11] John Goutsias and Chuanju Wen. Discrete random set models for shape synthesis and analysis. In *Proc. SPIE Vol. 1606 Visual Communications and Image Processing '91: Image Processing*, pages 174–185. Society for Optical Engineering, 1991.

[12] 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.

[13] R.M. Haralick, X. Zhuang, C. Lin, and J.S.J. Lee. The Digital Morphological Sampling Theorem. *IEEE Trans. Acoustics, Speech and Signal Processing*, 37(12):2067–2089, 1989.

[14] 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.

[15] T.Y. Kong and A. Rosenfeld. Digital topology: Introduction and survey. *Computer Vision, Graphics and Image Processing*, 48:357–393, 1989.

[16] 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.

[17] 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.

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

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

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

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

[22] 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.

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

[24] N.D. Sidiropoulos, J. Baras, and C. Berenstein. The structure of divisible discrete random sets and their randomized superpositions. Technical Report TR 91-54, Systems Research Center, University of Maryland, April 1991.

[25] D. Stoyan, W.S. Kendall, and J. Mecke. *Stochastic Geometry and its Applications.* Wiley, Berlin, 1987.

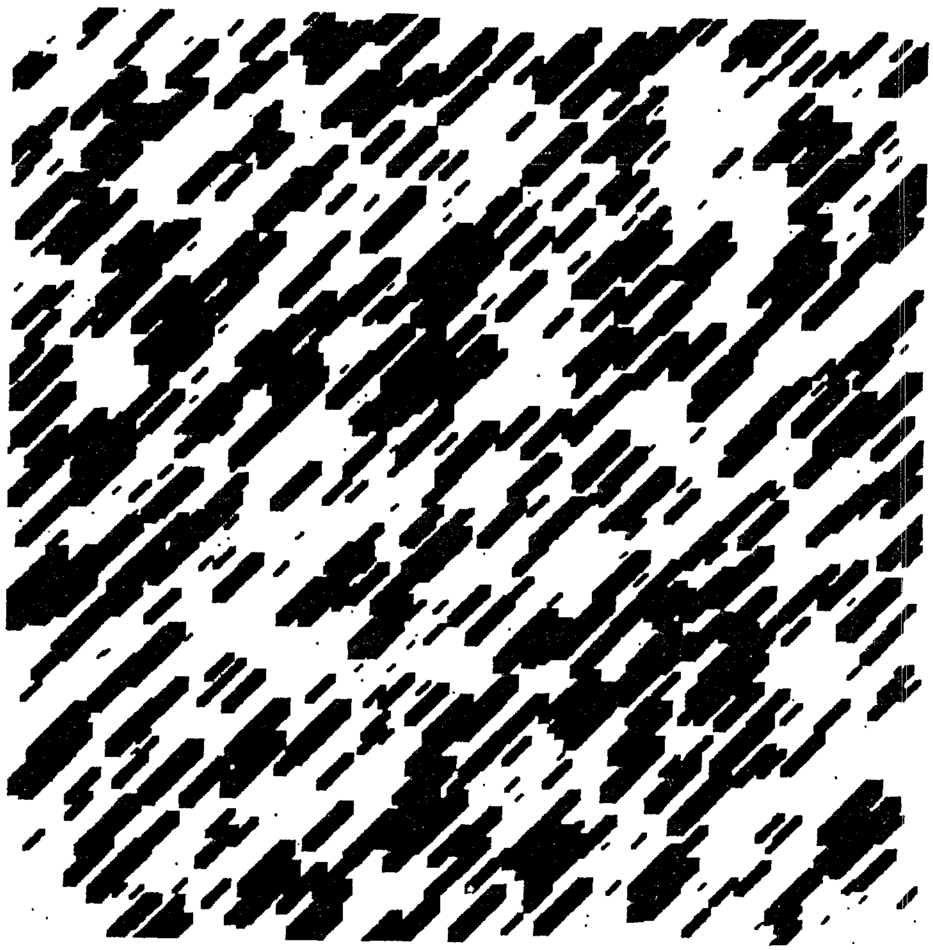

Figure 6: A realization of the signal DRS, X.

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

Table 1: Estimated values of the generating functional $Q_{X^c}(\cdot)$.

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

Table 2: Estimated values of the generating functional $Q_{N^c}(\cdot)$.

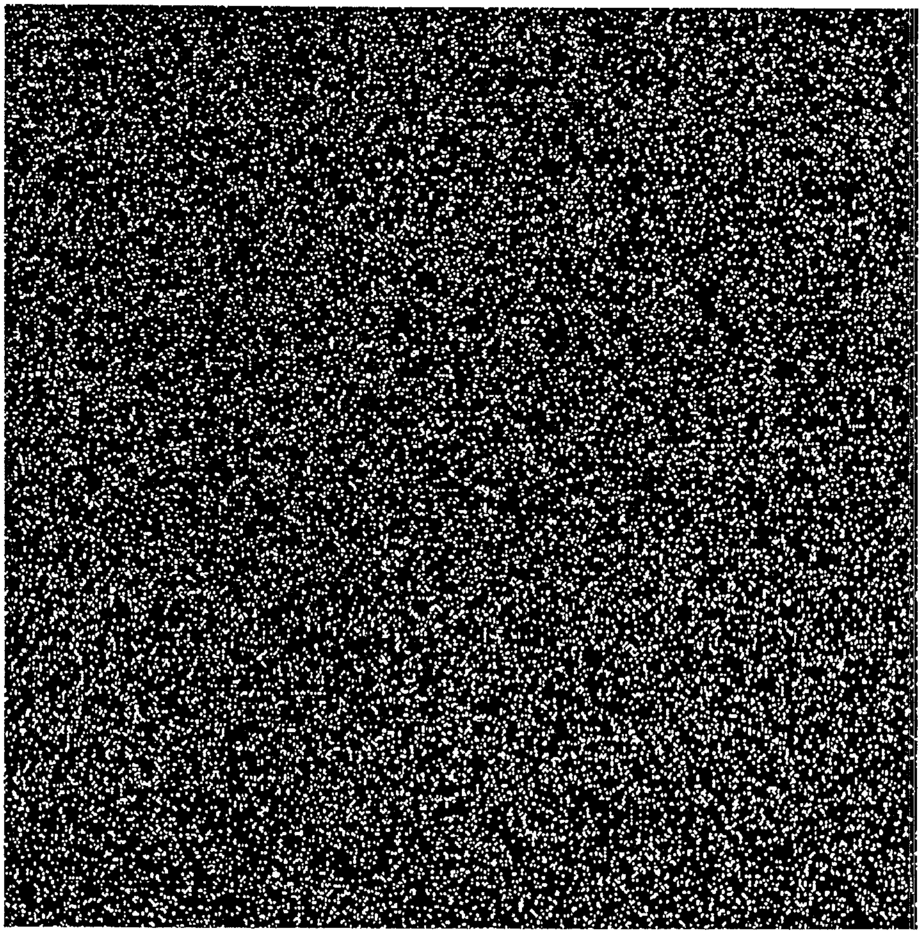

Figure 7: A realization of the noise DRS, N.

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

Table 3: Estimated values of the probability of pixel error.

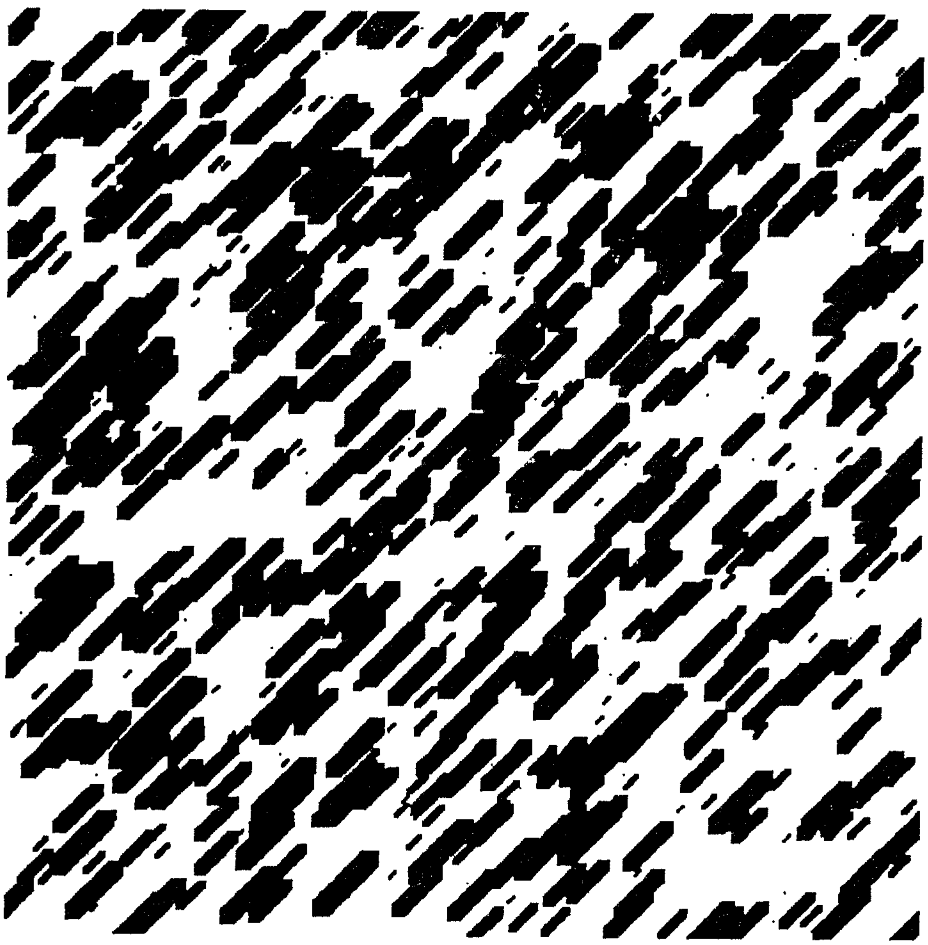

Figure 8: Another (independent) realization of the signal DRS, X.

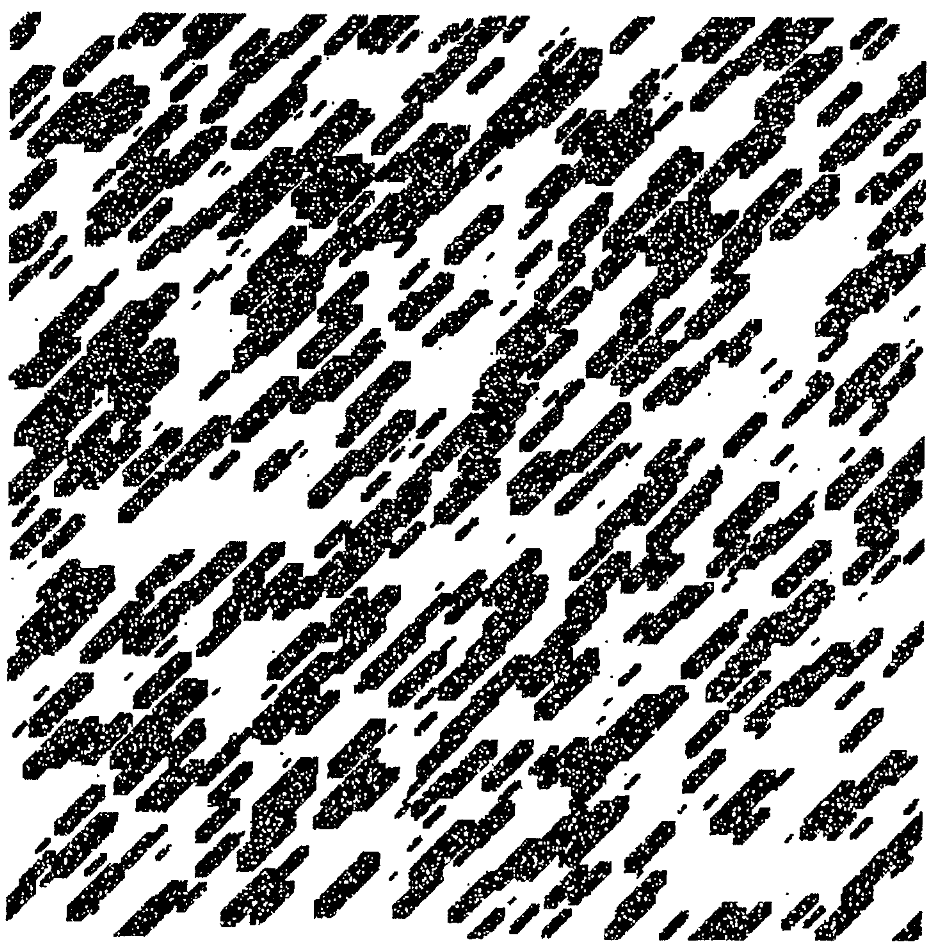

Figure 9: The result of intersecting the DRS realization of figure (8) with another (independent) realization of the noise DRS, N.

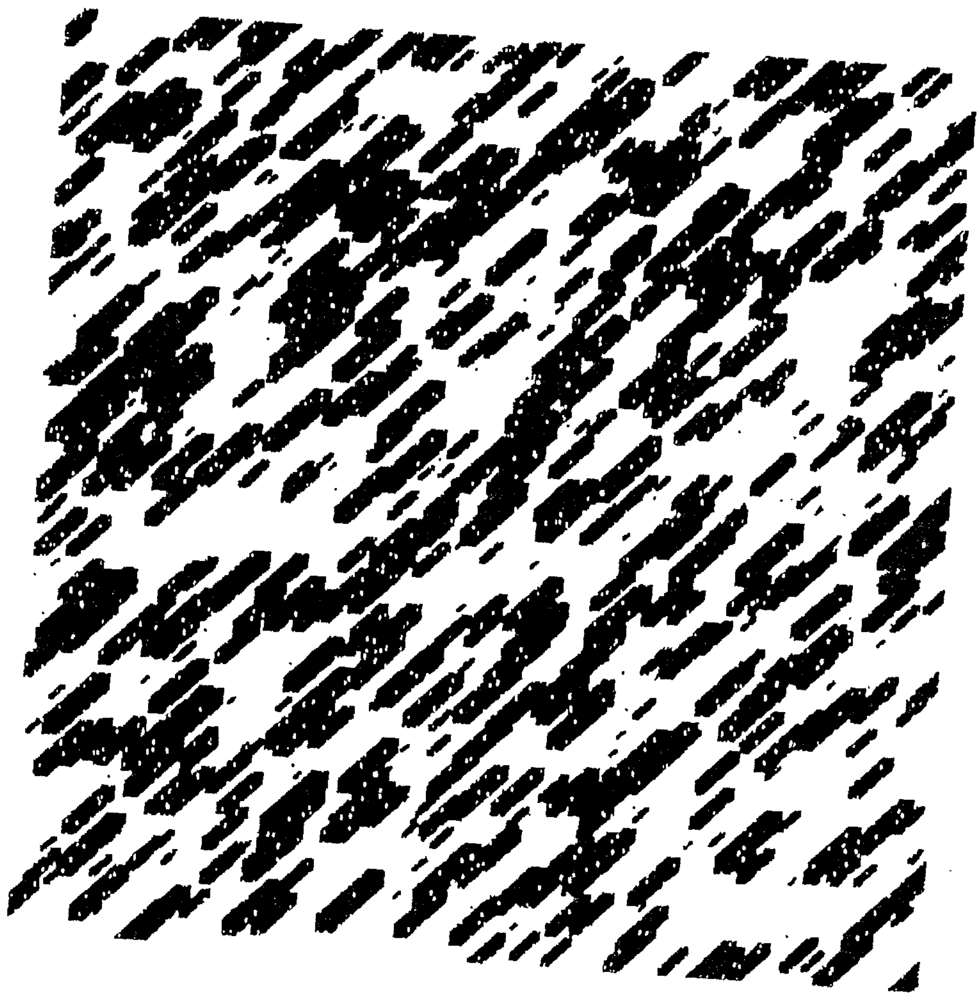

Figure 10: Restored image, obtained by filtering the DRS realization of figure (9) using structuring element W_2 (the best one).

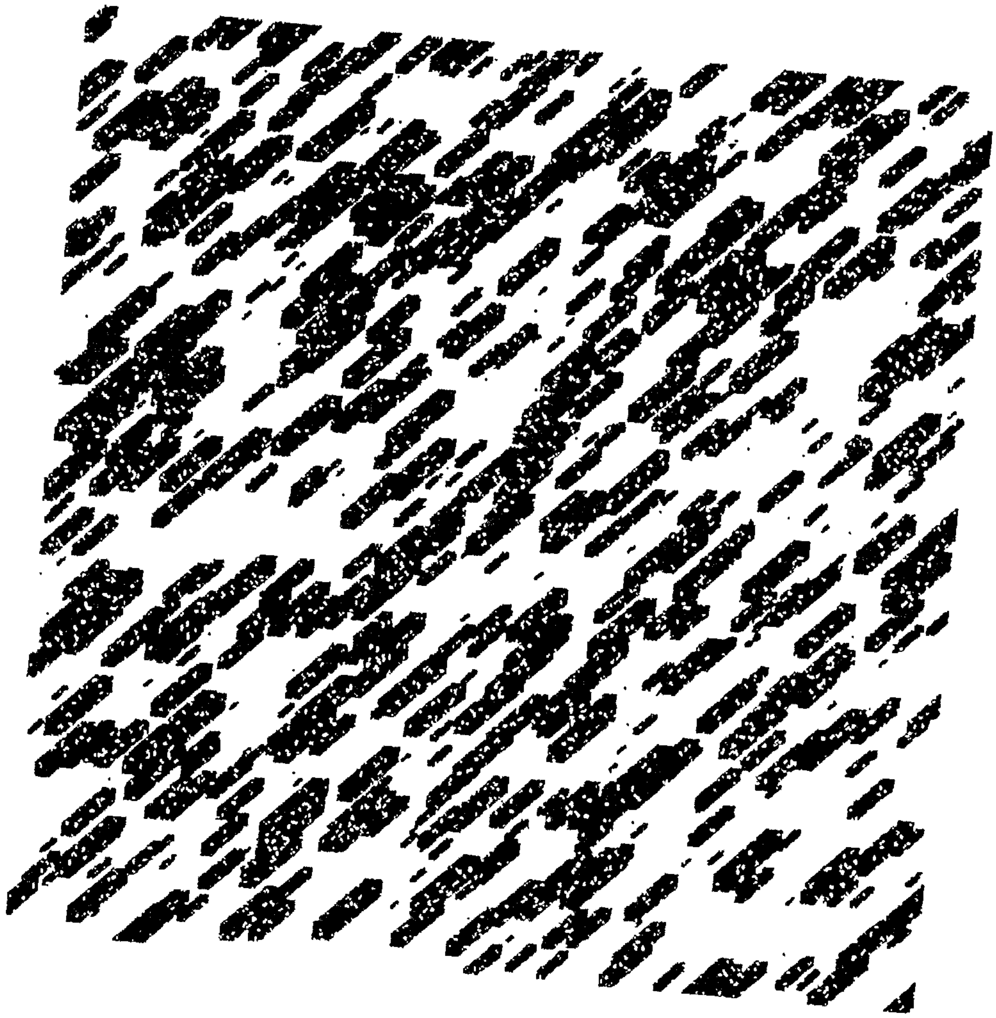

Figure 11: Restored image, obtained by filtering the DRS realization of figure (9) using structuring element W_4 (the worst one).