MAP Signal Estimation in Noisy Sequences of Morphologically Smooth Images

N. D. Sidiropoulos, D. Meleas, and T. Stragas

Abstract— **In a recent paper [1], it has been demonstrated that morphological openings and closings can be viewed as Maximum *a posteriori* (MAP) estimators of morphologically smooth signals in signal-independent i.i.d. noise. In this correspondence, we extend these results to the M-fold independent observation case, and show that the aforementioned estimators are strongly consistent. We also demonstrate the validity of a thresholding conjecture [2] by simulation, and use it to evaluate estimator performance. Taken together, these results can help determine the least upper bound, $\bar{M}$, on M, which guarantees virtually error-free reconstruction of morphologically smooth images.**

I. Introduction

In a recent paper [1], it has been demonstrated that morphological openings and closings can be viewed as Maximum *a posteriori* (MAP) estimators of morphologically smooth signals in signal-independent i.i.d. noise. These results were made possible by casting the filtering problem within a general framework of uniformly bounded discrete random set (or, discrete random set (DRS), for short) theory [4].

A DRS X is simply defined as a measurable mapping from some probability space to a measurable space $(\Sigma(B), \Sigma(\Sigma(B)))$, where $\Sigma(B)$ is a complete lattice with a finite least upper bound (usually, the power set, $\mathcal{P}(B)$, of some finite $B \subset \mathbf{Z}^2$), and $\Sigma(\Sigma(B))$ is a σ-field over $\Sigma(B)$ (usually, $\mathcal{P}(\mathcal{P}(B))$, the power set of the power set of B). A DRS X induces an associated probability structure $P_X(\cdot)$ on $\Sigma(\Sigma(B))$. DRS's can be viewed as finite-alphabet random variables, taking values in a finite partially ordered set (poset). Thus the basic difference with ordinary finite-alphabet random variables is that the DRS alphabet naturally possesses only a partial order relation, instead of a total order relation.

The foundations of mathematical morphology have been laid out by Matheron [5], [6], Serra [7], [8], and their collaborators. Morphological filtering [9] is one of the most popular and successful branches of this theory, due in part to the excellent shape-preservation (syntactic) properties of morphological filters [10]–[13]. Another aspect of filter behavior is revealed through statistical analysis. We are mostly interested in optimizing filter behavior with respect to some statistical measure of goodness [1], [2], [4]. Dougherty *et al.* [14]–[19], Schonfeld *et al.* [20]–[22], and Goutsias [23] have worked on several related problems, using different measures of optimality and/or families of filters. We concentrate on MAP optimality and strong consistency.

We will use $\oplus, \ominus, \circ, \bullet$ to denote Minkowski addition and subtraction, and morphological opening and closing, respectively; refer to [7] for definitions.

Manuscript received November 15, 1994; revised September 28, 1995. A portion of this work was presented at the IEEE 1995 Spring Workshop on Nonlinear Signal and Image Processing, Halkidiki, Greece.

N. D. Sidiropoulos was with the Hellenic Air Force Academy and G-Systems Ltd., Athens, Greece. He is now with the Institute for Systems Research, University of Maryland, College Park, MD 20742 USA (e-mail: nikos@Glue.umd.edu).

D. Meleas and T. Stragas were with the Hellenic Air Force Academy. They are now with Hellenic Air Force's Corps of Engineers, TK 11743 Athens, Greece.

Publisher Item Identifier S 1057-7149(96)04193-0.

A digital image $I \in \Sigma(B)$ is said to be W-*open iff* I is invariant under opening by W, i.e., $I \circ W = I$. It has been shown [6] that this latter condition is satisfied *iff* I is a union of replicas of the structural element W, i.e., iff I is spanned by translates of W. The interior of such an I can be perfectly traced by W, which can be thought of as a discrete counterpart to the notion of a ball (according to some distance metric) in $\mathbf{R}^2$; thus the term "smooth" in the title of this correspondence. We now state a backbone result of [1] which provides a starting point for our current discussion.

Theorem 1: Let $O_W(B)$ denote the collection of all W-open subsets of B. Assume that the signal DRS, X, on B, induces the following probability mass function on $\Sigma(B)$:

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

where $||$ stands for set cardinality. Furthermore, assume that the observable DRS is $Y = X \cup N$, where N is i.i.d. 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; N is, in effect, what is known as a *homogeneous Bernoulli lattice process* of intensity r), 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.

This latter result provided the first rigorous proof of a widespread "folk" theorem in the image processing community that opening is suitable, robust, and very effective in removing impulsive one-sided additive (union) noise. A union noise model is appropriate for modeling random clutter, partial occlusion, and obscuration. By duality and distributivity, similar results hold for the characterization of closings, unions of openings, and intersections of closings as MAP signal estimators under suitable—and plausible—statistical scenarios.

Even though the structure of the MAP estimator turns out to be independent of the noise level, r, the fidelity of the estimate, quite naturally, depends on r. In [2] it has been implicitly suggested that the fidelity of the MAP estimate (as measured by $E|\hat{X}_{MAP}(Y) \backslash X|$, where $\backslash$ stands for symmetric set difference) behaves as a steplike function of r, i.e., for $r < r^*$ the estimate is virtually error-free, while for $r \geq r^*$ the quality of the estimate deteriorates rapidly with r. However, an analytical proof of this "thresholding effect" seems to be very difficult.

In this correspondence, we extend the results of [1] to the M-fold case. In particular, we show that if X takes on values in the collection of all W-open subsets of some *finite* lattice, the observable record is $\mathbf{Y}^{(M)} = [Y_1, \cdots, Y_M]$, where $Y_i = X \cup N_i$, and $\{N_i\}_{i=1}^M$ is i.i.d., then $\cap_{i=1}^M Y_i$ is a sufficient statistic for the estimation of X on the basis of $\mathbf{Y}^{(M)}$, $(\cap_{i=1}^M Y_i) \circ W$, can be interpreted as a MAP signal estimator, the effective noise level is r^M, and this MAP estimator is strongly consistent. We also demonstrate the validity of the aforementioned thresholding conjecture by simulation, and conclude that, for all practical purposes, $r^* \geq 0.65$ [24].

Several related results follow by duality and distributivity of certain classes of morphological operators. Taken together, these results can help determine the least upper bound, $\bar{M}$, on M, which guarantees virtually error-free reconstruction of morphologically smooth images.

II. Multiframe Observations

The idea behind our present study of the case of multiple observations is that sometimes it pays off to consider combining a, usually small, number of degraded replicas of an original image, X, to produce a more accurate estimate $\widehat{X}$ of X. We start with the following setup. Let us assume for a moment that the signal (original) X is some unknown but deterministic "constant" in $O_W(B)$, and we observe a vector (sequence) of degraded images $\mathbf{Y}^{(M)} = [Y_1, \cdots, Y_M]$, where $Y_i = X \cup N_i$, $\{N_i\}_{i=1}^M$ is i.i.d. and independent of X, and each N_i is a Bernoulli lattice process of intensity $r \in [0, 1)$. We have the following result.

Lemma 1: $\cap_{i=1}^M Y_i$ is a sufficient statistic for the estimation of the unknown constant X on the basis of $\mathbf{Y}^{(M)}$.

Proof: Let $1(\cdot)$ denote the *indicator function* on the σ-field $\Sigma(\Sigma(B))$. Consider

$$\begin{aligned}\Pr(\mathbf{Y}^{(M)}|X) =& \begin{cases}\prod_{i=1}^M r^{|Y_i - X|}(1-r)^{|B-Y_i|}, & \text{if } X \subseteq \cap_{i=1}^M Y_i \\ 0, & \text{otherwise}\end{cases}\\ =& 1(X \subseteq \cap_{i=1}^M Y_i)\prod_{i=1}^M r^{|Y_i - X|}(1-r)^{|B-Y_i|}\\ =& 1(X \subseteq \cap_{i=1}^M Y_i) r^{\sum_{i=1}^M |Y_i - X|}(1-r)^{\sum_{i=1}^M |B-Y_i|}.\end{aligned}$$

Observe that $X \subseteq \cap_{i=1}^M Y_i$ implies $X \subseteq Y_i, \forall i = 1, \cdots, M$, therefore, $|Y_i - X| = |Y_i| - |X|$, and so,

$$\begin{aligned}&\Pr(\mathbf{Y}^{(M)}|X)\\ &= 1(X \subseteq \cap_{i=1}^M Y_i) r^{\left(\sum_{i=1}^M |Y_i|\right) - M|X|}(1-r)^{M|B| - \sum_{i=1}^M |Y_i|}\\ &= 1(X \subseteq \cap_{i=1}^M Y_i)\frac{(1-r)^{M|B|}}{r^{M|X|}}\left(\frac{r}{1-r}\right)^{\sum_{i=1}^M |Y_i|}.\end{aligned}$$

Let

$$g(X, \cap_{i=1}^M Y_i) \triangleq 1(X \subseteq \cap_{i=1}^M Y_i)\frac{(1-r)^{M|B|}}{r^{M|X|}}$$

and

$$h(\mathbf{Y}^{(M)}) \triangleq \left(\frac{r}{1-r}\right)^{\sum_{i=1}^M |Y_i|}$$

and the result follows from the Factorization Theorem [25]. ■

Next, let us consider MAP Signal Estimation. Toward this end, we need to introduce a prior distribution for X. This prior should reflect our knowledge about the statistical behavior of X. If all we know is that X takes on values in some collection of subsets of B (in our case $O_W(B)$), then it is reasonable to model this knowledge using a uniform distribution (in our case over $O_W(B)$). So let us additionally assume that X induces the following probability mass function on $\Sigma(B)$:

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

In addition, we assume that the noise sequence, $\{N_i\}_{i=1}^M$, is independent of X, for all $M \geq 1$. We have the following result.

Theorem 2: Under our foregoing assumptions

$$\widehat{X}_{MAP}(\mathbf{Y}^{(M)}) = \left(\cap_{i=1}^M Y_i\right) \circ W$$

for all $r \in [0, 1)$.

Proof:

$$\widehat{X}_{MAP}(\mathbf{Y}^{(M)}) = \text{argmax}_{S \in O_W(B)} \Pr(X = S|\mathbf{Y}^{(M)}).$$

Observe that any $S \in O_W(B)$ which does not satisfy $S \subseteq \cap_{i=1}^M Y_i$ has zero-conditional probability. Let $T = T(W, \cap_{i=1}^M Y_i)$ denote the restriction of $O_W(B)$ to $\cap_{i=1}^M Y_i$ (i.e., $S \in T$ iff $S \in O_W(B)$ and $S \subseteq (\cap_{i=1}^M Y_i)$). This is the signal subspace which now carries all the conditional measure. Then

$$\begin{aligned}&\widehat{X}_{MAP}(\mathbf{Y}^{(M)})\\ &= \text{argmax}_{S \in T}\frac{\Pr(\mathbf{Y}^{(M)}|X = S)\Pr(X = S)}{\Pr(\mathbf{Y}^{(M)})}\\ &= \text{argmax}_{S \in T}\Pr(\mathbf{Y}^{(M)}|X = S)\Pr(X = S)\\ &= \text{argmax}_{S \in T}\Pr(\mathbf{Y}^{(M)}|X = S)\frac{1}{|O_W(B)|}\\ &= \text{argmax}_{S \in T}\Pr(\mathbf{Y}^{(M)}|X = S)\\ &= \text{argmax}_{S \in T}\prod_{i=1}^M r^{|Y_i \setminus S|}(1-r)^{|B \setminus Y_i|}\\ &= \text{argmax}_{S \in T}\prod_{i=1}^M r^{|Y_i \setminus S|}\\ &= \text{argmax}_{S \in T} r^{\sum_{i=1}^M |Y_i \setminus S|}\\ &= \text{argmin}_{S \in T}\sum_{i=1}^M |Y_i \setminus S|\\ &= \text{argmax}_{S \in O_W(B), S \subseteq (\cap_{i=1}^M Y_i)}|S|\end{aligned}$$

So, $\widehat{X}_{MAP}(\mathbf{Y}^{(M)})$ is the largest $W-open$ subset of $\cap_{i=1}^M Y_i$, which is exactly the opening of $\cap_{i=1}^M Y_i$ by W [6], [7]. ■

Theorem 3—Strong Consistency: Under the foregoing assumptions

$$\widehat{X}_{MAP}(\mathbf{Y}^{(M)}) \longrightarrow X, a.s. as M \to \infty$$

where a.s. means *almost surely*, i.e., convergence almost everywhere, except for a set of measure zero.

Proof: We start by showing that, in the pathwise sense, and for all $M \geq 1$

$$X \subseteq \widehat{X}_{MAP}(\mathbf{Y}^{(M)}) \subseteq \cap_{i=1}^M Y_i$$

and complete the proof by showing that

$$\cap_{i=1}^M Y_i \longrightarrow X, a.s. as M \to \infty.$$

For the first step, observe that

$$\widehat{X}_{MAP}(\mathbf{Y}^{(M)}) = \left(\cap_{i=1}^M Y_i\right) \circ W \subseteq \cap_{i=1}^M Y_i$$

since $\circ$ is an anti-extensive operator [6], [7]. Also,

$$\begin{aligned}\widehat{X}_{MAP}(\mathbf{Y}^{(M)}) &= \left(\cap_{i=1}^M (X \cup N_i)\right) \circ W\\ &= \left(X \cup \cap_{i=1}^M N_i\right) \circ W\\ &\supseteq (X \circ W) \cup \left(\left(\cap_{i=1}^M N_i\right) \circ W\right)\\ &= X \cup \left(\left(\cap_{i=1}^M N_i\right) \circ W\right) \supseteq X\end{aligned}$$

since X is W-open. For the second part of the proof, look at

$$\Pr\left(lim\left\{\cap_{i=1}^M Y_i \setminus X = \emptyset\right\}\right).$$

Here, lim stands for limit as $M \to \infty$. The events in curly brackets form an increasing sequence, therefore [26],

$$\begin{aligned}&\Pr\left(\lim\left\{\cap_{i=1}^M Y_i \setminus X = \emptyset\right\}\right)\\ &= \lim \Pr\left(\cap_{i=1}^M Y_i \setminus X = \emptyset\right)\\ &= \lim \Pr\left(\cap_{i=1}^M N_i \setminus X = \emptyset\right)\\ &\geq \lim \Pr\left(\cap_{i=1}^M N_i = \emptyset\right) = \lim\left(1 - r^M\right)^{|B|} = 1.\end{aligned}$$

So, since $\Pr\left(\lim\left\{\cap_{i=1}^{M} Y_i \backslash X = \emptyset\right\}\right)$ is a valid probability, we conclude that $\Pr\left(\lim\left\{\cap_{i=1}^{M} Y_i \backslash X = \emptyset\right\}\right) = 1$ which completes the proof. ■

Since the sufficient statistic itself a.s. converges to the true underlying signal DRS X, why need we study the performance of the MAP estimator? The answer, of course, is that (as we will soon see) the MAP estimator *converges much faster* than the sufficient statistic does.

Corollary 1: Given our on-going setup, we can conclude that, for the purposes of MAP signal estimation, the effective noise level is r^M.

Proof: Note that

$$\cap_{i=1}^{M} Y_i = \cap_{i=1}^{M} (X \cup N_i) =$$

$$X \cup \left(\cap_{i=1}^{M} N_i\right) \sim X \cup N$$

where N is i.i.d. of intensity r^M, and $\sim$ means equivalent in the sense of distributions. ■

Thus, in so far as estimator performance is concerned, we can restrict our study to the $M = 1$ case.

III. Sampling and Simulation

In order to study estimator performance, and subsequently validate the thresholding conjecture suggested in [2], we would like to devise a way to sample from our target distribution i.e., the uniform distribution over $O_W(B)$. However, this turns out to be a formidable task. Observe, in particular, that a uniform distribution over $O_W(B)$ is not equivalent to, and, in fact, cannot be approximated by, the distribution of a Boolean model (cf. [27] and references therein for background on Boolean models; there exists a substantial amount of literature on this subject) having W as its (deterministic) primary grain, for even though individual samples of the two distributions may look very much alike, the ensemble properties of the two distributions differ (e.g., by virtue of the law of large numbers, and for sufficiently wide B, a Boolean model of constant intensity will tend to produce realizations with an approximately constant number of grains). In what follows, we focus on *worst case behavior* and provide a reasonable estimate of r^*, which is then validated by further simulation.

The fidelity of the MAP estimate (as measured via $E|\widehat{X}_{MAP}(Y)\backslash X| = E|Y \circ W \backslash X| = E|(X \cup N) \circ W \backslash X|$) depends on the noise intensity, r, and the size and shape of the signal primitive, W, relative to B. We have found that, for convex W, and for all practical purposes, the shape factor can be safely ignored. Observe that

$$E|(X \cup N) \circ W \backslash X| = \phi(r, W)$$

$$= E_X\{E_{N|X}\{|(X \cup N) \circ W \backslash X|\}\}$$

$$\leq \max_{X \in O_W(B)} \{E_{N|X}\{|(X \cup N) \circ W \backslash X|\}\}$$

$$= \bar{\phi}(r, W).$$

Let us fix W. For any *given* realization of the signal DRS X, and noise intensity r, the conditional expectation $E_{N|X}\{|(X \cup N) \circ W \backslash X|\}$ can be estimated by an appropriate sample average; that is, by producing sufficiently many independent realizations of N, each time merely generating an i.i.d. sequence of binary variables to simulate N, and computing the resulting error $|(X \cup N) \circ W \backslash X|$, then averaging the respective errors. Thus, by

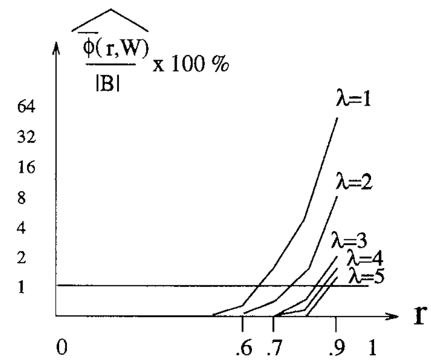

Fig. 1. Estimator performance plots.

repeating this process for a range of values of r, we can plot the value of $E_{N|X}\{|(X \cup N) \circ W \backslash X|\}$ as a function of r, for a given X.

The next step is to compute $\bar{\phi}(r, W)$, for which we need to generate a total of $|O_W(B)|$ plots like the one above, then take the pointwise (in r) maximum of these plots. This is not feasible in practice, for in general we cannot even *enumerate* the elements of $O_W(B)$, let alone generate them. Nevertheless, the goal is a worthy one, and we therefore take on a reasonable, yet not entirely satisfactory approach, that is *we only generate a subcollection* of these plots corresponding to a representative array of realizations of X, and take the pointwise maximum (in r) *over this subcollection only* as our estimate of $\bar{\phi}(r, W)$ [24]. This latter estimate clearly underestimates $\bar{\phi}(r, W)$, yet if the utilized subset of realizations of X is indeed representative of the entire signal space, $O_W(B)$, then the effect of this bias in our final estimate of r^* will be more than accounted for by the fact that we estimate an *upper bound* on expected error, instead of expected error *per se*. This is a key point, and it is validated by further simulation in Section V.

The experimental setup is as follows (further details can be found in [24]). B is taken to be a 146×146 square lattice, whereas $W = \lambda\Theta = \Theta \oplus \Theta \oplus \cdots \oplus \Theta(\lambda - 1 times)$, where $\lambda \in \{1, \cdots, 5\}$, and Θ is a 21-point discrete octagon. For each λ we run a separate experiment.

The most important consideration in generating a *representative* subcollection of signal realizations is to account for signal variability, i.e., *number* and *relative position* of W primitives in X. To do this, we take the following approach. We fix the number of primitives, ν_p, in X, and generate 50 realizations of X with the given number of primitives, each time placing ν_p random translates of W within the interior of B, according to a uniform probability distribution over $B \ominus W^s$ for the translation vectors (here, ^{s} stands for reflection about the origin). The process is repeated for 10 equispaced values of ν_p, the maximum of which is chosen to ensure almost complete coverage of B (this strongly depends on λ). Thus for each λ, we end up with 500 realizations of X, for each one of which a plot of $E_{N|X}\{|(X \cup N) \circ W \backslash X|\}$ as a function of r is generated, as explained earlier on. Finally, we take the pointwise (in r) maximum of these 500 curves. The overall process results in one performance curve for each λ. These curves are plotted in Fig. 1.

The results are consistent with intuition, for the bigger λ is, the smoother the signal DRS X relative to the noise DRS N is, and therefore, the more noise is removed by projecting the statistic onto the signal subspace, which is essentially what the MAP estimator is doing. As it can be clearly seen from this figure, if we define "virtually error-free" reconstruction to mean that the residue noise rate *after MAP filtering* is below 1%, then 0.65 appears to be a tight lower bound on r^* for all $\lambda \geq 1$. What this means is that, in the context of multiframe observations of noise intensity r each, by virtue of Corollary 1, and assuming that r is known *a priori*, one can pick the length, M, of the observation sequence such that the effective noise level *prior to filtering*, r^M, drops below the 0.65 threshold, and therefore guarantee that the residue noise rate *after filtering* is below 0.01, i.e., virtually error-free reconstruction. This will also be the case if r can be adaptively estimated by occasionally transmitting a test pattern.

IV. Extensions

We now present two more theorems. The first can be easily established by following the same lines of proof as in Theorems 2 and 3. The second can be established from the first by appealing to duality (note that closing is the dual of opening with respect to lattice complementation).

Theorem 4: Assume we observe $\mathbf{Y}^{(M)} = [Y_1, \cdots, Y_M]$, where $Y_i = X \cup N_i$, $\{N_i\}_{i=1}^M$ is i.i.d. and independent of X for all $M \geq 1$, and each N_i is i.i.d. of intensity $r \in [0, 1)$. Let us further assume that X is uniformly distributed over the collection of all subsets K of B which are spanned by unions of translates of W_l, i.e., those $K \subseteq B$ which can be written as

$$K = \cup_{l=1}^L K_l, K_l \in O_{W_l}(B), l = 1, \cdots, L$$

Then

$$\widehat{X}_{MAP}(\mathbf{Y}^{(M)}) = \bigcup_{l=1}^{L} \left(\left(\cap_{i=1}^M Y_i \right) \circ W_l \right)$$

and

$$\widehat{X}_{MAP}(\mathbf{Y}^{(M)}) \longrightarrow X, a.s. as M \rightarrow \infty$$

We now state the dual theorem. Observe that here we deal with intersection noise, which can be interpreted as a formal mechanism to consider random sampling of DRS's. The reader is referred to [28] for an account of an interesting approach when N is assumed to be a deterministic regularly spaced grid which undersamples the observation.

Theorem 5: Assume we observe $\mathbf{Y}^{(M)} = [Y_1, \cdots, Y_M]$, where $Y_i = X \cap N_i$, $\{N_i\}_{i=1}^M$ is i.i.d. and independent of X for all $M \geq 1$, and each N_i is i.i.d. of intensity $r \in [0, 1)$. Let us further assume that X is uniformly distributed over the collection of all subsets K of B which can be written as

$$K = \cap_{l=1}^L K_l, K_l \in C_{W_l}(B), l = 1, \cdots, L$$

where $C_{W_l}(B)$ denotes the set of all W_l-closed subsets of B. Then

$$\widehat{X}_{MAP}(\mathbf{Y}^{(M)}) = \bigcap_{l=1}^{L} \left(\left(\cup_{i=1}^M Y_i \right) \bullet W_l \right)$$

and

$$\widehat{X}_{MAP}(\mathbf{Y}^{(M)}) \longrightarrow X, a.s. as M \rightarrow \infty.$$

V. An Example

Let us now present a simple complete simulation example to enable us to appreciate the kind of gain one might achieve by using more than one observation. Fig. 2 depicts a 512×512 raster-scanned portion of a typeset document, which has been preconditioned to be open with respect to a discrete octagon of size $\lambda = 2$; this is assumed to be a realization of the signal DRS X. In practice, given a characterization of the class of noise-free images under consideration (e.g., font type(s) and size(s) for typeset documents), and for a union noise model, the structural element W should be chosen to maximize $|W|$ (i.e., maximize noise elimination capability), subject to the constraint that

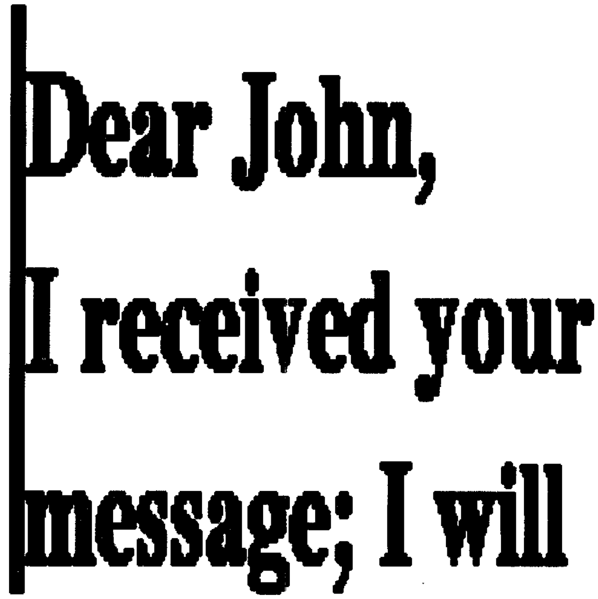

Fig. 2. X is a raster-scanned portion of a typeset document.

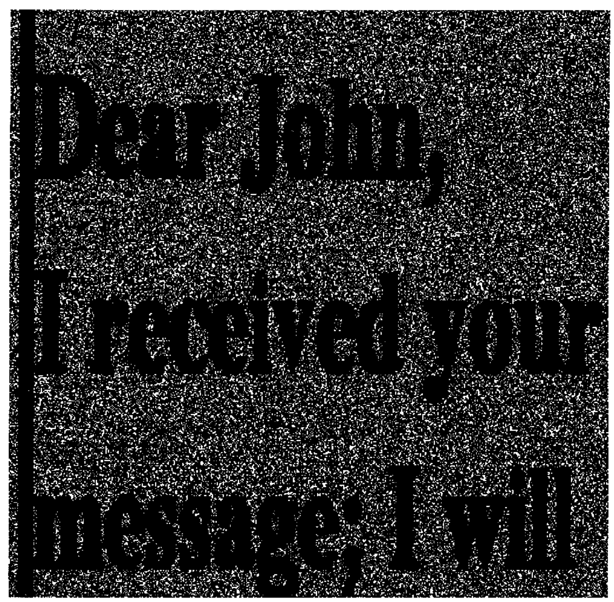

Fig. 3. Y_1 is X corrupted by i.i.d. union noise of intensity $r = 0.75$ (slightly above threshold).

the given class of noise-free images is a subset of $O_W(B)$ (i.e., subject to the constraint that the signal is not affected by the filtering operation). If the latter constraint is not satisfied, then the resulting filter may distort the noise-free signal itself, and it may converge to a subset of the true signal. In general, it pays to consider multiple structuring elements, as in Theorem 4, to achieve the best results, in which case one follows the same basic principle for choosing an appropriate collection of structural elements. To avoid unnecessary complication, we only use one structural element for the purposes of illustration.

Figs. 3 and 4 depict two independently degraded versions of Fig. 2, each heavily corrupted by i.i.d. union noise of intensity $r = 0.75$,

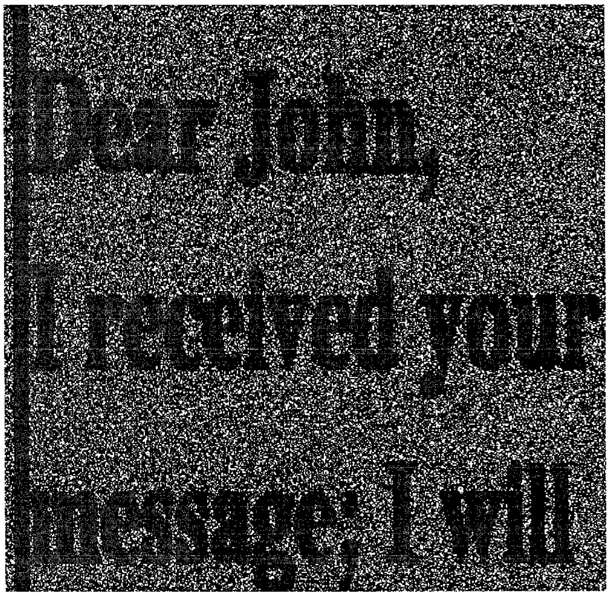

Fig. 4. Y_2 is X corrupted by i.i.d. union noise of intensity $r = 0.75$ (slightly above threshold).

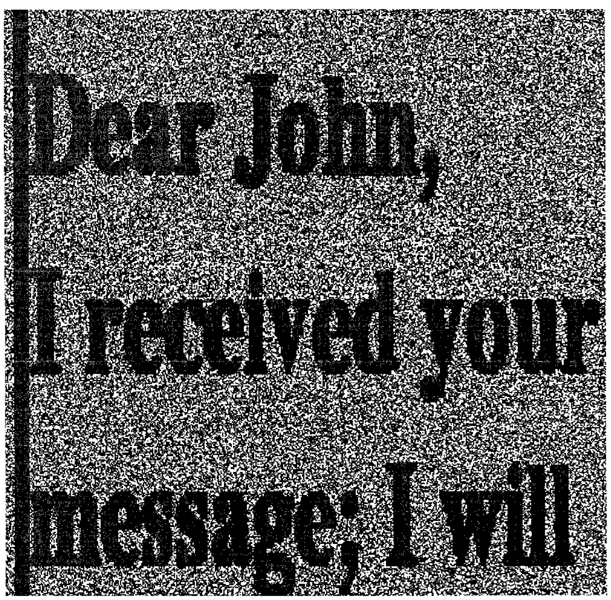

Fig. 5. $Y_1 \cap Y_2$.

Fig. 6. $\widehat{X}_{MAP}(Y_1)$ is MAP estimate based on Y_1.

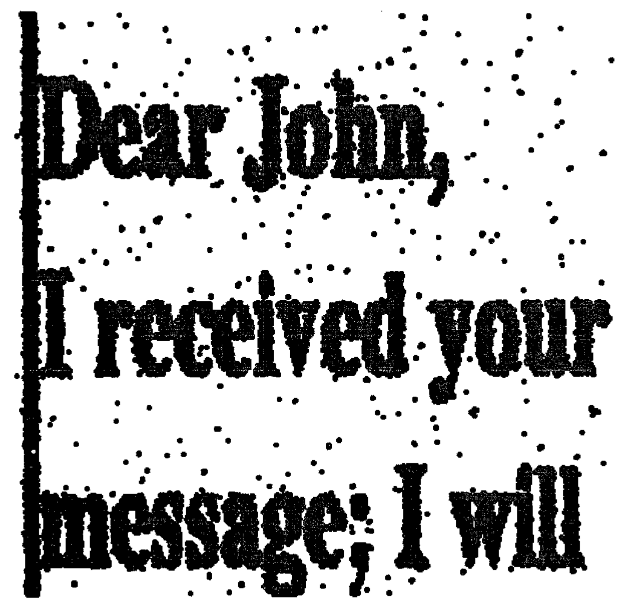

Fig. 7. $\widehat{X}_{MAP}(Y_2)$ is MAP estimate based on Y_2.

which is slightly above threshold. Call these Y_1, and Y_2, respectively. Fig. 5 depicts the resulting realization of the sufficient statistic, $Y_1 \cap Y_2$. Figs. 6 and 7 depict the optimal MAP estimate of X based on Y_1, Y_2, respectively. Both estimates suffer from noise artifacts due to excessive clustering. Fig. 8 depicts the optimal MAP estimate of X on the basis of the joint observation vector (Y_1, Y_2). This estimate does not suffer from noise clustering artifacts; this could have been predicted by our earlier discussion of the thresholding effect. Observe that the effective noise power in the observation statistic is $0.75^2 = 0.562$, i.e., below threshold.

Given that we operate in a very noisy environment, the estimate in Fig. 8 seems very good. Yet, how does it compare to standard median filtering? Observe that, unlike opening, the median is not guaranteed to converge when applied to the sufficient statistic, the reason being that the "dominated convergence" argument in the proof of Theorem 3 fails in the case of the median. Still, one might be able to get good results by choosing the median window to make the class of input signals a subset of the *root set* of the median, i.e., the class of all signals that are invariant under median filtering with respect to the given median window. However, median roots are much harder to characterize in two dimensions [13], and may include oscillatory patterns, which are probably not suitable for modeling typeset documents. In contrast, as mentioned before, the class of images invariant under opening by W is precisely the class of images spanned by translates of W; this class is easy to visualize and work with, and it does not include any oscillatory patterns. The net result is

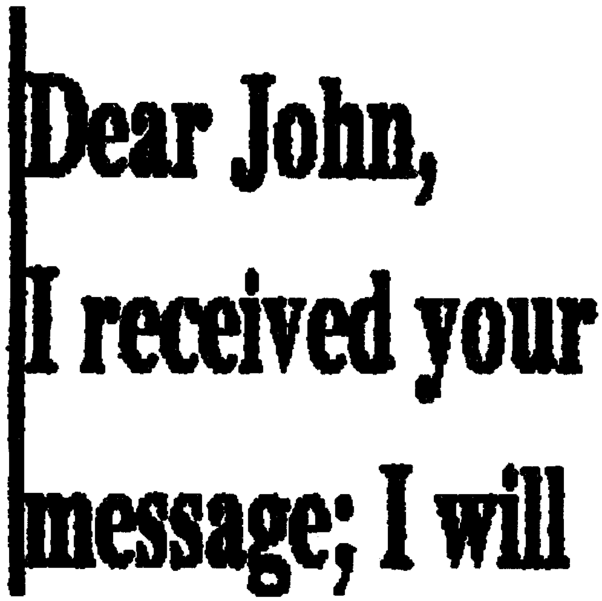

Fig. 8. $\widehat{X}_{MAP}(Y_1, Y_2)$ is MAP estimate based on the vector (Y_1, Y_2).

that while a properly chosen median may also give reasonable results when M is small, its asymptotic behavior will be worse than that of the opening, which is MAP-optimal and guaranteed to converge in this case, and for a good reason: the median is an *unbiased* estimator, and unbiased estimators are not suitable for union noise, for they treat an object and its background in a balanced fashion, whereas there exists a very clear imbalance between the two in *one-sided* noise; good estimators should capitalize on this fact.

VI. Conclusions

In this correspondence we presented proofs of MAP optimality and strong consistency of certain classes of morphological filters acting on noisy *sequences* of morphologically smooth images. We also demonstrated the validity of a thresholding conjecture suggested in [2] by simulation, and used it to evaluate estimator performance. Taken together, these results can help determine the least upper bound, $\bar{M}$, on M, which guarantees virtually error-free reconstruction of morphologically smooth images.

Acknowledgment

The authors would like to thank the anonymous reviewers, whose comments have helped them improve this manuscript.

References

[1] N. D. Sidiropoulos, J. S. Baras, and C. A. Berenstein, "Optimal filtering of digital binary images corrupted by union/intersection noise," *IEEE Trans. Image Processing*, vol. 3, pp. 382–403, 1994.
[2] ——, "An algebraic analysis of the generating functional for discrete random sets, and statistical inference for intensity in the discrete Boolean random set model," *J. Math. Imaging Vision*, vol. 4, pp. 273–290, 1994.
[3] N. D. Sidiropoulos, D. Meleas, and T. Stragas, "Multiframe MAP signal estimation in morphologically smooth images," in *Proc. IEEE Spring Workshop in Nonlinear Signal and Image Processing*, Halkidiki, Greece, June 1995.
[4] N. D. Sidiropoulos, "Statistical inference, filtering, and modeling of discrete random sets," Ph.D. dissertation, Univ. of Maryland, College Park, June 1992.
[5] G. Matheron, *Elements pour une theorie des Milieux Poreux*. France: Masson, 1967.
[6] ——, *Random Sets and Integral Geometry*. New York: Wiley, 1975.
[7] J. Serra, *Image Analysis and Mathematical Morphology*. New York: Academic, 1982.
[8] J. Serra, Ed., *Image Analysis and Mathematical Morphology, vol. 2, Theoretical Advances*. San Diego, CA: Academic, 1988.
[9] J. Serra and L. Vincent, "An overview of morphological filtering," *Circuits, Syst. Signal Processing*, vol. 11, no. 1, pp. 47–108, Jan. 1992.
[10] J. Song and E. J. Delp, "A study of the generalized morphological filter," *Circuits, Syst. Signal Processing*, vol. 11, no. 1, pp. 229–252, Jan. 1992.
[11] P. Maragos and R. W. Schafer, "Morphological filters—Part I: Their set-theoretic analysis and relations to linear shift-invariant filters," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, pp. 1153–1169, Aug. 1987.
[12] P. Maragos and R. W. Schafer, "Morphological filters—Part II: Their relations to median, order-statistic, and stack filters," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, pp. 1170–1184, Aug. 1987.
[13] I. Pitas and A. N. Venetsanopoulos, *Nonlinear Digital Filters: Principles and Applications*. Boston, MA: Kluwer, 1990.
[14] E. Dougherty, "Optimal mean square N-Observation Digital morphological filters—I. Optimal binary filters," *Computer Vision, Graphics, and Image Processing: Image Understanding*, vol. 55, no. 1, pp. 36–54, Jan. 1992.
[15] R. P. Loce and E. R. Dougherty, "Using structuring element libraries to design suboptimal morphological filters," in *Proc. SPIE Conf. on Image Algebra and Morphological Image Processing II*, vol. 1568, July 1991, pp. 233–246.
[16] R. M. Haralick, E. R. Dougherty, and P.L. Katz, "Model-based morphology," in *Proc. SPIE*, vol. 1472, Orlando, FL, Apr. 1991.
[17] 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, CA, Feb. 1991.
[18] E. R. Dougherty and R. P. Loce, "Constrained optimal digital morphological filters," in *Proc. 25th Ann. Conf. on Information Sciences and Systems*, Baltimore, MD, Mar. 1991.
[19] 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, MA, Nov. 1991.
[20] D. Schonfeld and J. Goutsias, "Optimal morphological pattern restoration from noisy binary images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 13, pp. 14–29, Jan. 1991.
[21] D. Schonfeld, "Optimal structuring elements for the morphological pattern restoration of binary images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 16, pp. 589–601, June 1994.
[22] D. Schonfeld and J. Goutsias, "On the morphological representation of binary images in a noisy environment," *J. Visual Communi. Image Rep.*, vol. 2, no. 1, pp. 17–30, Mar. 1991.
[23] J. Goutsias, "Morphological analysis of discrete random shapes," *J. Math. Imaging Vision*, vol. 2, pp. 193–215, 1992.
[24] D. Meleas and T. Stragas, "A simulation study of the noise-removal capability of morphological openings as a function of noise intensity in the context of MAP image estimation," Diploma thesis, Hellenic Air Force Academy, May 1994.
[25] H. V. Poor, *An Introduction to Signal Detection and Estimation*. New York: Springer-Verlag, 1988.
[26] R. M. Gray and L. D. Davisson, *Random Processes*. Englewood Cliffs, NJ: Prentice-Hall, 1986.
[27] D. Stoyan, W. S. Kendall, and J. Mecke, *Stochastic Geometry and its Applications*. Berlin, Germany: Wiley, 1987.
[28] R. M. Haralick, X. Zhuang, C. Lin, and J. S. J. Lee, "The digital morphological sampling theorem," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 2067–2089, 1989.