JB 94-27

Entitled:

"Further Results on MAP Optimality and Strong Consistency of Certain Classes of Morphological Filters"

Authors:

(with N.D. Sidiropoulos and C.A. Berenstein)

IEEE Transactions on Image Processing,

available as a technical report
from the Institute of Systems Research
University of Maryland

IEEE TRANSACTIONS ON IMAGE PROCESSING

A PUBLICATION OF THE IEEE SIGNAL PROCESSING SOCIETY

MAY 1996 VOLUME 5 NUMBER 5 IIPRE4 (ISSN 1057-7149)

PAPERS

CORRESPONDENCE

Correspondence

Further Results on MAP Optimality and Strong Consistency of Certain Classes of Morphological Filters

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

Abstract— **Morphological openings and closings can be viewed as consistent MAP estimators of smooth random binary image signals immersed in i.i.d. clutter, or suffering from i.i.d. random dropouts. We revisit this viewpoint under much more general assumptions and show that, quite surprisingly, the above interpretation is still valid.**[1]

I. Introduction

In recent work [3], [4], Sidiropoulos *et al.* have obtained statistical proofs of MAP optimality and strong consistency of certain classes of morphological filters, namely, morphological openings, closings, unions of openings, and intersections of closings. These results were made possible by casting the filtering problem within a general framework of uniformly bounded discrete random set (DRS) theory [5], [6].

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 optimality results of [3], [4] critically depend on the assumption that B is *finite*; they further assume that the noise process is i.i.d., both within a given observation (pixel-wise), and across a sequence of observations (sequence-wide). As it turns out, the pixel-wise i.i.d. assumption as well as the sequence-wide assumption of identical distribution can both be removed, as long as the sequence-wide independence assumption is maintained. The net result is that we end up with a rather general set of optimality conditions, which includes the previous set as a special case. The most interesting feature of this new set of conditions is that it allows the explicit incorporation of geometric and probabilistic constraints into the noise model, thereby establishing optimality in a significantly more flexible environment.

II. Background

The fundamental elements of mathematical morphology have been developed by Matheron [7], [8], Serra [9], [10], and their collaborators. Morphological filtering is one of the most popular and successful branches of this theory.[2] One good reason for the widespread use of morphological filters is their excellent shape-preservation (syntactic) properties. Important characterizations (e.g., root signal structure, relations to other filter classes, basis representation theory) are well developed and understood [11], [13]–[16]. 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 [3]–[6]. Dougherty *et al.* [17]–[22] Schonfeld *et al.* [23]–[25] and Goutsias [26] have worked on several related problems using different measures of optimality and/or families of filters. We concentrate on MAP optimality and strong consistency.

In morphological image analysis, structural and geometric image constraints are often expressed in terms of domains of invariance under certain morphological lattice operators. A digital image $I \in \Sigma(B)$ is said to be *smooth* with respect to a given operator f *iff* it is invariant under that operator, i.e., $f(I) = I$, in which case I is also called a *root* of f. The collection of all roots of f is the *domain of invariance* (or *root set*) of f. In what follows $\circ, \bullet$ denote morphological opening, and closing, respectively, whereas $O_W(B), C_W(B)$ denote the root sets of opening by W, and closing by W, respectively. It is well known [8] that $O_W(B)$ is the collection of all images (subsets of B), which are spanned by unions of translates of W, and, in view of duality, a dual interpretation is available for $C_W(B)$. We can also fit more complicated image structure by allowing composite constraints, e.g., consider the class of all images that are invariant under a union of openings with respect to a family of structural elements.

A drawback of the optimality results of [3] and [4] was that the noise process could not be "smooth:" e.g., one could not accommodate a composite noise process resulting by taking the union of translated replicas of some noise "primitives." In effect, one could not accommodate colored and/or geometrically structured noise. In what follows, this restriction is completely eliminated. Furthermore, the sequence-wide assumption of identical noise distribution is removed. In fact, we show that, modulo a relatively mild condition on marginal noise statistics that is needed for consistency, as long as the noise process consists of a sequence of independent DRS's, which is independent of the signal DRS, it can otherwise be arbitrary, and the results still hold.

III. Results

Theorem 1—MAP Optimality: Assume we observe $\mathbf{Y}^{(M)} = [Y_1, \cdots, Y_M]$, where $Y_i = X \cup N_i$, $\{N_i\}_{i=1}^M$ is an independent *but not necessarily identically distributed* sequence of noise DRS's, which is independent of X, and each N_i is an *otherwise arbitrary DRS* taking values in some arbitrary collection $\Psi_i(B) \subseteq \Sigma(B)$ of subsets of the observation lattice B. Let us further assume that X is uniformly distributed over a collection $\Phi(B) \subseteq \Sigma(B)$, of all subsets K of B, which are spanned by unions of translates of a family of structural elements W_l, $l = 1, \cdots, L$, i.e., those $K \subseteq B$, which can be written as[3] $K = \cup_{l=1}^L K_l$, $K_l \in O_{W_l}(B)$, $l = 1, \cdots, L$. Then

Manuscript received January 31, 1995; revised September 18, 1995. The associate editor coordinating the review of this work and approving it for publication was Prof. John Goutsias.

N. D. Sidiropoulos and J. S. Baras are with the Institute for Systems Research and the Department of Electrical Engineering, University of Maryland, College Park, MD 20742 USA (e-mail: nikos@Glue.umd.edu).

C. A. Berenstein is with the Department of Mathematics and the Institute for Systems Research of the University of Maryland, College Park, MD 20742 USA.

Publisher Item Identifier S 1057-7149(96)03171-5.

[1] Restricted earlier versions of these results were presented in [1] and [2].

[2] See [11] and [12] for recent surveys of the status of morphological filtering.

[3] Note that one or more of the K_l's can be empty, since $\emptyset \in O_W(B), \forall W$.

$\hat{X}_{\mathrm{MAP}}(\mathbf{Y}^{(M)}) = \bigcup_{l=1}^{L}((\cap_{i=1}^{M} Y_i) \circ W_l)$ is a MAP estimator of X on the basis of $\mathbf{Y}^{(M)}$.

Proof: Following some manipulations, the MAP principle reduces to $\hat{X}_{\mathrm{MAP}}(\mathbf{Y}^{(M)}) = \mathrm{argmax}_{S \in \Phi(B) \cap \mathcal{P}(\cap_{i=1}^{M} Y_i)} \Pr(\mathbf{Y}^{(M)} \mid X = S)$, where $\mathcal{P}(\cap_{i=1}^{M} Y_i)$ (the power set of $\cap_{i=1}^{M} Y_i$) is the sub-σ-field restriction imposed by the observations. By independence

$$\hat{X}_{\mathrm{MAP}}(\mathbf{Y}^{(M)}) = \underset{S \in \Phi(B) \cap \mathcal{P}(\cap_{i=1}^{M} Y_i)}{\mathrm{argmax}} \prod_{j=1}^{M} \Pr(S \cup N_j = Y_j).$$

Now, $S \in \mathcal{P}(\cap_{i=1}^{M} Y_i)$ implies that $S \subseteq Y_j$ for every $j = 1, 2, \cdots, M$, and $S \cup N_j = Y_j \Leftrightarrow Y_j \backslash S \subseteq N_j \subseteq Y_j$. For $j = 1, 2, \cdots, M$, define $G(\cdot, j)$: $\Phi(B) \cap \mathcal{P}(\cap_{i=1}^{M} Y_i) \mapsto \mathcal{P}(\Psi_j(B))$ by

$$G(S, j) \triangleq \{V \in \Psi_j(B) \mid Y_j \backslash S \subseteq V \subseteq Y_j\}.$$

This is the set of all N_j-realizations, which are consistent with the jth observation (i.e., Y_j) under the hypothesis that the true signal is S, which is necessarily in $\Phi(B) \cap \mathcal{P}(\cap_{i=1}^{M} Y_i)$. Clearly, as a function of $S \in \Phi(B) \cap \mathcal{P}(\cap_{i=1}^{M} Y_i)$, and for every $j = 1, 2, \cdots, M$, the set $G(S, j)$ is monotone nondecreasing. Therefore, the functional

$$\prod_{j=1}^{M} \Pr(S \cup N_j = Y_j) = \prod_{j=1}^{M} \Pr(N_j \in G(S, j))$$

is monotone nondecreasing with respect to $S \in \Phi(B) \cap \mathcal{P}(\cap_{i=1}^{M} Y_i)$.

The MAP optimality result then follows trivially from the fact that $\bigcup_{l=1}^{L}((\cap_{i=1}^{M} Y_i) \circ W_l)$ is the maximal element of $\Phi(B) \cap \mathcal{P}(\cap_{i=1}^{M} Y_i)$. Nonuniqueness of the functional form of the MAP estimator is a direct consequence of the fact that the functionals $G(S, j)$ are generally not *strictly* increasing. □

Theorem 2—Strong Consistency: In addition, assume the following Condition 1 holds.

Condition 1: $\forall z \in B$ there exists $0 \leq r < 1$, such that $\Pr(z \in N_j) \leq r$, for infinitely many indices j.[4] In other words, for every $z \in B$, there exists $0 \leq r < 1$, and an infinitely long subsequence $\mathcal{I}$ of observation indices (both possibly dependent on z), such that $\Pr(z \in N_j) \leq r$, $\forall j \in \mathcal{I}$.

Then, under the foregoing assumptions, $\hat{X}_{\mathrm{MAP}}(\mathbf{Y}^{(M)}) \longrightarrow X$, a.s. as $M \to \infty$, i.e., this MAP estimator is strongly consistent.

Proof: The proof involves three steps. We start by showing that, in the pathwise sense, and for all $M \geq 1$, $X \subseteq \hat{X}_{\mathrm{MAP}}(\mathbf{Y}^{(M)}) \subseteq \cap_{i=1}^{M} Y_i$. The next step is to show that $\cap_{i=1}^{M} Y_i \longrightarrow X$, a.s. as $M \to \infty$ is implied by $\Pr(\cap_{j=1}^{M} N_j = \emptyset) \to 1$, as $M \to \infty$, and complete the proof by showing that this latter condition is implied by Condition 1 in the statement of the theorem.

The first two steps can be found in [4]. We now proceed to prove the third step. Equivalently, we need to show that $\Pr(\cap_{j=1}^{M} N_j \neq \emptyset) \to 0$, as $M \to \infty$. Now

$$\Pr(\cap_{j=1}^{M} N_j \neq \emptyset) = \Pr(z \in \cap_{j=1}^{M} N_j \text{ for some } z \in B)$$

(by the union bound)

$$\leq \sum_{z \in B} \Pr(z \in \cap_{j=1}^{M} N_j)$$

and by independence

$$\sum_{z \in B} \Pr(z \in \cap_{j=1}^{M} N_j) = \sum_{z \in B} \prod_{j=1}^{M} \Pr(z \in N_j).$$

Since $|B| < \infty$, it suffices that $\prod_{j=1}^{M} \Pr(z \in N_j) \to 0$, as $M \to \infty$ for each and every $z \in B$. This is clearly implied by Condition 1, and the proof is complete. □

[4]Observe that this is a condition on marginal noise statistics only.

We now present two more theorems. They can both be established by appealing to duality (note that closing is the dual of opening with respect to lattice complementation). Observe that here we deal with intersection noise, which can be interpreted as a formal mechanism to consider random sampling of DRS's.

Theorem 3—MAP Optimality Dual: Assume we observe $\mathbf{Y}^{(M)} = [Y_1, \cdots, Y_M]$, where $Y_i = X \cap N_i$, $\{N_i\}_{i=1}^{M}$ is an independent *but not necessarily identically distributed* sequence of noise DRS's, which is independent of X, and each N_i is an *otherwise arbitrary DRS* taking values in some arbitrary collection $\Psi_i(B) \subseteq \Sigma(B)$ of subsets of the observation lattice B. Let us further assume that X is uniformly distributed over a collection $\Phi(B) \subseteq \Sigma(B)$ of all subsets K of B, which can be written as $K = \cap_{l=1}^{L} K_l$, $K_l \in C_{W_l}(B)$, $l = 1, \cdots, L$. Then $\hat{X}_{\mathrm{MAP}}(\mathbf{Y}^{(M)}) = \bigcap_{l=1}^{L}((\cup_{i=1}^{M} Y_i) \bullet W_l)$ is a MAP estimator of X on the basis of $\mathbf{Y}^{(M)}$.

Theorem 4—Strong Consistency Dual: In addition, assume the following holds.

Condition 2: $\forall z \in B$ there exists $0 \leq r < 1$, such that $\Pr(z \notin N_j) \leq r$, for infinitely many indices j[5]. In other words, for every $z \in B$, there exists $0 \leq r < 1$, and an infinitely long subsequence $\mathcal{I}$ of observation indices (both possibly dependent on z), such that $\Pr(z \notin N_j) \leq r$, $\forall j \in \mathcal{I}$.

Then, under the foregoing assumptions, $\hat{X}_{\mathrm{MAP}}(\mathbf{Y}^{(M)}) \longrightarrow X$, a.s. as $M \to \infty$, i.e., this MAP estimator is strongly consistent.

IV. Discussion

A little reflection on the above results is in order. The discussion will focus on Theorems 1 and 2, but the remarks are equally applicable to the case of Theorems 3 and 4.

The first observation is that both theorems crucially depend on B being finite[6]. This is obvious at several points in the proofs. We view this as further evidence of the utility of this restriction. The second observation is that the results are fairly general: apart from (mild) Condition 1, which is needed for consistency, and the requirement that $\{N_j\}$ is a sequence of independent DRS's, which is independent of X, we have imposed absolutely no other restrictions on the sequence of noise DRS's $\{N_j\}$.

In general, we cannot derive analytical formulas for some standard measures of estimator performance, such as bias and variance, without specifying the sequence of noise DRS's $\{N_j\}$; this is obvious, since these measures strongly depend on the structure of this sequence. Based on our experience in [4], our feeling is that these derivations are going to be nasty, except in some limited cases. However, it should be noted that the MAP principle leads to optimal estimators in a particular Bayesian sense: it minimizes the total probability of error, P_e [27]. In other words, even though the MAP estimator may not be unbiased and/or minimize the error variance (as a MMSE estimator typically does) it is optimal in the sense that, for each and every M, it minimizes the total probability of error. This is just an alternative concept of optimality.

Let us now consider three special cases.

- The sequence of noise DRS's $\{N_j\}$ is i.i.d., and each noise DRS is i.i.d., i.e., a Bernoulli lattice process of constant intensity. This particular noise process is compatible with earlier results in [4]. In addition to MAP optimality and strong consistency, compatibility with [4] buys *uniqueness* of the functional form of the MAP estimator, and a handle on the bias [4].
- $\{N_j\}$ is a sequence of independent DRS's, which is independent of X, and each N_j is uniformly distributed over $\Psi_j(B) = \Psi(B)$, $\forall j \geq 1$, where $\Psi(B) \subseteq \Sigma(B)$ is a collection of all

[5]Again, this is a condition on marginal noise statistics only.

[6]The size of B can be made as large as one wishes, as long as it is finite.

subsets K of B which are spanned by unions of translates of a family of structural elements, V_l, $l = 1, \cdots, \Lambda$, i.e., those $K \subseteq B$ that can be written as $K = \cup_{l=1}^{\Lambda} K_l$, $K_l \in O_{V_l}(B)$, $l = 1, \cdots, \Lambda$. The noise is now a system of overlapping particles of several different types, i.e., constrained to be smooth with respect to a union of openings by an appropriately chosen family of structural elements. Noise particles overlap with signal particles. It is easy to see that this composite noise process satisfies Condition 1. Regardless of the degree of overlap and the particular types of signal and noise particles, we can claim optimality and strong consistency. However, small sample behavior will be governed by the interplay between the two families of structural elements that span the signal and noise DRS's ($\{W_l\}$, $\{V_l\}$, respectively). For example, if $|V_l| < |W_m|$, $\forall m = 1, \cdots, L$, then application of the $M = 1$ MAP filter will eliminate all isolated instances of V_l noise patterns. This may well be the case in applications, where the signal is usually associated with the more prominent image structures.

- $\{N_j\}$ is a sequence of independent discrete Boolean random sets [5], [6], which is independent of X. This particular case is of great interest, since the Boolean union noise model is arguably one of the best models for clutter. For all practical purposes (i.e., all Boolean models of practical interest), Condition 1 is satisfied, and, therefore, optimality and strong consistency can be warranted.

V. Conclusion

We have revisited the problem of estimating realizations of random sets immersed in random clutter, or suffering from random dropouts, under a much more general set of assumptions, which allows the explicit incorporation of geometric and probabilistic constraints into the noise model, i.e., the noise may now exhibit geometric and probabilistic *structure*; surprisingly, it turns out that this affects neither the optimality nor the consistency of appropriate morphological estimators.

Acknowledgment

The authors are indebted to one anonymous reviewer, whose insightful comments have been very helpful in improving this manuscript.

References

[1] N. D. Sidiropoulos, J. S. Baras, and C. A. Berenstein, "New results on optimality and consistency of morphological openings and closings," in *Proc. Conf. Stochastic, Morphologic., Neural Methods Signal Image Processing, SPIE*, San Diego, CA, July 1995, vol. 2568.

[2] ——, "On morphological openings and closings of signals in shaped noise," in *Proc. IEEE Int. Conf. Image Processing*, Washington, DC, Oct. 1995.

[3] ——, "Optimal filtering of digital binary images corrupted by union/intersection noise," *IEEE Trans. Image Processing*, vol. 3, no. 4, pp. 382–403, 1994.

[4] N. D. Sidiropoulos, D. Meleas, and T. Stragas, "MAP signal estimation in noisy sequences of morphologically smooth images," *IEEE Trans. Image Processing*, to be published. (Also available as Institute for Systems Research ISR TR 94-83.)

[5] N. D. Sidiropoulos, J. S. Baras, and C. A. Berenstein, "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.

[6] N. D. Sidiropoulos, "Statistical inference, filtering, and modeling of discrete random sets," Ph.D. dissertation, Univ. of Maryland, College Park, MD, June 1992.

[7] G. Matheron, *Elements Pour Une Theorie Des Milieux Poreux*. France: Masson, 1967.

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

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

[10] J. Serra, Ed., "Image analysis and mathematical morphology," *Theoretical Advances*. San Diego, CA: Academic, 1988, vol. 2.

[11] H. J. A. M. Heijmans, *Morphological Image Operators*. Boston, MA: Academic, 1994.

[12] J. Serra and L. Vincent, "An overview of morphological filtering," *Circuits, Syst., Signal Processing*, vol. 11, no. 1, pp. 47–108, Jan. 1992.

[13] 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.

[14] 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, no. 8, pp. 1153–1169, Aug. 1987.

[15] 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, no. 8, pp. 1170–1184, Aug. 1987.

[16] I. Pitas and A. N. Venetsanopoulos, *Nonlinear Digital Filters: Principles and Applications*. Boston, MA: Kluwer, 1990.

[17] 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.

[18] R. M. Haralick, E. R. Dougherty, and P. L. Katz, "Model-based morphology," in *Proc. SPIE, vol. 1472*. Orlando, FL: Soc. Optical Eng., Apr. 1991.

[19] 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: Soc. Opt. Eng., Feb. 1991.

[20] E. R. Dougherty, Ed.,

[21] *Mathematical Morphology in Image Processing*. New York: Markel Dekker, 1993.

[22] E. R. Dougherty, "Optimal mean-absolute-error filtering of gray-scale signals by the morphological hit-or-miss transform," *J. Math. Imaging Vision*, vol. 4, pp. 255–271, 1994.

[23] E. R. Dougherty, J. Newell, and J. Pelz, "Morphological texture-based maximum-likelihood pixel classification based on local granulometric moments," *Pattern Recog.*, vol. 25, pp. 1181–1198, 1992.

[24] D. Schonfeld and J. Goutsias, "Optimal morphological pattern restoration from noisy binary images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 13, no. 1, pp. 14–29, Jan. 1991.

[25] D. Schonfeld, "Optimal structuring elements for the morphological pattern restoration of binary images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 16, no. 6, pp. 589–601, June 1994.

[26] D. Schonfeld and J. Goutsias, "On the morphological representation of binary images in a noisy environment," *J. Visual Commun. Image Representation*, vol. 2, no. 1, pp. 17–30, Mar. 1991.

[27] J. Goutsias, "Morphological analysis of discrete random shapes," *J. Math. Imaging Vision*, vol. 2, pp. 193–215, 1992.

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