Technical Research Report

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

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

T.R. 94-84

INSTITUTE FOR SYSTEMS RESEARCH

*Sponsored by
the National Science Foundation
Engineering Research Center Program,
the University of Maryland,
Harvard University,
and Industry*

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— **In two recent papers [1], [2], Sidiropoulos et al. have obtained statistical proofs of Maximum *A Posteriori* (MAP) optimality and strong consistency of certain classes of Morphological filters, namely, Morphological Openings, Closings, unions of Openings, and intersections of Closings, under i.i.d. (both pixel-wise, and sequence-wide) assumptions on the noise model. In this paper we revisit this classic filtering problem, and prove MAP optimality and strong consistency under a different, and, in a sense, more appealing set of assumptions, which allows the explicit incorporation of geometric and Morphological constraints into the noise model, i.e., the noise may now exhibit *structure*; surprisingly, it turns out that this affects neither the optimality nor the consistency of these filters.**

Keywords— **Morphological Image Processing and Analysis, Opening, Closing, Statistical Optimization of Nonlinear Filters, MAP Optimality, Consistency of MAP estimator**

I. Introduction

In two recent papers [1], [2], 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 (or, Discrete Random Set (DRS), for short) theory [3], [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 of some finite $B \subset \mathbf{Z}^2$), and $\Sigma(\Sigma(B))$ is a σ-field over $\Sigma(B)$ (usually, the power set of the power set of B). A DRS X induces an associated probability structure $P_X(\cdot)$ on $\Sigma(\Sigma(B))$.

The optimality results of [1], [2] 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, and a uniformity condition (to be specified) is imposed. The net result is that we end up with a new set of optimality conditions, which neither implies, nor is implied by the previous set. The most interesting feature of this new set of conditions is that it allows the explicit incorporation of geometric and morphological constraints into the noise model, thus establishing optimality in a more flexible and interesting environment.

N.D. Sidiropoulos is with the Institute for Systems Research, University of Maryland, College Park, MD 20742 U.S.A. He can be reached via e-mail at nikos@Glue.umd.edu

J.S. Baras is with the Dept. of Electrical Engineering and the Institute for Systems Research of the University of Maryland

C.A. Berenstein is with the Dept. of Mathematics and the Institute for Systems Research of the University of Maryland

II. Background

The theory of Mathematical Morphology has been developed mainly by Matheron [5], [6], Serra [7], [8], and their collaborators during the 70's and early 80's. morphological Filtering is one of the most popular and successful branches of this theory[1]. 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) are well developed and understood [10], [11], [12], [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], [3], [4]. Dougherty et al. [14], [15], [16], [17], [18], [19], Schonfeld et al. [20], [21], [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 do not reproduce the definitions of basic morphological operators $\oplus, \ominus, \circ, \bullet$ (Minkowski addition and subtraction, and morphological opening and closing, respectively) here; we follow the conventions of [7].

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 (filter) f *iff* it is invariant under that operator, i.e., $f(I) = I$. For example, an image I is smooth with respect to morphological opening by a structural element W *iff* $I \circ W = I$. 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. We shall use $O_W(B)$ to denote the domain of invariance of opening by W, i.e., the collection of all images (subsets of B) which are invariant under opening by W, i.e., spanned by translates of W. Note that $\emptyset \in O_W(B)$, $\forall\, W$. Similarly, we shall use $C_W(B)$ to denote the domain of invariance of closing by W. At times we may also abuse terminology and say that an image is "smooth

[1] See [9] for a recent survey of the status of morphological Filtering

with respect to W"; the meaning should be clear from context. We can also fit more complicated image structure by allowing composite constraints, e.g., consider the class of all images which are invariant under a union of openings with respect to a family of structural elements. Actually this is going to be one of two signal DRS models in what follows.

A drawback of the optimality results of [1], [2] 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 noise. In what follows, this restriction is considerably relaxed by imposing a milder uniformity condition. Furthermore, the sequence-wide assumption of identical noise distribution is completely removed.

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 uniformly distributed over *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[2]

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

Then

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

is *a* MAP estimator of X on the basis of $\mathbf{Y}^{(M)}$.

Proof: Following some manipulations, the MAP principle reduces to

$$\widehat{X}_{MAP}(\mathbf{Y}^{(M)}) =$$

$$argmax_{S \in \Phi(B) \cap \Sigma\left(\cap_{i=1}^M Y_i\right)} Pr(\mathbf{Y}^{(M)} \mid X = S)$$

where $\Sigma\left(\cap_{i=1}^M Y_i\right)$ (the power set of $\cap_{i=1}^M Y_i$) is the sub-σ-field imposed by the observations. By independence

$$\widehat{X}_{MAP}(\mathbf{Y}^{(M)}) =$$

$$argmax_{S \in \Phi(B) \cap \Sigma\left(\cap_{i=1}^M Y_i\right)} \prod_{j=1}^{M} Pr(S \cup N_j = Y_j)$$

$$= argmax_{S \in \Phi(B) \cap \Sigma\left(\cap_{i=1}^M Y_i\right)} G(S)$$

where the *gain functional*, $G : \Phi(B) \cap \Sigma\left(\cap_{i=1}^M Y_i\right) \longrightarrow \mathbf{Z}_+$ is defined as

$$G(S) \triangleq \prod_{j=1}^{M} Card\left\{N_j \in \Psi_j(B), \; N_j \subseteq Y_j \mid S \cup N_j = Y_j\right\}$$

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

where $Card\{\cdot\}$ stands for set cardinality. The following Lemma is elementary:

Lemma 1: $G(\cdot)$ is a non-decreasing functional on the complete lattice $\left(\Phi(B) \cap \Sigma\left(\cap_{i=1}^M Y_i\right), \subseteq\right)$.

The MAP optimality result then follows trivially from the fact that

$$\bigcup_{l=1}^{L} \left(\left(\cap_{i=1}^M Y_i \right) \circ W_l \right)$$

is the maximal element of this lattice. Non-uniqueness of the functional form of the MAP estimator is a direct consequence of the fact that the gain functional is generally not *strictly* increasing. ■

The assumption of uniform distribution clearly buys a lot; it reduces optimization to a counting argument. The range spaces $\{\Psi_j(B)\}$, $\Phi(B)$ are not *quite* as important here; it is the principle of uniformity that counts. The natural question then is what do we really model by using a uniform distribution? A simple answer is that we model a random variable whose range is completely known, but no other piece of information concerning its probabilistic structure is available. Alternatively, we may think of it as modeling an "unbiased" or "fair" adversary. If the noise is "biased", then, depending on the particular type of probabilistic noise structure, and assuming we can uncover this structure, we might well be able to construct better estimators, or, we might not even be able to guarantee consistency. We believe that, in the absence of such information, a uniform distribution approach is both reasonable and prudent.

Theorem 2: (Strong Consistency) In addition, if $\emptyset \in \Psi_i(B)$, $\forall\, i \geq 1$, then, under the foregoing assumptions

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

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 $M \geq 1$

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

The next step is to show that

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

is implied by

$$lim Pr(\cap_{j=1}^M N_j = \emptyset) = 1$$

and complete the proof by showing that if $\emptyset \in \Psi_i(B)$, $\forall\, i \geq 1$, then the latter condition is satisfied.

The essential elements of the first two steps can be found in [2]. We now proceed to prove the third step. Observe that

$$Pr(\cap_{j=1}^M N_j = \emptyset) = Pr(\cap_{j=1}^{M-1} N_j = \emptyset) +$$

$$\sum_{R \in \Theta_{M-1}} Pr(N_M \cap R = \emptyset \mid \cap_{j=1}^{M-1} N_j = R) Pr(\cap_{j=1}^{M-1} N_j = R)$$

where

$$\Theta_M \triangleq \left\{K = \cap_{j=1}^M K_j \neq \emptyset \mid K_j \in \Psi_j(B)\right\}$$

By independence it follows

$$Pr(\cap_{j=1}^{M} N_j = \emptyset) = Pr(\cap_{j=1}^{M-1} N_j = \emptyset) +$$

$$\sum_{R \in \Theta_{M-1}} Pr(N_M \cap R = \emptyset) Pr(\cap_{j=1}^{M-1} N_j = R)$$

Now observe that

$$Pr(N_j \cap R = \emptyset) \geq Pr(N_j = \emptyset), \ \forall R, \ \forall j \geq 1$$

and, under the uniformity assumption, $\emptyset \in \Psi_i(B), \ \forall \ i \geq 1$ implies that[3]

$$Pr(N_j = \emptyset) = \frac{1}{|\Psi_j(B)|} \triangleq \pi_j 0$$

therefore, taking $r = inf_j \pi_j 0$

$$Pr(N_j \cap R = \emptyset) \geq r, \ \forall R, \ \forall j \geq 1$$

we obtain

$$Pr(\cap_{j=1}^{M} N_j = \emptyset) \geq Pr(\cap_{j=1}^{M-1} N_j = \emptyset) +$$

$$r \sum_{R \in \Theta_{M-1}} Pr(\cap_{j=1}^{M-1} N_j = R)$$

$$= Pr(\cap_{j=1}^{M-1} N_j = \emptyset) + r \left(1 - Pr(\cap_{j=1}^{M-1} N_j = \emptyset)\right)$$

For notational convenience we define

$$p_M \triangleq Pr(\cap_{j=1}^{M} N_j = \emptyset)$$

and the latter inequality becomes

$$p_M \geq p_{M-1} + r(1 - p_{M-1})$$

i.e.,

$$p_M \geq (1-r) p_{M-1} + r$$

Solving this inequality we get

$$p_M \geq 1 - (1-r)^{M-1} + (1-r)^{M-1} p_1$$

$$= 1 - (1-r)^{M-1}(1 - p_1) \longrightarrow 1, \ as \ M \to \infty$$

Since p_M is a valid probability, we conclude that $p_M \longrightarrow 1, \ as \ M \to \infty$, and the proof is complete. ■

Remark: In fact we may slightly relax the condition $\emptyset \in \Psi_i(B), \ \forall \ i \geq 1$, by allowing it to be violated for finitely many i's.

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[4].

[3]Note that this step crucially depends on B (and, therefore, $\Psi_i(B)$) being *finite*.

[4]See [24] for an account of an interesting approach when N is assumed to be a deterministic regularly spaced grid which undersamples the observation.

Theorem 3: 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 uniformly distributed over *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

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

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

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

where $\bullet$ stands for the closing operation, is *a* MAP estimator of X on the basis of $\mathbf{Y}^{(M)}$.

Theorem 4: In addition, if $B \in \Psi_i(B), \ \forall \ i \geq 1$, then, under the foregoing assumptions

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

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 1,2, but the remarks are equally applicable to the case of Theorems 3,4.

The first observation is that both theorems crucially depend on B being finite[5]. 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 the mild condition $\emptyset \in \Psi_i(B), \ \forall \ i \geq 1$, which is needed for consistency, we have imposed absolutely no other restrictions on the sequence of range spaces $\{\Psi_j(B)\}$ of the noise DRS's $\{N_j\}$; some particular examples will be given in the subsection that follows. Given the generality of the results, the proofs appear to be surprisingly simple.

In general, we cannot derive analytical formulas for some standard measures of estimator performance, such as bias and variance, without specifying the sequence of range spaces $\{\Psi_j(B)\}$ of the noise DRS's $\{N_j\}$; this is obvious, since these measures strongly depend on the structure of this sequence. Based on our experience in [2], 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 [25]. 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.

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

A. Some Special Cases

Let us now consider two special cases. Again, our discussion will focus on Theorems 1,2, but the remarks are equally applicable to the case of Theorems 3,4.

• $\Psi_j(B) = \Sigma(B), \ \forall\ j \geq 1$: The noise DRS's are identically distributed, each noise DRS is uniformly distributed over the power set of B. This is in fact the only nontrivial noise distribution compatible both with our earlier results in [2], and with our results herein. This corresponds to the case of an i.i.d. sequence of i.i.d. DRS's, each being a Bernoulli lattice process of constant intensity $\lambda = \frac{1}{2}$. In addition to MAP optimality and strong consistency, compatibility with [2] buys *uniqueness* of the functional form of the MAP estimator, and a handle on the bias.

• $\Psi_j(B) = \Psi(B), \ \forall\ j \geq 1$, where $\Psi(B) \subseteq \Sigma(B)$, is a collection of all 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 $K \subseteq B$ which can be written as

$$K = \cup_{l=1}^{\Lambda} K_l, \quad K_l \in O_{V_l}(B), \ l = 1, \cdots \Lambda$$

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. Regardless of the degree of overlap and the particular types of signal and noise particles, we can claim optimality and strong consistency[6]. However, small sample behavior will be governed by the interplay between the two families of structural elements which 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.

V. Conclusions

In this paper we have revisited a classic filtering problem, that of estimating realizations of random sets immersed in random clutter, or suffering from random dropouts. We have established MAP optimality and strong consistency of certain classes of morphological filters under a new, and, in a sense, more appealing set of assumptions, which allows the explicit incorporation of geometric and morphological constraints into the noise model, i.e., the noise may now exhibit *structure*; Surprisingly, it turns out that this affects neither the optimality nor the consistency of these filters.

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, no. 4, pp. 382–403, 1994.

[2] N.D. Sidiropoulos, D. Meleas, and T. Stragas, "MAP Signal Estimation in Noisy Sequences of Morphologically Smooth Images", Submitted for publication, IEEE Trans. Image Processing. Also available as Institute for Systems Research ISR TR 94-83.

[3] 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", *Journal of Mathematical Imaging and Vision*, vol. 4, pp. 273–290, 1994.

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

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

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

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

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

[9] J. Serra and L. Vincent, "An Overview of Morphological Filtering", *Circuits, Systems and Signal Processing*, vol. 11, no. 1, pp. 47–108, January 1992.

[10] J. Song and E.J. Delp, "A Study of the Generalized Morphological Filter", *Circuits, Systems and Signal Processing*, vol. 11, no. 1, pp. 229–252, January 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. on Acoustics, Speech and Signal Proc.*, vol. ASSP-35, no. 8, 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. on Acoustics, Speech and Signal Proc.*, vol. ASSP-35, no. 8, pp. 1170–1184, Aug. 1987.

[13] I. Pitas and A. N. Venetsanopoulos, *Nonlinear Digital Filters: Principles and Applications*, Kluwer, Boston, 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, January 1992.

[15] 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.

[16] 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.

[17] E. R. Dougherty, Ed., *Mathematical Morphology in Image Processing*, Marcel Dekker, New York, 1993.

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

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

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

[21] D. Schonfeld, "Optimal Structuring Elements for the Morphological Pattern Restoration of Binary Images", *IEEE Transactions on Pattern Anal. Mach. Intell.*, vol. 16, no. 6, pp. 589–601, June 1994.

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

[23] J. Goutsias, "Morphological Analysis of Discrete Random Shapes", *Journal of Mathematical Imaging and Vision*, vol. 2, pp. 193–215, 1992.

[24] R.M. Haralick, X. Zhuang, C. Lin, and J.S.J. Lee, "The Digital Morphological Sampling Theorem", *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. 37, no. 12, pp. 2067–2089, 1989.

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

[6] Observe that the consistency condition is automatically satisfied here, since $\emptyset$ is open with respect to *all* structural elements.