DESIGN OF MORPHOLOGICAL SET OPERATORS BY STATISTICAL OPTIMIZATION

J. Barrera

University of São Paulo. Computer Science Department.
São Paulo, SP, Brazil. e-mail: jb@ime.usp.br

E. R. Dougherty

Texas A&M University. Texas Center for Applied Technology. Department of Electrical Engineering.
College Station, Texas, USA. e-mail: edward@ee.tamu.edu

ABSTRACT

The design of operators (e.g. filters) for Signal or Image Processing requires an algebraic decomposition structure, to represent the family of operators considered, and statistical optimization techniques, defined over the space of operator decompositions or over another isomorphic structure. This paper surveys a technique for designing set operators based on Mathematical Morphology decompositions and optimization algorithms over Boolean lattices. This technique has an intrinsic discrete nature and usually depends on combinatorial optimization algorithms. Its most remarkable quality is the conjunction of formality and pragmatism. In fact, it is both a strong Mathematical theory and a powerful computational tool.

1. INTRODUCTION

A central problem in Binary Image Analysis is the design of efficient image processing procedures to perform desired tasks. Recent research has addressed the construction of computational systems to design automatically image procedures. These systems require suitable knowledge representation formalisms, for high level of abstraction description of the desired procedure, and computational tools, for translating the formal description into a computational realization. In this paper, we survey a theory for the construction of a family of such systems, that uses examples (i.e., collections of input-output image pairs) as the knowledge representation formalism [2][9][14].

A natural model for a procedure used in binary image analysis is a set mapping (operator) applied to a discrete random set. Procedure design from examples can be modeled by statistical estimation of set operators from observations of input-output image pairs. Formulation of this statistical approach can be decomposed in two basic steps: (1) estimation of the operator input **S** and ideal output **I** random sets; (2) design of a set operator Ψ such that $\Psi(\mathbf{S})$ is statistically close to **I**. The second step can be formulated as an optimization problem: given a family $\Im$ of set operators, where the distance between the ideal **I** and estimator $\Psi(\mathbf{S})$ random sets is measured by a probabilistic error measure. An operator Ψ_{opt} is called *optimal* in $\Im$ if it posses minimum error. If all elements of $\Im$ can be characterized by some algebraic representation, then optimization can be viewed as finding an efficient representation defining an optimal operator.

Mathematical Morphology is a general framework to study set operators. A central paradigm in Mathematical Morphology is the representation of set operators by concatenations of erosions and dilations via the operations of composition, intersection, union and complementation. This paradigm can be stated by the use of a formal language: the *morphological language*, whose vocabulary are erosions, dilations, intersection, union and complementation. Using the morphological language to represent set operators, optimization consists of finding a phrase structure and estimating the parameters, called *structuring elements*, that characterize an optimal operator.

The set operators that are translation invariant and locally defined in a window *W* are called *W-operators*. Any W-operator can be represented by a single phrase structure, called the *standard morphological representation* [10]. Therefore, optimization in the space of W-operators consists in estimating the structuring elements of the standard morphological representation.

This paper reviews techniques for designing W-operators and sub-families of W-operators. Following this Introduction, Section 2 presents the system modeling adopted. Section 3 recalls the standard morphological representation of W-operators. Section 4 presents the statistical optimization formulation. Section 5 studies the estimation of optimal operators. Section 6 gives a detailed computational procedure for estimating operators. Section 7 presents an application example. Finally, Section 8 discusses the state of the art in the field.

2. SYSTEM MODELING

We call **S**, **I** and $\Psi(\mathbf{S})$ the *observation*, *ideal* and *estimator* random processes, respectively. If *S* is a realization of **S**, then $\Psi(S)$ is the corresponding realization of $\Psi(\mathbf{S})$. The distance of **I** and $\Psi(\mathbf{S})$ is measured by a probabilistic error $\varepsilon[\mathbf{I}, \Psi(\mathbf{S})]$.

Assuming that operators belong to an operator family $\Im$, an *optimal operator* relative to $\Im$ is an operator $\Psi_{opt} \in \Im$ such that $\varepsilon[\mathbf{I}, \Psi_{opt}(\mathbf{S})] \leq \varepsilon[\mathbf{I}, \Psi(\mathbf{S})]$, for all $\Psi \in \Im$.

Operationally, the optimization model includes a *system transformation* Ξ such that $\mathbf{S} = \Xi(\mathbf{I})$, that is, the observation process is assumed to be the output of some system operating on the ideal process. Optimization involves minimizing the error $\varepsilon[\mathbf{I}, \Psi(\Xi(\mathbf{I})]$. In general, Ξ is a multi-valued image operator, meaning that given a realization I of the ideal image, there are many possible output realizations $\Xi(I)$. Essentially, there is an inverse problem: we wish to design Ψ to recover information observed through Ξ.

A much studied application occurs when Ξ is a degradation transformation: the ideal image is obscured by noise. Ξ can take many forms, depending on the physical system causing the degradation. For instance, if $\mathbf{N}$ is a binary process, then the *signal-union noise model* is defined by the transformation $\mathbf{S} = \Xi(\mathbf{I}) = \mathbf{I} \cup \mathbf{N}$. Rather than union noise, the noise might be subtractive, so $\Xi(\mathbf{I}) = \mathbf{I} - \mathbf{N}$. There can be two noise processes $\mathbf{N_1}$ and $\mathbf{N_2}$, with the degradation process both adjoining and deleting pixels, so that $\Xi(\mathbf{I}) = (\mathbf{I} \cup \mathbf{N_1}) - \mathbf{N_2}$.

3. STANDARD REPRESENTATION OF W-OPERATORS

Let E denote the integer plane Z^2. *A binary image* X is a function from E to {0,1). Let $P(E)$ be the power set of E. An equivalent representation of the binary image X is the element X of $P(E)$., defined by, for any $x \in X \Leftrightarrow X(x)=1$. In this paper, we will use both representations for binary images, denoting the function and the set representation by the same upper case letter.

The set E is an Abelian group with respect to the vector addition, denoted by +. The zero element of $(E,+)$ is the origin of E and is denoted by o. Let $X \in P(E)$ and $u \in E$. The translation of X by u is the element of $P(E)$ defined by $X_u = \{x \in E: x-u \in X\}$.

A mapping from $P(E)$ to $P(E)$ is called a *binary image operator* or, for simplicity, just *operator*. We denote set operators by upper case Greek letters Λ, Ψ, ... A set operator Ψ is called *translation invariant* iff, for any $u \in E$ and $X \in P(E)$, $\Psi(X_u) = \Psi(X)_u$. Let W be a finite subset of E. A set operator is called *locally defined* in the window W iff, for any $x \in E$ and $X \in P(E)$, $u \in \Psi(X) \Leftrightarrow u \in \Psi(X \cap W_u)$. A set operator that is both translation invariant and locally defined in W is called a W-operator. The family of W-operators is denoted Ψ_W. The W-operators have a standard morphological representation [1],[10].

The next definitions are necessary for presenting this representation. The *kernel* of an operator $\Psi \in \Psi_W$ is the subcollection of $P(W)$ defined by $K(\Psi)=\{X \in P(W): o \in \Psi(X)\}$.

Let $A,B \in P(W)$ such that $A \subset B$. The *sup-generating operator* characterized by the structuring elements A and B is the set operator defined by, for any $X \in P(E)$,

$$\Lambda_{A,B}(X)=\{x \in E: A \subset X_{-x} \cap W \subset B\}.$$

The subcollection $[A,B]$ of $P(W)$ defined by $[A,B]=\{X \in P(W): A \subset X \subset B\}$ is called an *interval*. Any W-operator Ψ can be represented by an union of sup-generating operators characterized by structuring elements in $K(\Psi)$, that is, for any $X \in P(E)$,

$$\Psi(X) = \cup\{\Lambda_{A,B}(X): [A,B] \subset K(\Psi)\}.$$

Therefore, in order to characterize a W-operator it is enough to have its kernel. This representation can also be simplified by the introduction of the notion of basis of a W-operator.

Any W-operator Ψ can also be represented by, for any $X \in P(E)$,

$$\Psi(X) = \{x \in E: \psi(X_{-x} \cap W) = 1\},$$

where ψ is a Boolean function from $P(W)$ to {0,1} such that, for any $Y \in P(W)$,

$$\psi(Y) = 1 \Leftrightarrow Y \in K(\Psi).$$

ψ is called the *window function* of Ψ.

4. STATISTICAL OPTIMIZATION FORMULATION

Estimation of $\mathbf{I}$ from $\Xi(\mathbf{I})$ by a W-operator Ψ requires finding a Boolean function ψ to minimize error [2]. Since Ψ is translation-invariant, we make the modeling assumption that $\mathbf{I}$ and $\Xi(\mathbf{I})$ are jointly strict-sense stationary. This means that, if $\mathbf{X}$ is the random vector of binary values in W_z and $Y = \mathbf{I}(z)$, then the joint probability distribution for $\mathbf{X}$ and Y is independent of z, so that estimating Y from $\mathbf{X}$ yields a translation-invariant operator.

For operator optimization, we require a *loss function* $l: \{0,1\}^2 \rightarrow [0, \infty)$, where $l(a, b)$ measures the cost of the difference between a and b, with $l(0, 0) = l(1, 1) = 0$. Relative to the loss function (and owing to stationarity), operator error, $Er\langle\Psi\rangle$, is given by the expected loss from estimating $\mathbf{I}(z)$ by $\Psi(\Xi(\mathbf{I}))(z)$),

$$Er\langle\Psi\rangle = Er[\mathbf{I},\Psi(\Xi(\mathbf{I}))] = E[l(\mathbf{I}(z), \Psi(\Xi(\mathbf{I}))(z))],$$

where z is an arbitrary pixel.

An optimal image operator is one whose Boolean function ψ minimizes $Er\langle\Psi\rangle$. Although there can be more than one operator achieving minimal error, we shall denote "the" optimal operator and its window function by Ψ_{opt} and ψ_{opt}, respectively, the convention being that, from the standpoint of operator optimization, all operators possessing minimal error are equivalent.

The mean-absolute-error (MAE) loss function is defined by

$$l(y, \psi(\mathbf{x})) = |y - \psi(\mathbf{x})|$$

Since y and $\psi(\mathbf{x})$ are binary valued, the loss function is numerically defined by $l(1, 0) = l(0, 1) = 1$ and $l(0, 0) = l(1, 1) = 0$. The associated error is the *mean absolute error* (*MAE*) and is denoted by $MAE\langle\Psi\rangle$. Because $l(1, 0) = l(0, 1)$, it follows that the optimal Boolean function and the error of the corresponding optimal set operator are given, respectively, by

$$\psi_{MAE}(\mathbf{x}) = \begin{cases} 1, \text{if } P(Y=1\,|\,\mathbf{x}) > 0.5 \\ 0, \text{if } P(Y=1\,|\,\mathbf{x}) \le 0.5 \end{cases}$$

and

$$MAE\langle\Psi_{MAE}\rangle = E[|Y - \psi(\mathbf{X})|] =$$

$$\sum_{\{\mathbf{x}: P(Y=1|\mathbf{x})>0.5\}} P(\mathbf{x})P(Y=0\,|\,\mathbf{x}) + \sum_{\{\mathbf{x}: P(Y=1|\mathbf{x})\le 0.5\}} P(\mathbf{x})P(Y=1\,|\,\mathbf{x}) .$$

5. ESTIMATION OF OPTIMAL OPERATORS

In practice, the optimal operator is statistically estimated from image realizations by estimating the conditional expectations composing the decision criterion [2]. This is accomplished by taking image-pair realizations (I_1, S_1), (I_2, S_2),.., (I_m, S_m) of $\mathbf{I}$ and $\mathbf{S} = \Xi(\mathbf{I})$, and forming estimators

$$\hat{e}_{l,\mathbf{x}}(0) = \hat{E}\,[l(Y, 0)\,|\,\mathbf{x}]$$

$$\hat{e}_{l,\mathbf{x}}(1) = \hat{E}\,[l(Y, 1)\,|\,\mathbf{x}],$$

where $\hat{E}$ is the estimation of the expectation.

The designed estimate of the optimal operator, $\hat{\Psi}_{opt}$ (with window function $\hat{\psi}_{opt}$), is determined by the set $K[\hat{\psi}_{opt}]$ of observation vectors $\mathbf{x}$ for which $\hat{e}_{l,\mathbf{x}}(1) < \hat{e}_{l,\mathbf{x}}(0)$

There are two types of estimation error: $\mathbf{x} \in K[\Psi_{opt}]$ but $\mathbf{x} \notin K[\hat{\Psi}_{opt}]$; and $\mathbf{x} \notin K[[\Psi_{opt}]$ but $\mathbf{x} \in K[\hat{\Psi}_{opt}]$. Therefore, $\hat{\Psi}_{opt}$ is the suboptimal operator being used in place of Ψ_{opt}.

For the MAE loss function, we use the estimator

$$\hat{P}(Y = k\,|\,\mathbf{x}) = \frac{Card[Y = k\,|\,\mathbf{x}]}{Card[\mathbf{x}]}$$

for $k = 0, 1$, where *Card* denotes set cardinality and the numerator and denominator give the number of times the sample ideal images are k-valued given $\mathbf{x}$ and the number times $\mathbf{x}$ is observed across the sample, respectively. $\mathbf{x} \in K[\hat{\Psi}_{MAE}]$ if and only if $\hat{P}(1 = k\,|\,\mathbf{x}) > \hat{P}(0 = k\,|\,\mathbf{x})$.

The *error increase* is the error for taking the suboptimal operator $\hat{\Psi}_{opt}$ in place of the optimal operator Ψ_{opt}, that is, the difference $Er\langle\hat{\Psi}_{opt}\rangle - Er\langle\Psi_{opt}\rangle$]. In fact, $K[\hat{\Psi}_{opt}]$ is a random collection depending on the realizations selected. Thus, $Er\langle\hat{\Psi}_{opt}\rangle - Er\langle\Psi_{opt}\rangle$ is a random variable and we measure the precision with which $\hat{\Psi}_{opt}$ estimates Ψ_{opt} by the expected error increase, $\rho_{opt} = E[Er\langle\hat{\Psi}_{opt}\rangle - Er\langle\Psi_{opt}\rangle]$.

6. DETAILED DESIGN PROCEDURE

Design procedure of set operators for binary image analysis is composed of two main steps [2]: (1) estimation of conditional probabilities; (2) computation of an operator of minimal cost in accordance with a given loss function and the estimated conditional probabilities. The detailed design procedure may be outlined as follows:

1. Shift the window to all pixel locations within the observation image;
2. At each location, record the observed shape;
3. At each location, record the value of the pixel in the ideal image that is colocated with the window origin in the observation image;
4. For each shape, tally the number of times a one is observed in the ideal image and the number of times a zero is observed;
5. For the operator representation select, for each shape observed, the value (0 or 1) of minimal cost;
6. To reduce the representation cost, perform logic minimization assuming that the unobserved templates are don’t cares;

Except for the fact that a large sample may be required, the computational cost of the first five steps is low; however, the computational cost of the sixth step may be high. Step 5 creates a truth table defining a family of statistically equivalent Boolean functions. Each function in the family possesses a large number of representations. An ideal procedure for logical reduction would give the best representation (i.e., the one using a minimal number of logic gates) among all possible representations of all equivalent functions. In practical applications, such a procedure is usually not available. Typically, the best canonical representation of the equivalent Boolean functions (or, equivalently, the basis of the corresponding set operator) is found. The Incremental Splitting of Intervals (ISI) algorithm was created to perform that efficiently on functions with a large number of variables [2][9].

7. AN APPLICATION EXAMPLE

In this section, we apply the theory presented to design a W-operator that detects a texture in the binary image of a map. The training example is a pair of images, where the input (Fig. 1) is a rectangular region of the map and the output (Fig. 2) is the desired texture found in this region.

The application of the designed operator on the test image (i.e., the complete image of the map - Fig. 3) detects, with small error, the desired texture (Fig. 4).

The window used is the 5x5 square centered at the origin. The training sample is of size 90,126. There are 5,098 distinct observed examples, 5,024 negatives and 74 positives. Learning time is 1s. The resulting basis is composed by 12 intervals.

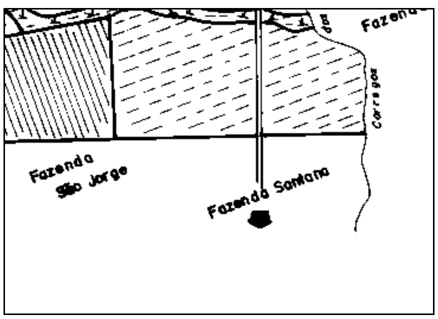

Figure 1. Observed image.

8. DISCUSSION

The theory presented gives a structure to the problem of designing set operators over discrete Boolean lattices, but it is far from being complete.

The main subjects under research nowadays are the design of operators that depend on large windows, the search of optimal morphological language structures to represent W-operators

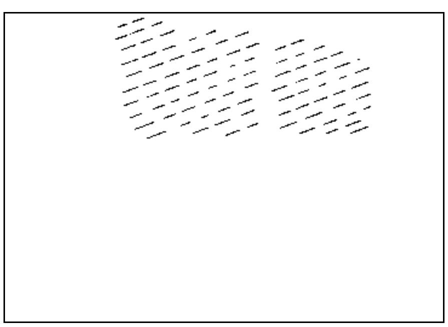

Figure 2. ideal image.

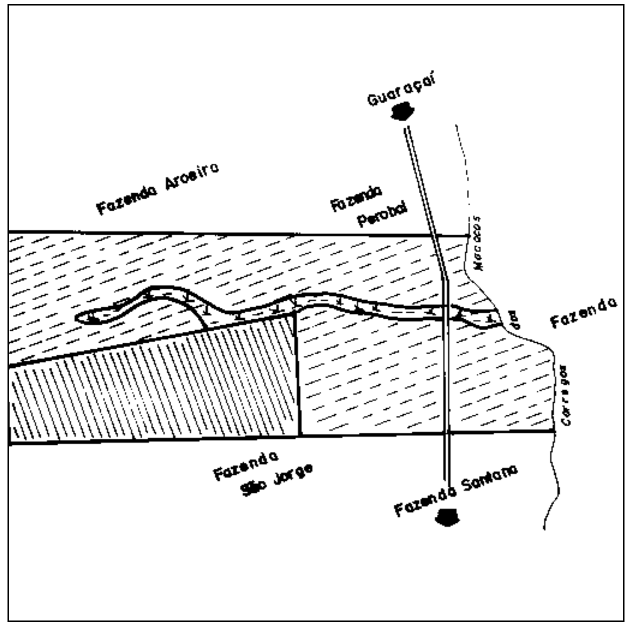

Figure 3. Test image.

and the generalization of the design approach to Signal and Gray-Scale Image Processing.

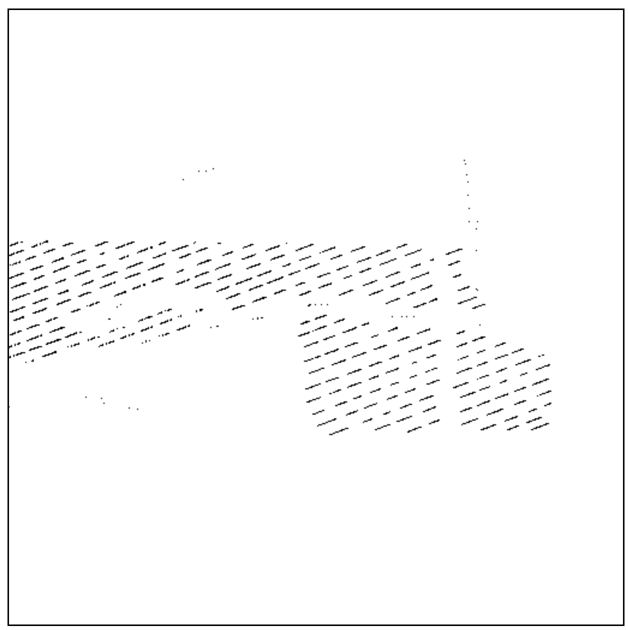

Figure 4. Resulting image.

The problem of designing operators that depend on large windows is classifying correctly a very large number of "don't cares" that appear by the relatively small number of training data available. The solution for this is adding external knowledge about the family of W-operators considered. This is done technically by introducing algebraic constraints on the family of W-operators [8][9]. Results in this direction have been obtained for increasing operators [4][19][15], envelope constraint [13] and iterative operators [5]. Particularly, the increasing and iterative constraints can lead to hard combinatorial problems. Besides algebraic constraint, one can also employ prior information for either the operator [20] or the probability structure [21].

Optimizing the representation structure of Boolean functions, in general, is not the best that can be done for optimizing the representation of W-operators. These operators have local and translation invariant properties that can not be explored when looking just to Boolean functions. The proper view of the problem is looking for efficient structures for phrases of the morphological language from syntactical transformations of the standard representation. However, this formulation leads to extremely hard combinatorial problems. Some preliminary results on this subject are the automatic approach for computing the equivalence between morphological representations [10], the compact representation [12] and the generic algorithm for finding equivalent compositions of dilations and erosions [11].

The generalization of the standard morphological representation to operators defined between complete lattices [3] gives the algebraic framework to the generalization of the statistical design technique to discrete function (i.e., signals or images) operators. In this more general case, we do not have a Boolean lattice, but we still have a finite complete lattice well suited for statistical estimation and combinatorial optimization techniques. Some preliminary results on representation [6] and design of discrete functions operators [7][16][17][18][22] show the potential of the approach and give a new dimension to non linear operator design.

9. REFERENCES

[1] Banon, G. J. F., and J. Barrera, "Minimal Representation of Translation-Invariant Set Mappings by Mathematical. Morphology," *SIAM J. Applied Mathematics*," **51** (6), 1991.

[2] Barrera, J., E. R. Dougherty, and N. S. Tomita, Automatic Programming of Binary Morphological Machines by Design of Statistically Optimal Operators in the Context of Computational Learning Theory," *Journal of Electronic Imaging*, **6** (1), pp. 54-67, 1997.

[3] Banon, G. J. F., and J. Barrera, "Decomposition of Mappings Between Complete Lattices by Mathematical Morphology, Part I. General Lattices," *Signal Processing*, **30**, 1993.

[4] Hirata, N. S. T., J. Barrera, and E. R. Dougherty, "An Efficient Switching Algorithm for Designing Increasing Binary Filters," *Proc. SPIE Nonlinear Image Processing*, January, 1999.

[5] Hirata, N. S. T, E. R. Dougherty, and J. Barrera, "Design of Large-Window Binary Filters via Iteration," *Proc. SPIE Intelligent Robots and Computer Vision Algorithms, Techniques,and Active Vision*, **3522**, November, 1998.

[6] Barrera, J., and E. R. Dougherty, "Representation of Gray-Scale Windowed Operators," *Fourth ISMM*,, pp. 19-26, Amsterdam, June, 1998.

[7] Hirata, R., E. R. Dougherty, and J. Barrera, "Optimal Range-Domain Window Filters," *Proc. SPIE Nonlinear Image Processing*, January, 1999.

[8] Dougherty, E. R., and J. Barrera, "Prior Information in the Design of Optimal Binary Filters," *Fourth ISMM*, pp. 259-266, Amsterdam, June, 1998.

[9] Dougherty, E. R., and J. Barrera, "Logical Image Operators," in *Nonlinear Filters in Image Processing* eds. E. R. Dougherty and J. T. Astola, SPIE and IEEE, Presses, Belingham, 1999.

[10] Barrera, J. and G. P. Salas Set operations on collections of closed intervals and their applications to the automatic programming of Morphological Machines. *Journal of Electronic Imaging*, special section on Non Linear Image Processing, pp 335-352, July 1996.

[11] Barrera, J., C. E. Ferreira, R. Hashimoto Finding optimal sequential decompositions of erosions and dilations. *Fourth ISMM*, pp. 299-306 , Amsterdam, June, 1998.

[12] Barrera, J. and R. F. Hashimoto Compact representation of W-operators. *Proc. SPIE on Non Linear Image Processing IX*, pp 84-94, San Jose, February, 1998.

[13] Barrera, J., E. R. Dougherty, and M. Brun Hybrid human-machine morphological operator design. *Proc. SPIE Nonlinear Image Processing X*, January, 1999.

[14] Dougherty, E. R. and R. P. Loce "Optimal mean-absolute-error hit-or-miss filters: morphological representation and estimation of the binary conditional expectation," *Optical Engineering*, 32(4), pp. 815-827, 1993.

[15] Loce, R. P. and E.R. Dougherty, "Optimal morphological restoration: the morphological filter mean-absolute-error theorem," *Visual Communication and Image Representation*, 3(4), pp. 412-432, 1992.

[16] Dougherty E. R. and D. Sinha, "Computational gray-scale mathematical morphology on lattices (A comparator-based image algebra) Part I: architecture, *Real-Time Imaging*, 1 (1), pp. 69-85, 1995.

[17] Dougherty, E. R. and D. Sinha. "Computational mathematical morphology," Signal Processing, 38, pp. 21-29, 1994.

[18] Loce, R. P., and E. R. Dougherty, "Mean-absolute-error representation and optimization of computational-morphological filters," *CVGIP Graphical Models and Image Processing*, 57(1), 1995.

[19] Dougherty, E. R., "Optimal Mean-Square *N*- Observation Digital Morphological Filters – Part I: Optimal Binary Filters," *CVGIP: Image Understanding*, **55** (1), 1992

[20] Barrera, J., E. R. Dougherty and N. S. Hirata, "Design of Optimal Morphological Operators Using Prior Filters," *Acta Stereologica*, **16** (3), pp. 193-200, 1997.

[21] Dougherty, E. R. and J. Barrera, "Bayesian Design of Optimal Morphological Operators Based on Prior Distributions for Conditional Probabilities", *Acta Stereologica*, **16** (3), pp. 167-174, 1997.

[22] Dougherty, E. R. and J. Barrera, "Computational Gray-scale Operators," in *Nonlinear Filters in Image Processing*, eds. E. R. Dougherty and J. T. Astola, SPIE and IEEE Presses, Belingham, 1999.