Optimal Single Stage Restoration of Subtractive Noise

Larry R. Rystrom\textsuperscript{1,2} \hspace{1cm} Robert M. Haralick\textsuperscript{2,1} \hspace{1cm} Philip L. Katz\textsuperscript{1,2}

\textsuperscript{1}Applied Physics Laboratory, \\
\textsuperscript{2}Department of Electrical Engineering \\
University of Washington, Seattle, WA 98195.

Abstract

This paper analyzes restoration of subtractive noise on a binary image by a single morphological operation, dilation. Restoration by dilation alone is appropriate under particular explicitly defined random noise models, based respectively on erosion, independent pixel subtractive noise, and independent pixel subtractive noise followed by dilation. Since in general it is not possible to perfectly restore subtractive noise we use the Hausdorff metric to measure the residual error in restoration. This metric is the appropriate one because of its geometric interpretation in terms of set coverings.

We describe a search procedure to find a structuring element for dilation that is optimal in the sense of minimizing the mean Hausdorff error. The search procedure's utility function is based on the calculation of certain probabilities related to the noise model, namely the probability of one set being the subset of another set and some related probabilities.

1 Introduction

After a brief review of morphology, we give three different noise models that describe the noise or image corruption process. It is only subtractive noise that motivates restoration by dilation alone. Since in general it is not possible to perfectly restore subtractive noise we use the Hausdorff metric to measure the residual error in restoration. After describing the properties of the Hausdorff metric we then describe a combinatorial search procedure to find a structuring element for dilation that minimizes the mean Hausdorff error. We describe how to calculate the mean Hausdorff error in terms of set coverage probabilities as well as describe how to compute an underestimate of the coverage probabilities (necessary for optimal search). We also tell how to find an initial starting point for the search and a sufficient termination condition.

2 Brief Review of Mathematical Morphology

The two basic operations of Mathematical Morphology are dilation and erosion. References [1], [2], [3] and [4] each give a comprehensive review of the mathematical properties associated with each.

Definition Binary dilation of a set $C$ by a set $L$ is defined as $C \uplus L = \{x|z = c + l \text{ for some } c \in C \text{ and for some } l \in L\}$.

Definition Binary erosion of a set $C$ by a set $L$ is defined as $C \ominus L = \{x|z + l \in C \text{ for all } l \in L\}$.

Definition The reflection of set $L$, $\tilde{L}$, is defined as $\tilde{L} = \{x|\text{for some } l \in L, x = -l\}$.

Definition Translation of a set $C$ by element $t$ is defined as $C_t = \{x|\text{for some } c \in C, x = c + t\}$.

Translation is useful because binary dilation can be written as a union of translates $C \uplus L = \bigcup_{t \in L} C_t$.

It is important to note that unlike erosion, dilation is commutative, that is $C \uplus L = L \uplus C$. Thus there is no difference between specifying $L$ and trying to find an "optimal" $C$ and the problem of specifying $C$ and trying to find $L$.

Both erosion and dilation preserve order. Operators that have this property are increasing, thus if $C \subseteq B$ then $C \ominus L \subseteq B \ominus L$ and $C \uplus L \subseteq B \uplus L$. Erosion and dilation are duals in the sense that $(C \ominus L)^c = C^c \uplus L$.
HAUSDORF SET METRIC
(STICK-OUT)

ONE-SIDED HAUSDORF PSEUDOMETRIC \( \rho \)

\[
\rho(X, Y) = \max \text{ distance any point in } X \text{ lies from its closest point in } Y
\]

HAUSDORF (SYMmetric) METRIC \( \rho_M \)

\[
\rho_M(X, Y) = \max \{ \rho(X, Y), \rho(Y, X) \}
\]

Figure 1: Illustrating the Hausdorff distance between two sets.

The primitives of erosion and dilation can be combined into more complex operations. An opening is an erosion followed by a dilation, \( C \circ L = (C \ominus L) \oplus L \), and closing is a dilation followed by an erosion, \( C \bullet L = (C \oplus L) \ominus L \).

Opening is anti-extensive: the input always contains the output, i.e. \( C \circ L \subseteq C \). Closing is extensive, i.e. \( C \subseteq C \bullet L \). Erosions and dilations are necessarily anti-extensive and extensive respectively only when they contain the origin.

3 Hausdorff Metric

The Hausdorff metric ([1], [4], p.73, [5], [6]) measures the distance between two sets in a way that directly relates to mathematical morphology. Following [5], we first define a one sided pseudodistance between two sets \( X \) and \( Y \).

**Definition** \( \rho(X, Y) = \max_{x \in X} \min_{y \in Y} \| x - y \| \).

\( \rho(X, Y) \) measures the amount of “stick out” of \( X \) from \( Y \), see Figure 1. The Hausdorff distance between two sets is defined as

**Definition** \( \rho_M(X, Y) = \max \{ \rho(X, Y), \rho(Y, X) \} \).

The Hausdorff metric (see [5] for proof that it is a metric) is intimately connected to dilation with a disk, with the disk radius depending on the amount of “stick out” as the following proposition tells us.

**Proposition 1** Let \( \text{disk}_a(r) = \{ x \| \| x \| \leq r \} \), and \( X, Y \in F^N \).

Then \( \max_{x \in X} \min_{y \in Y} \| x - y \| = \inf \{ r | X \subseteq Y \oplus \text{disk}_a(r) \} \).

**Proof:** See [5].

Here \( \text{disk}_a(r) \) is a closed (in a topological sense) analog disk.

We can restate the previous proposition as follows:

**Proposition 2** If \( \rho(X, Y) = r \) then \( X \subseteq Y \oplus \text{disk}_a(r) \) and \( X \not\subseteq Y \oplus \text{disk}_a(r') \) for \( r' < r \).
Proof: See proposition 1. ■

Thus the following proposition holds.

Proposition 1. \( \rho_M(X, Y) \leq r \) if and only if \( X \subseteq Y \triangleq disk_a(r) \) and \( Y \subseteq X \triangleq disk_a(r) \).

Proof: See proposition 1 and the definition of the Hausdorff distance. ■

At this point we adopt a slightly different convention on how we specify disk sizes because we really want to perform these manipulations on a computer with digital disks. Various approaches to digitization are discussed in [7]-[9]. We also desire to index a family of these disks based on a one to one correspondence with the integers. Our approach will be to quantize the Hausdorff metric, after first multiplying by a factor of two. The factor of two allows better quantization and corresponds to an easily interpretable quantity, the diameter.

In all discussions hereafter we are operating in the digital domain where every set has been sampled by \( S \), the discrete pixel set. Therefore define \( disk_{\text{num}}(X, Y) = \lceil 2\rho(X, Y) \rceil \), and \( d(X, Y) = \lceil 2\rho_M(X, Y) \rceil \) where the ceiling operation, \( \lceil \cdot \rceil \), rounds upward to the next largest integer (returns the smallest integer greater than or equal to \( x \)).

Proposition 4. \( d(X, Y) \) is a metric.

Proof:

1. If \( X = Y \) then \( \rho_M(X, Y) = 0 \) since it is a metric. So \( 2\rho_M(X, Y) = 0 \) thus \( d(X, Y) = 0 \) since \( \lceil 0 \rceil = 0 \).

   If \( d(X, Y) = 0 \) then \( -1 < 2\rho_M(X, Y) \leq 0 \) by definition of the ceiling operator. So \( -\frac{1}{2} < \rho_M(X, Y) \leq 0 \). But \( \rho_M(X, Y) \geq 0 \) always since it is a metric. Thus \( \rho_M(X, Y) = 0 \) which implies that \( X = Y \).

2. \( d(X, Y) = \lceil 2\rho_M(X, Y) \rceil = \lceil 2\rho_M(X, Y) \rceil = d(Y, X) \).

3. \( \rho_M(X, Y) \leq \rho_M(X, Z) + \rho_M(Z, Y) \) implies \( d(X, Y) \leq d(X, Z) + d(Z, Y) \).

   But \( \lceil \rho_M(X, Z) + \rho_M(Z, Y) \rceil \leq \lceil \rho_M(X, Z) \rceil + \lceil \rho_M(Z, Y) \rceil \). ■

An alternative approach (not followed) would be to use the \( L^\infty \) norm in the definition of Hausdorff. In the \( L^\infty \) norm, the shape of balls is square.

We can restate proposition 2 and 3 in this convention.

Proposition 5 (For Sets in the Digital Domain)

1. If \( d_{\text{num}}(X, Y) = i \) then \( X \subseteq Y \triangleq disk(i) \) and \( X \not\subseteq \text{Y} \triangleq disk(i') \) for \( i' < i \).

2. \( d(X, Y) \leq i \) if and only if \( X \subseteq Y \triangleq disk(i) \) and \( Y \subseteq X \triangleq disk(i) \).

Proof: Proposition 2, 3, the definition of \( d(X, Y) \), and the definition of \( disk(i) \). ■

This proposition relates (quantized) Hausdorff distance to dilation with digital disks. The only situation not yet handled is when either \( X \) or \( Y \) is the empty set (in the context of image restoration, this occurs when a binary image is completely obliterated by noise). One palliative is to define the distance to be some suitably large number if either one of the inputs is the empty function. The number functions as a "badness" weight in this case.

3.1 Comparison of Square Error distance to Hausdorff distance

The Square Error distance and the Hausdorff distances both satisfy the three criteria for a distance relationship to be a metric. But there is no relationship between values of these metrics in any particular case. The Square Error measure in the binary image context is just a count of the number of excess pixels summed with the number of missing pixels. The Hausdorff distance on the other hand measures a maximum error between two sets.

The appropriate choice of distance measure depends on the application as the following quotation from [13], p 221, suggests:

"In principle, there is a great latitude available in choosing the criterion of optimality. In a practical sense, however, it should be a meaningful measure of "goodness" for the problem at hand and should correspond to equations that are mathematically tractable."

We chose the Hausdorff metric as our error measure because it has a natural interpretation in terms of set coverings and dilation. These qualities lend themselves to simplified design in the morphological context. The Square Error metric on the other hand has no interpretation in terms of set coverings and does not easily fit into either the dilation or erosion set equation.
4 Single Stage Restoration and Appropriate Noise Models

The purpose of any model is to describe or represent reality. Noise models for images describe the noise or image corruption process. We propose three different descriptions of the image corruption process, and we will assume the parameters associated with each noise model are known. If this is not the case, they may have to be estimated from training data. We will use the terminology of Figure 2; random sets will be denoted by underscoring them.

The input image $A$ is corrupted into noise image $C$ by the noise process. Restoration is assumed to be a single dilation with structuring element $L$ which is under our control. Image $A$ is random, however it will be a constrained randomness. For our purposes, $A$ will be some unknown translate of a known set $A$ (random location) or possibly $A$ will be an unknown translate of one of a fixed number of underlying known sets (random location and random shape), each known to occur with a specified probability. For example, the underlying known set could be a line where the length is random according to some distribution. Given the form assumed for the restoration, it is appropriate to describe forms of noise (noise models) that can reasonably be restored.

4.1 Erosion

For the erosion noise model, the input image $A$ is eroded with one of $n$ possible structuring elements, $K_j$, $j = 1, \ldots, n$. $K_j$ is chosen according to probability $p_j$ with $\sum_{j=1}^{n} p_j = 1$. Then $C = A \ominus K_j$. We impose no other restrictions on the $K_j$'s, such as a requirement that they contain the origin.

Figure 3 shows some X-ray images of a human heart. The erosion noise model is applicable as a description of change between images. In this model noise can be viewed as "eating away" at the edges. One might consider this model as a representation of errors due to improperly thresholding a grey scale image to binary.
4.2 Independent Pixel Subtractive Noise

The independent pixel subtractive noise model is a form of subtractive noise different from erosion. Only pixels that had value one in the original binary image, are potentially changed. Each pixel changes independently from one to a zero with probability 1 - \( p \). (The probability a pixel in \( A \) will survive in \( C \) is \( p \)). Figure 4 shows a portion of a simulated split beam sonar image. Unlike most sonar images, split-beam sonar tends to produce sparse returns, which makes this model appropriate.

4.3 Independent Pixel Subtractive Noise with Dilation

This model is similar to the previous model, but after the independent pixel subtractive noise the resulting (intermediate) image \( B \) is dilated with known structuring element \( J \) to produce the final damaged image \( C \).

See Figure 5.

5 Search Procedure

Our goal is to find a structuring element \( L \) to optimize the reconstruction of \( C \) by dilating with \( L \) so that,

\[
E\{d(A, C \oplus L)\}
\]

is minimized. \( L \) must be found by a combinatorial search since it is a set.

Conceptually the combinatorial search takes place as follows. First an initial starting set \( I \) is formed with \( I \) "large enough" so that no points outside of \( I \) need to be considered. Next all \#I descendant nodes are formed from \( I \), each with one pixel less than \( I \). For each of these descendant nodes, the expectation is calculated. The node with the smallest expectation is selected for further expansion, and so on until the search returns the subset of \( I \) that is best in sense of minimizing the expectation.

For the problem at hand, the search procedure undergoes some modifications due to the availability of underestimates of the utility function. We shall see that the underestimates dramatically speed up the (now trivial) search process as well as tell when to terminate the search.
The organization of this section will be to first provide all definitions necessary to understand the search procedure, secondly give an exact statement of the search procedure, then provide a justification of the search procedure in detail.

5.1 Definitions

For the following definitions, let \( N \) be given.

**Definition** \( U_i(L) = p(C \oplus L \subseteq A \Rightarrow disk(i)) \) for \( i = 1 \ldots N \).

\( U_i(L) \) is the probability the restored set is "under" or covered by the (enlarged) ideal set \( A \). The index \( i \) determines the size of the disk that \( A \) is dilated with.

**Definition** \( O_i(L) = p(A \subseteq (C \oplus L) \supseteq disk(i)) \) for \( i = 1 \ldots N \).

\( O_i(L) \) is the probability that the (enlarged) restored set is "over" or covers the ideal set \( A \).

**Definition** \( H_i(L) = p(C \subseteq L \subseteq A \Rightarrow disk(i) \text{ and } A \subseteq (C \oplus L) \supseteq disk(i)) \) for \( i = 1 \ldots N \).

\( H_i(L) \) is the probability that the restored set is under the (enlarged) ideal set \( A \) and that the (enlarged) restored set is over the ideal set \( A \). Note that \( H_i(L) = q \) is the probability that the quantized Hausdorff distance is less than or equal to \( q \).

**Definition** \( E(L) = p(A \cap (C \oplus L) = 0) \) for \( i = 1 \ldots N \).

\( E(L) \) is the probability that the intersection of the ideal set and the restored set is empty.

**Definition** \( \epsilon(L) = \sum_{i=0}^{N-1} H_i(L) \).

\( -\epsilon(L) \) is an error function directly related to the expectation, \( E\{d(A, C \oplus L)\} \).

**Definition** \( B_{\epsilon}(L) = \sum_{i=0}^{N-1} O_i(L) \).

\( B_{\epsilon}(L) \) is an upper bound on \( \epsilon(L) \) based on calculating \( O_i(L) \), \( i = 0 \ldots N - 1 \).

**Definition** \( S(l) = \sum_{i=0}^{N-1} U_i(\{l\}) \).

\( S(l) \) calculates a "score" (larger is better) for every pixel \( l \in L \) that measures how likely it is that a given translate of \( C \) is under the (enlarged) ideal set.

**Definition** \( B_{U}(L) = \min_{l \in L} S(l) \).

\( B_{U}(L) \) is an upper bound on \( \epsilon(L) \) based on calculating \( U_i(\{l\}) \) for every \( l \in L \).

5.2 Optimal Search Procedure

1. Determine the constant \( N \).
   
   (a) \( M=0 \).
   
   (b) if \( H_M(\{0\}) = 1.0 \) then goto done
      
      else \( M = M + 1 \)
   
   (c) repeat step b.
   
   (d) done: \( N = 2 \ast M \).

2. Determine initial starting point, \( I = disk(M) \).
3. Determine reduced starting point, $L$.
   (a) $L = \emptyset$
   (b) For each $i \in I$
   (c) if $E(i) < 1$ then $L = L \cup \{i\}$.
4. Calculate for every pixel $l \in L$ its score function $S(l)$.
5. Remove those pixels which have a score lower than $\epsilon(\{0\})$.
6. Sort the pixels in $L$ by score function in ascending order (put pixels with smaller score functions at top of list).
7. Remove pixels from $L$ starting at top of list. For each smaller $L$, calculate $\epsilon(L)$, record the maximum. Stop if $\epsilon(L) = B\epsilon(L)$.

5.3 Utility Function
In this section we derive how to compute $E\{d(A, C \oplus L)\}$, which is the utility function that must be minimized. We begin by recognizing that $d = d(A, C \oplus L)$ is a discrete random variable and has a density function that will depend on the noise model and choice of $L$. From the definition of expected value,

$$E\{d(A, C \oplus L)\} = 0p(d = 0) + 1p(d = 1) + 2p(d = 2) + \cdots + Np(d = N).$$

The sum is taken over the range of potential disk diameter sizes; the upper limit is an integer above which the density function is zero. Determination of this integer, $N$, is explained later.

Since the $p$'s form a probability mass function for disk diameter, they sum to one, i.e.

$$\sum_{i=0}^{N} p(d = i) = 1.$$

Thus the probability of the last diameter bin can be rewritten as

$$p(d = N) = 1 - (p(d = 0) + p(d = 1) + \cdots + p(d = N - 1)).$$

So

$$E\{d(A, C \oplus L)\} = 1p(d = 1) + 2p(d = 2) + \cdots +$$

$$N[1 - (p(d = 0) + p(d = 1) + p(d = 2) + \cdots + p(d = N - 1))]$$

$$= N - [Np(d = 0) + (N - 1)p(d = 1) + \cdots + p(d = N - 1)]$$

$$= N - [p(d \leq 0) + p(d \leq 1) + \cdots + p(d \leq N - 1)]$$

The last equality is just a restatement in terms that are part of the cumulative distribution function of the random variable of distance, $d$. Now we recognize via proposition 5 and the definition of $H$ that

$$p(d \leq i) = p(A \subseteq C \oplus L \ominus disk(i) \text{ and } C \ominus L \subseteq A \ominus disk(i)) = H_{i}(L)$$

Recalling that $\epsilon(L)$ was the sum over the $H_{i}(L)$’s leads to

$$E\{d(A, C \oplus L)\} = N - \epsilon(L)$$

Minimizing the expected value corresponds to maximizing $\epsilon(L)$. In order to perform a search to find the optimal $L$, it is necessary to be able to calculate $\epsilon(L)$, which in turn is based on calculating $H_{i}(L)$.

Now we show how to calculate the variance in a related calculation.

$$Var\{d(A, C \oplus L)\} = E\{(d(A, C \oplus L)^{2}\}} - E\{d(A, C \oplus L)\}^{2}$$

and
\[
E\{(d(A, C \oplus L))^2\} = 1^2p(d = 1) + 2^2p(d = 2) + \cdots + N^2p(d = N)
\]
\[
= 1^2[p(d \leq 1) - p(d \leq 0)] + 2^2[p(d \leq 2) - p(d \leq 1)] + \cdots + N^2[p(d \leq N) - p(d \leq N - 1)]
\]
\[
= p(d \leq 0)(-1^2) + p(d \leq 1)(1^2 - 2^2) + \cdots + p(d \leq N - 1)((N - 1)^2 - N^2) + p(d \leq N)(N^2)
\]
\[
= p(d \leq N)N^2 - \sum_{i=0}^{N-1} p(d \leq i)(i^2 - (i + 1)^2)
\]
\[
= p(d \leq N)N^2 - \sum_{i=0}^{N-1} p(d \leq i)(2i + 1)
\]

### 5.4 Bounds on the Utility Function

We will calculate two different kinds of upper bounds on \( \epsilon(L) \). (An upper bound on \( \epsilon(L) \) corresponds to a lower bound on \( E\{d(A, C \oplus L)\} = N - \epsilon(L) \). Whichever \( L \) is found by the search procedure can expect to have no better performance than the least upper bound (greatest lower bound) on \( \epsilon(L) \) \( (E\{d(A, C \oplus L)\}) \). The bounds are useful first because they provide underestimates of the search distance (necessary for optimal search) and secondly signal termination of the search if no better structuring element can be found.

The first upper bound on \( \epsilon(L) \) is \( B_U(L) \) which is a bound based on the restored set being underneath the ideal set.

The bound follows in the following manner:

\[
p(\mathcal{C} \ominus L \subseteq A \ominus \text{disk}(i) \text{ and } A \subseteq (\mathcal{C} \ominus L) \ominus \text{disk}(i)) \leq p(\mathcal{C} \ominus L \subseteq A \ominus \text{disk}(i))
\]

because \( p(E \cap F) \leq p(E) \) always for any sets of events \( E \) and \( F \) (See [12], p.26). Also by the same reasoning

\[
p(\mathcal{C} \ominus L \subseteq A \ominus \text{disk}(i)) \leq p(\mathcal{C} \ominus K \subseteq A \ominus \text{disk}(i)) \text{ for } K \subseteq L
\]

In terms of the previous functions we defined,

\[
H_l(L) \leq U_l(L)
\]

and

\[
U_l(L) \leq U_l(K) \text{ for } K \subseteq L
\]

It follows that

\[
U_l(L) \leq U_l(\{l\}) \text{ for all } l \in L
\]

since \( \{l\} \subseteq L \) always holds. Since the relationship holds for every \( l \in L \) it holds for the smallest of them or

\[
U_l(L) \leq \min_{l \in L} U_l(\{l\})
\]

Combining,

\[
\epsilon(L) = H_0(L) + H_1(L) + \cdots + H_{N-1}(L)
\]
\[
\leq U_0(L) + U_1(L) + \cdots + U_{N-1}(L)
\]
\[
\leq \min_{l \in L} U_l(\{l\}) + \min_{l \in L} U_l(\{l\}) + \cdots + \min_{l \in L} U_{N-1}(\{l\})
\]
\[
\leq \min_{l \in L} S(l)
\]
\[
\leq B_U(L)
\]

This bound is important for several reasons. First, the score function \( S(l) \) can be precalculated for each pixel \( l \in L \) that is a member of whichever structuring element \( L \) is under consideration, it need only be done once then it can be stored in a table.
Secondly, it suggests a search order if the search is organized so that the largest possible structuring element is considered first, then various subsets of it.

The second upper bound on \( \epsilon(L) \), \( B_O(L) \), is based on the restored set being over or covering the ideal set. Since

\[
p(C \ominus L \subseteq A \ominus disk(i) \text{ and } A \subseteq (C \ominus L) \ominus disk(i) \leq p(A \subseteq C \ominus L \ominus disk(i))
\]

and

\[
p(A \subseteq C \ominus K \ominus disk(i)) \leq p(A \subseteq C \ominus L \ominus disk(i)) \text{ for } K \subseteq L
\]

inequalities similar to the previous ones also hold.

\[H_i(L) \leq O_i(L)\]

and

\[O_i(K) \leq O_i(L) \text{ for } K \subseteq L\]

So

\[
\epsilon(L) = H_0(L) + H_1(L) + \cdots + H_{N-1}(L) \\
\leq O_0(L) + O_1(L) + \cdots + O_{N-1}(L) \\
\leq B_O(L)
\]

This second bound provides a useful stopping condition on the search.

### 5.5 Search Starting Point

In this section we discuss how to find the initial shape to start the combinatorial search on. Determination of \( N \) will also be explained in this section.

We shall show that a disk can be found such that the search need only be restricted to its subsets. If points outside the disk are considered then the results are actually worse than doing no filtering at all. No filtering corresponds to the case when \( L = \{0\} \) since \( C \ominus \{0\} = C \). We first temporarily find \( N \) by repeatedly calculating \( H_i(\{0\}) \), \( i = 0, 1, \ldots \) for ever increasing values of \( i \) and set \( N \) to be the smallest integer \( M \) such that \( H_M(\{0\}) = 1 \). This is an acceptable choice for \( N \) when \( L = \{0\} \) because if the distribution function is equal to one (and \( H_i(L) \) is the distribution for \( d \) evaluated at point \( i \) at value \( N \), then the density function is zero for all values greater than \( N \).

So

\[E(d(A, C \ominus \{0\})) = M - \epsilon(\{0\}) \leq M\]

holds.

We now state a proposition which relates translation with points from outside \( disk(M) \) to subset containment.

**Proposition 6** If \( C \subseteq A \ominus disk(M) \) and \( A \subseteq C \oplus disk(M) \) then either \( C_1 \subseteq A \oplus disk(M) \) or \( A \subseteq C_1 \oplus disk(M) \) for all \( t \notin disk(M) \).

**Proof:** Suppose not. Suppose there exists a \( t \notin disk(M) \) and \( C_1 \subseteq A \ominus disk(M) \) and \( A \subseteq C_1 \ominus disk(M) \). Let \( c \in C \). Then \( c + t = a + d \) for some \( a \in A \) and for some \( d \in disk(M) \). Let \( a' \in A \). Then \( a' = c + t + d' \) for some \( c' \in C \), \( d' \in D \). Rearranging produces \( t = a - c + d \) and \( t = a' - c' \). Thus \( 2t \leq\| a - c \| + \| a' - c' \| + \| d - d' \| \). By the conditions of the hypothesis, proposition 5, and definition of \( disk(M) \) we have \( \| a - c \| \leq M/2 \) and \( \| a' - c' \| \leq M/2 \). Since \( disk(M) = disk(M) \) then \( c + t = a + (d - c) \) for some \( a \in A \) and for some \( d \in disk(M) = disk(M) \). Or \( a = c + t + d \). So the triple of \( a = a' \), \( c = c' \) and \( d = d' \) satisfies both equations. For this case \( d = d' \) implies \( \| d - d' \| = 0 \) so \( \| 2t \| \leq M/2 + M/2 = M \) which implies \( \| t \| \leq M/2 \). Since \( t \notin disk(M) \) then \( \| t \| > M/2 \) which is a contradiction. 

Now we are in a position to make a statement about any \( L \) that contains points outside \( disk(M) \). Since \( H_M(\{0\}) = 1 \) implies that the hypothesis of proposition 6 is the certain event, it follows that \( C \ominus L \subseteq A \ominus disk(M) \) or \( A \subseteq C \ominus L \ominus disk(M) \) occurs. This implies that for \( i \leq M \),

\[p(C \ominus L \subseteq A \ominus disk(i) \text{ and } A \subseteq C \ominus L \ominus disk(i) = p(d(A, C \ominus L) = i) = 0.0\]
Figure 6: Illustrating removing points in $K$ from $L$. No point in any $C, C \subseteq A$, when dilated by any pixel in $K$, will cover any of $A$.

Thus

$$E\{d(A, Q \oplus L)\} = 0 + 1p(d = 1) + \ldots + Mp(d = M) + (M + 1)p(d = M + 1) + \ldots$$

$$= 0 + 0 + \ldots + 0 + (M + 1)p(d = M + 1) + \ldots$$

$$\geq M + 1$$

so it is always the case that

$$E\{d(A, Q \oplus \{0\})\} \leq M < M + 1 \leq E\{d(A, Q \oplus L)\}$$

when $L$ contains points outside $\text{disk}(M)$. So we shall restrict our search to subsets of $\text{disk}(M)$. The next proposition states that if $L$ is restricted to being a subset of a disk, then we can guarantee a subset relation.

**Proposition 7** If $C \subseteq A \oplus \text{disk}(M)$ and $A \subseteq C \oplus \text{disk}(M)$ then $C \oplus L \subseteq A \oplus \text{disk}(2M)$ and $A \subseteq C \oplus L \oplus \text{disk}(2M)$ for $L \subseteq \text{disk}(M)$.

**Proof:** Since $C \subseteq A \oplus \text{disk}(M)$ then $C \supseteq \text{disk}(M) \subseteq A \oplus \text{disk}(M) \oplus \text{disk}(M)$ but $L \subseteq \text{disk}(M)$ so that $C \supseteq L \subseteq C \oplus \text{disk}(M) \subseteq A \oplus \text{disk}(M) \oplus \text{disk}(M)$. Now we show that $\text{disk}(M) \oplus \text{disk}(M) \subseteq \text{disk}(2M)$. Let $x \in \text{disk}(M) \oplus \text{disk}(M)$. Then $z = a + b$ for some $a, b \in \text{disk}(M)$. $a, b \in \text{disk}(M) = S \cap \text{disk}_a(M/2)$ imply $a, b \in S$ and $a, b \in \text{disk}_a(M/2)$. So $a + b \in S$ and $a + b \in \text{disk}_a(M/2)$. Therefore $a + b \in S \cap \text{disk}_a(M) = \text{disk}(2M)$.

To show $C \subseteq A \oplus L \oplus \text{disk}(2M)$ we begin by showing that $0 \in L \oplus \text{disk}(M)$. Let $l \in L$. Then $l \in \text{disk}(M)$ since $L \subseteq \text{disk}(M)$. But $\text{disk}(M) = \text{disk}(M)$ so that $-l \in \text{disk}(M)$. Thus $0 \in L \oplus \text{disk}(M)$ since $0 = -l + l$. Hence $A \subseteq (C \oplus \text{disk}(M)) \oplus (L \oplus \text{disk}(M))$. Rearranging by the distributive and associative properties of dilation, and using the results of the first part of the proof, we arrive at the desired result.

Thus take $N = 2M$ in the general case. If it is known from the noise model that $C \subseteq A$ always holds then $N = M$ is a smaller $N$ that will suffice.

Now we consider how the search starting point can be reduced further from a disk of size $M$. Suppose there were some points $K$, such that $A \cap (C \oplus K) = \phi$. Then none of the points in $L \cap K$ need be considered further in the search. This is because the Hausdorff metric is concerned with covering and set containment. If some points are in both $L$ and $K$ they can obviously be removed because they don't help with the covering. See Figure 6.

In what follows we show that $L - K$ is at least as good a starting starting point as $L$. Consider the following proposition.

**Proposition 8** If $A \subseteq C \oplus L$ and $A \cap (C \oplus K) = \phi$ then $A \subseteq C \oplus (L - K)$.

**Proof:** Since $A \subseteq C \oplus L$ and $A \cap (C \oplus K) = \phi$ then $A \subseteq (C \oplus L) - (C \oplus K)$. Let $x \in (C \oplus L) - (C \oplus K)$. Then $x \in C \oplus L$ and $x \notin C \oplus K$. Thus $x = c + l$ for some $c \in C$ and for some $l \in L$ and $x \notin c + k$ for every $c \in C$ and for every $k \in K$. So $x = c + j$ for some $c \in C$ and for some $j \in L - K$. So $x \in C \oplus (L - K)$. ■
The proposition states that a set containment relationship will still hold if the translates in \( L \) that don't cover anything are removed, provided that \( L \) was able to cover all of \( A \) in the first place. Thus we use the proposition to make the assertion that \( \text{H}_1(L - K) \geq \text{H}_1(L) \) where \( L = \text{disk}(N) \) and \( A \cap (C \oplus K) = \emptyset \) is certain to hold by only taking those points \( K \) such that \( E(K) = p(A \cap (C \oplus K) = \emptyset) = 1 \). The argument can easily be extended to \( \text{H}_1(L - K) \geq \text{H}_1(L) \) and thus \( \epsilon(L - K) \geq \epsilon(L) \).

The test if a point in \( L \) can be eliminated is done by checking if \( A \cap (C \oplus \{l\} = \emptyset) \) holds for every \( l \in L \). This test is located in the algorithm in step 2, where \( E(\{l\}) = p(A \cap (C \oplus \{l\})) \) computed for every pixel. If this probability is 1.0 then the event of the intersection being empty is the certain event and that point need not be included.

One further test is done at this point. If the score function function for a pixel is less than \( \epsilon(\{0\}) \) then the pixel must be removed. For in this case

\[
\epsilon(L) \leq B_U(L) = \min_{i \in L} S(I) \leq \epsilon(\{0\})
\]

and the structuring element cannot possibly be optimal. The pixels that have a low score function guarantee that the results will be worse (should they be included) than doing no filtering at all (equivalent to filtering with the origin).

### 5.6 Search Stopping Point

Suppose we are at some point in the search, we are evaluating structuring element \( L \). Given that we know how well \( L \) does, we are asking the question, is there any way to eliminate some of the descendants \( K \) of \( L, K \subseteq L \). The answer is yes, under some conditions, based on the value of the second upper bound, \( B_0(L) \). Earlier we argued that \( O_i(K) \leq O_i(L) \) for \( K \subseteq L \). But \( B_0(L) \) is just the sum of the \( O_i(L) \)'s therefore, \( B_0(K) \leq B_0(L) \) for \( K \subseteq L \).

Then when \( \epsilon(L) = B_0(L) \), we have \( \epsilon(K) \leq B_0(K) \leq B_0(L) = \epsilon(L) \) holds. So no matter which subset \( K \) of \( L \) is chosen it can do no better than the current structuring element \( L \). Thus no further checking of the descendants of \( L \) need be done.

### 5.7 Optimality of Search Procedure

In this section we justify the (trivial) search procedure described as the optimal algorithm. In fact no real searching is done. In order to justify this procedure we need to show that we are always choosing the path that has the smallest estimated total distance.

In terms of combinatorial search \( -\epsilon(L) \) is the objective function we want to minimize. It is a well known result of Artificial Intelligence (see [10] and [11]) that if an accurate underestimate of the distance remaining is available, and that if the node with the smallest estimated total distance is always expanded, then the search procedure will be optimal. Since \( \epsilon(L) \leq B_0(L) \) then \( -\epsilon(L) \geq -B_0(L) \) which shows that \( -B_0(L) = -\min_{i \in L} S(I) \) is an underestimate of \( \epsilon(L) \).

Because of the form of \( B_0(L) \), choosing the pixel with the smallest score function is always the correct decision, since only with this choice will the underestimate decrease. This is true because \( B_0(L) = \min_{i \in L} S(I) \); its value is given by the smallest score of all the pixels contained in \( L \). If the wrong pixel is removed the underestimate is guaranteed to remain the same, not decrease. Thus optimal search, which consists of always expanding the node with the smallest underestimate of total distance, requires removing pixels with the smallest scores first. The search process has been trivialized to removing the pixels in a predefined order.

As an example of this principle consider an \( L \) which has three pixels in it with scores \{0.2, 0.3, 0.4\}. The underestimate of distance is \(-0.2 = -\min\{0.2, 0.3, 0.4\}\). If the pixel with score of 0.2 is removed the underestimate becomes \(-0.3 = -\min\{0.3, 0.4\}\) (smaller than \(-0.2\)), otherwise it remains at \(-0.2\).

This result can be understood intuitively in terms of

\[
\text{H}_1(L) = p(C \oplus L \subseteq A \oplus \text{disk}(i) \text{ and } A \subseteq (C \oplus L) \oplus \text{disk}(i)).
\]

Essentially, the algorithm removes those pixels in \( L \) which are more likely to "stick out" when dilated by \( C \) and not be underneath \( A \oplus \text{disk}(i) \).

### 6 Conclusion

We have described three different noise models in which it makes sense to restore the image by dilation (maximum filtering). Given the noise model, we describe a search procedure that finds the optimum structuring element in the sense of minimizing the mean Hausdorff error. We described both a search starting point and a search stopping condition.
7 Acknowledgements

Mr. Rystrom gratefully acknowledges the support he received from the Applied Physics Laboratory of the University of Washington.

References


