Chapter 1

Introduction to Mathematical Morphology

In this chapter we endeavor to introduce in a concise way the main aspects of Mathematical Morphology, as well as what constitutes its field. This question is difficult, not so much as a technical matter but as a question of starting point. Historically, mathematical morphology began as a technique to study random sets with applications to the mining industry. It was rapidly extended to work with two-dimensional (2D) images in a deterministic framework first with binary images, then gray-level and later to color and multispectral data and in dimensions > 2. The framework of mathematical morphology encompasses many various mathematical disciplines from set theory including lattice theory, random sets, probabilities, measure theory, topology, discrete and continuous geometry, as well as algorithmic considerations and finally applications.

The main principle of morphological analysis is to extract knowledge from the response of various transformations which are generally nonlinear.

One difficulty in the way mathematical morphology has been developed and expanded [MAT 75, SER 82, SER 88c] (see also [HEI 94a, SCH 94, SOI 03a]) is that its general properties do not fall within the general topics taught at school and universities (with the exception of relatively advanced graduate-level courses). *Classical* mathematics define a function as an operator associating a single point in a domain with a single value. *A contrario*, in morphology we associate whole sets with other whole sets. The consequences of this are important. For instance, if a point generally has zero measure, this is not generally the case for sets. Consequently, while a probability of the presence of a point may be zero, this is not the case for a set.

Chapter written by Laurent NAJMAN and Hugues TALBOT.

Mathematical Morphology: From Theory to Applications Edited by Laurent Najman and Hugues Talbot

In addition, we can compare morphology to other image processing disciplines. For instance, linear operator theory assumes that images are merely a multidimensional signal. We also assume that signals combine themselves additively. The main mathematical structure is the vector space and basic operators are those that preserve this structure and commute with basic rules (in this case, addition and multiplication by a constant). From this point deriving convolution operators is natural; hence it is also natural to study Fourier or wavelet transforms. It is also natural to study decomposition by projections on basis vectors. This way is of course extremely productive and fruitful, but it is not the complete story.

Indeed, very often a 2D image is not only a signal but corresponds to a projection of a larger 3D 'reality' onto a sensor via an optical system of some kind. Two objects that overlap each other due to the projections do not add their feature but, on the contrary, create occlusions. The addition is not the most natural operator in this case. It makes more sense to think in terms of overlapping objects and therefore, in terms of sets, their union, intersections and so on. With morphology, we characterize what is seen via geometrical transforms, taking into account shapes, connectivity, orientation, size, etc. The mathematical structure that is most adapted to this context is not the vector space, but the generalization of set theory to complete lattices [BIR 95].

1.1. First steps with mathematical morphology: dilations and erosions

In order to be able to define mathematical morphology operators, we need to introduce the abstract notion of complete lattice. We shall then be able to 'perform' morphology on any instance of such a lattice.

1.1.1. *The notion of complete lattice*

A *lattice* [BIR 95] $(E, \leq)$ is a set E (the space) endowed with an *ordering* relationship $\leq$ which is reflexive ($\forall x \in E, x \leq x$), anti-symmetric ($x \leq y$ and $y \leq x \Rightarrow x = y$) and transitive ($x \leq y$ and $y \leq z \Rightarrow x \leq z$). This ordering is such that for all x and y, we can define both a larger element $x \vee y$ and a smaller element $x \wedge y$. Such a lattice is said to be *complete* if any subset P of E has a *supremum* $\bigvee P$ and an *infimum* $\bigwedge P$ that both belong to E. The supremum is formally the smallest of all elements of E that are greater than all the elements of P. Conversely, the infimum is the largest element of E that is smaller than all the elements of P. In a lattice, *supremum* and *infimum* play symmetric roles. In particular, if we consider the lattice $\mathcal{P}[E]$ constituted by the collection of all the subsets of set E, two operators ψ and ψ^* are *dual* if, for all X, $\psi(X^c = [\psi^*(X)])$ where $X^c = E \setminus X$ is the *complement* of X in E.

1.1.2. *Examples of lattices*

Figure 1.1 is an example of a lattice. This instance is simple but informative, as it corresponds to the lattice of primary additive colors (red, green and blue). Each element of the lattice is a binary 3-vector, where 0 represents the absence of a primary color and 1 its presence. The color black is represented by $[0, 0, 0]$ and white by $[1, 1, 1]$. Pure red is $[1, 0, 0]$, pure green is $[0, 1, 0]$, and so on. Magenta is represented by $[1, 0, 1]$. In this lattice, there does not exist a way to directly compare pure green and pure blue or magenta and yellow: the order is not *total*. However, white is greater (brighter) than all colors and black is smaller (darker). Whatever subset of colors is chosen, it is always possible to define a *supremum* by selecting the maximal individual component among the colors of the set (e.g. the supremum of $[1, 0, 0]$ and $[0, 0, 1]$ is $[1, 0, 1]$). This supremum may not be in the subset, but it belongs to the original lattice. Similarly, the *infimum* is defined by taking the minimal individual component.

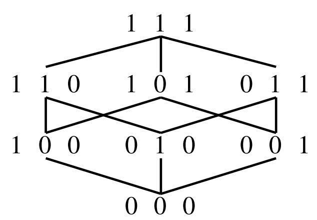


Figure 1.1. *An example of a lattice: the lattice of additive primary colors*

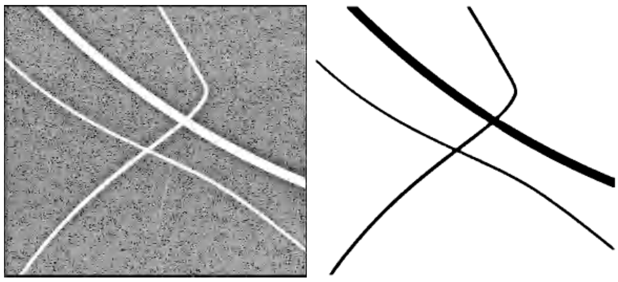
Another example of a lattice is the set of real numbers $\mathbb{R}$ endowed with the usual order relation. This lattice is not complete since, for instance, the subset of integer numbers has $+\infty$ as supremum but $+\infty$ is not part of $\mathbb{R}$. In contrast, $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$ is a complete lattice. Through these examples, we can see that the notion of complete lattice is not fundamentally difficult.

1.1.2.1. *Lattice and order*

Many morphological operators preserve the ordering structure. We call such operators Φ *increasing* and express it by $\forall x, y \in (E, \leq), x \leq y \Rightarrow \Phi(x) \leq \Phi(y)$. Others will transform input lattice elements into larger or smaller elements. If we have an operator Ψ which is such that $\forall x \in (E, \leq), x \leq \Psi(x)$ then the operator is called *extensive*: it will enlarge elements. Conversely, if $\Psi(x) \leq x$, then the operator is *anti-extensive*: it will shrink them. The simplest operators we can introduce on a lattice are those that commute with the *supremum* or the *infimum*. Respectively, these operators are called abstract *dilation* and *erosion*. Under various conditions such operators can combine some of these properties, as we will see shortly.

While these definitions are straightforward and relatively easily understood after some period of familiarization, there is a legitimate question as to why morphologists like to propose such abstract concepts. In order to answer this question, it is useful to think one level deeper and come back to the definition of an image.

Let us consider Figure 1.2a, which is a simple gray-level image. The content of this image may be technically interesting – it consists of glass fibers observed in an electron microscope – but it has no bearing here. We consider this image as a *function* $F : E \rightarrow T$, where E is the set of image points and T the set of possible values of F. In this case F is perhaps a set of discrete gray levels, possibly coded over 8 bits or 256 gray levels. The space T might instead be a subset of $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$. Conversely, the space E can be seen as continuous (for instance $E = \mathbb{R}^n$) or discrete (for instance $E = \mathbb{Z}^n$ or a suitable subset). We will denote the set of functions from E to T by T^E.



(a) (b)

Figure 1.2. *(a) A gray-level image and (b) a binary image obtained by thresholding (a)*

Depending on our application, it might be useful to consider one or the other of these definitions. How can we define operators that are in some way 'generic' and which will work irrespective of the precise definition of E and T? A benefit of using the lattice framework is precisely that we can define operators acting on images without specifying further the space of definition of these images. A more detailed description of lattices and algebraic morphology can be found in Chapter 2.

1.1.3. *Elementary operators*

It is possible to define morphological operators in many different ways. It is useful to consider the very simple case of *binary* images i.e. image that possess only two levels: strictly black with value 0 and strictly white with value 1. This framework is not the only one over which we can express morphology, but it has several advantages: it is relatively simple and intuitive but it is also sufficiently flexible for the further generalization of most operators to more complex lattices.

One of the simplest operators applicable to a gray-level image F is the *thresholding*. The threshold of F at *level* t is the set $X_t(F)$ defined by:

$$X_t(F) = \{p \in E | F(p) \leq t\}. \tag{1.1}$$

A threshold of image of Figure 1.2a is given in Figure 1.2b. The former is called a *gray-level* image and the latter a *binary* image. We can consider a binary image either as a subset of the continuous or discrete plane or, alternatively, as a function with values in $\{0, 1\}$. Once again, if we use the lattice framework this choice has little effect.

If we consider binary images as subsets of E the corresponding structure is the lattice $\mathcal{P}(E)$ endowed with the inclusion comparison operator, i.e. let X and Y be two subsets of E, then $X \leq Y \Leftrightarrow X \subseteq Y$. The *supremum* of a collection of sets $\{A, B, \ldots\}$ is given by the union operator $\bigvee\{A, B, \ldots\} = \bigcup\{A, B, \ldots\}$ and the *infimum* by the inclusion. This set lattice is very commonly used in practice, but it is not the only possible choice. For instance, if we seek to only work with convex sets it is much more appropriate to choose the convex set lattice with the usual inclusion operator as the *infimum*, but the convex hull of the union as the *supremum*.

1.1.3.1. *Structuring elements*

In the day-to-day practice of morphology, we often study binary or gray-level images using families of special sets B that are known *a priori* and can be adapted to our needs (in terms of size, orientation, etc.). These sets B are called *structuring elements*. They allow us to define the operators we evoked earlier (erosions and dilations) in a practical way. For instance, let X be a binary image i.e. a subset of E. The *translate* of X by $p \in E$ is the set $X_p = \{x + p | x \in X\}$. Here p defines a translation vector. The morphological *dilation* of X by B is given by:

$$\begin{aligned} \delta_B(X) = X \oplus B &= \bigcup_{b \in B} X_b \\ &= \bigcup_{x \in X} B_x \\ &= \{x + b | x \in X, b \in B\}. \end{aligned} \tag{1.2}$$

The resulting dilation is the union of the B_p such that p belongs to X: $\delta_B(X) = \bigcup\{B_p | p \in X\}$. As a consequence, the dilation of X by B 'enlarges' X, hence the name of the transform. In the formula, X and B play symmetric roles. Note also that when B is untranslated, (i.e. B_o), it is located somewhere relative to the origin of the

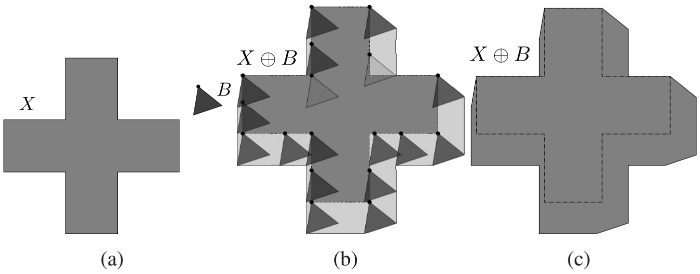


Figure 1.3. *The dilation of a cross by a triangle. The origin or the structuring element is one of the vertices of triangle B and is shown as a small black disk: (a) the original X (the light-gray cross) and B (the dark triangle); (b) the dilation taking place; and (c) the final result with the original set X overlaid*

coordinate system. We usually associate this point with B itself and call it the *origin* of the structuring element. When B is translated, so is its origin. An example of a dilation is shown in Figure 1.3.

The erosion of X by B is defined:

$$\begin{aligned} \varepsilon_B(X) = X \ominus B &= \bigcap_{b \in B} X_{-b} \\ &= \{p \in E | B_p \subseteq X\}. \end{aligned} \tag{1.3}$$

The erosion of X by B is the locus of the points p such that B_p is entirely included in X. An erosion 'shrinks' sets, hence its name. This is illustrated in Figure 1.4.

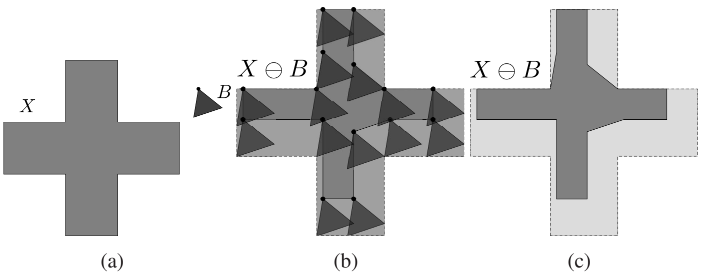


Figure 1.4. *The erosion of a cross by the same triangle structuring element as in Figure 1.3: (a) the original X (the light-gray cross) and B (the dark triangle); (b) the erosion taking place; and (c) the final result, overlaid within the original set X*

Erosion and dilation have opposite effects on images. More formally, they are *dual* by complementation: the dilation of a set X by B is the erosion of its complementary set X^c using the symmetric structuring element of B, denoted $\check{B}$. Let p, q be two points where $p \in B_q \Leftrightarrow q \in \check{B}_p$. This amounts to $\check{B} = \{-b | b \in B\}$:

$$(X \oplus B)^c = X^c \ominus \check{B}, \text{ and } (X \ominus B)^c = X^c \oplus \check{B}.$$

We illustrate this property in Figure 1.5 using the erosion as an example.

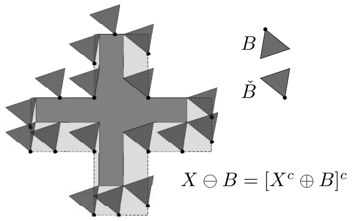


Figure 1.5. *The erosion of the cross of Figure 1.3, using the property that the dilation with the symmetric structuring element is the dual of this operation*

One way to extend the binary operators to the gray-level case is to take the hypograph $SG(F)$ of a function F:

$$SG(F) = \{(x, t) \in E \times T | t \leq F(x)\}.$$

Using this approach, dilating (respectively, eroding) a gray-level image is equivalent to dilating (respectively, eroding) each of its thresholds.

An equivalent approach consists of using the lattice of functions, using the order structure provided by the order on T. In particular, for two functions $F, G \in T^E$, we obtain:

$$F \leq G \Longleftrightarrow \forall x \in E, F(x) \leq G(x).$$

In this way, equations (1.2) and (1.3) translate in the following manner:

$$\delta_G(F)(x) = (F \oplus G)(x) = \sup_{y \in E}\{F(y) + G(x - y)\} \tag{1.4}$$

and

$$\varepsilon_G(F)(x) = (F \ominus G)(x) = \inf_{y \in E}\{F(y) - G(y - x)\}. \tag{1.5}$$

In these equations, function G is a *structuring function*. This function may be arbitrary, for instance sometimes parabolic functions are used in operations [BOO 96] such as the Euclidean distance transform [MEH 99].

1.1.3.2. *Flat structuring elements*

In practice, the most common structuring functions are the flat structuring elements (SEs). These are structuring functions which are identically equal to zero on a compact support K and that take the value $\bigwedge T$ elsewhere. In this case, equations (1.4) and (1.5) reduce to:

$$\varepsilon_K(F)(x) = \inf_{y \in E, y-x \in K} F(y) = \inf_{y \in K_x} F(y) \tag{1.6}$$

$$\delta_K(F)(x) = \sup_{y \in E, x-y \in K} F(y) = \sup_{y \in (\check{K})_x} F(y). \tag{1.7}$$

In this case, the alternative viewpoint is helpful: applying a flat morphological operator on a function F is equivalent to applying a morphological operator on all the thresholds $X_t(F)$ of F. For instance, in the case of the dilation by a flat structuring element K, this amounts to:

$$\delta_K(F) = \bigvee \{t \in T | p \in \delta_K(X_t(F))\}. \tag{1.8}$$

Figure 1.6 depicts an example of the dilation of a 1D signal by a structuring function. Figures 1.7a and b illustrate the 2D case.

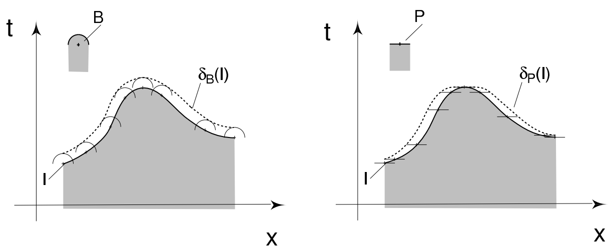
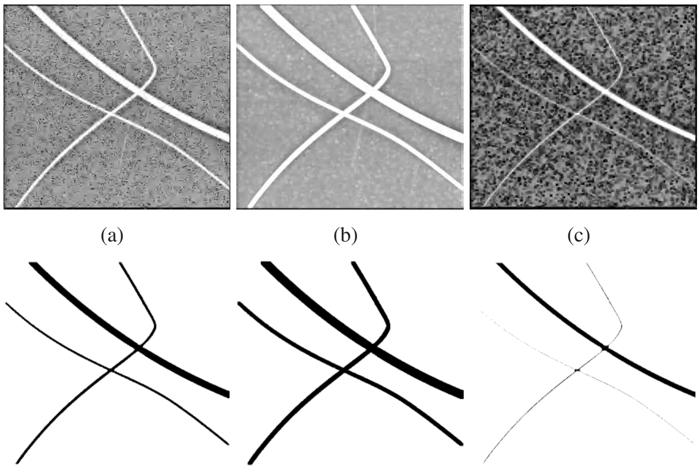


Figure 1.6. *Dilation of a signal (a 1D image) by a non-flat structuring element (a structuring function) and a flat structuring element. A dilation by a flat SE is the same as taking at every point the maximum of the function over the window defined by the symmetric SE*

We see here that morphological operators can readily be extended from the binary to the grayscale case. It is often easier to understand intuitively what an operator does in the binary case. It is also the case that, when working on gray-level images, it can be preferable to work in this mode for as long as possible and defer any thresholding. This way, the parameter of this operator can be chosen at a later stage when this decision might be easier.



(a) (b) (c) (d) (e) (f)

Figure 1.7. *Gray-level dilations and erosions of the images in Figure 1.2 by a symmetric* 5×5 *square structuring element: (a) gray-level original; (b) dilation; (c) erosion; (d) binary original; (e) dilation; and (f) erosion*

1.1.4. *Hit-or-miss transforms*

The erosion and dilation operators are useful by themselves (for instance to suppress some kinds of noise) but they are even more powerful when combined. For instance, we might want to consider some transforms that take into account both points that belong to a set and those that do not belong to it. We then need two structuring elements with a common origin. The first, denoted T_1, is applied to a set and the second, denoted T_2, is applied to its complementary set. We write:

$$X \circledast T = (X \ominus T_1) \cap (X^c \ominus T_2). \tag{1.9}$$

These operators are called *hit-or-miss transforms* or *HMT*. (Some authors also refer to this as the hit-*and*-miss transform. Both are acceptable and, as expressions, mean approximately the same thing. However, in the context of morphology, even if hit-*and*-miss is arguably better because we require one structuring element to fit in the foreground *and* the other to fit in the background, hit-or-miss is more usual.)

The operators are denoted $X \circledast T$, which is the locus of the points such that T_1 is entirely included in set X while T_2 is entirely included in the complement of X. These transforms can be used for pattern recognition, and many classical shape simplification procedures, such as skeletonization, use such techniques. Chapter 18 on document image processing describes some uses of HMTs. Chapter 15 presents an extension of HMT to grayscale images, and applies it in the context of medical image segmentation.

More generally, composing morphological operators such as dilations and erosions leads to morphological filtering.

1.2. Morphological filtering

In classical signal processing, the term ‘filter’ may mean any arbitrary processing procedure. In mathematical morphology, this terminology has a more precise meaning: a *morphological filter* is an operator that is both *increasing* and *idempotent*. We encountered the former insection 1.1.2.1: it means the order is preserved. The latter term means that if we repeat the operator, the result does not change after the first time. In other words, morphological filters respect the ordering and converge in one iteration.

In this context, the two most important operators are the *opening* and the *closing*. The opening is often denoted by γ and is a morphological filter (therefore increasing and idempotent) that is also *anti-extensive*. The closing is the complement of the opening; it is denoted most often by φ and is *extensive*. We also encountered extensivity and anti-extensivity in section 1.1.2.1. Respectively, they mean that the result is greater than the initial image, or smaller. In other words, openings make sets smaller and images darker, while closings make sets larger and images lighter. We shall now see examples of such operators.

1.2.1. *Openings and closings using structuring elements*

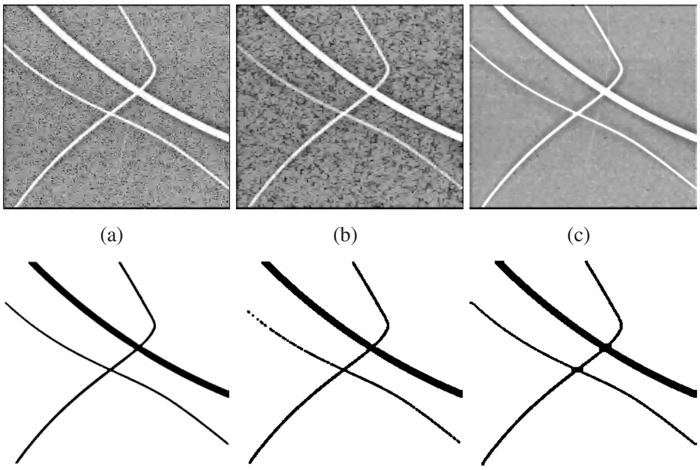
It is possible, as a particular case, to define morphological filters by composing dilations and erosions using structuring elements. For instance, the opening of set X by structuring element B may be defined:

$$\begin{aligned} \gamma_B(X) = X \circ B &= (X \ominus B) \oplus B \\ &= \bigcup \{B_p | p \in E \; et \; B_p \subseteq X\} . \end{aligned} \tag{1.10}$$

The closing of X by B is defined:

$$\varphi_B(X) = X \bullet B = (X \oplus B) \ominus B. \tag{1.11}$$

These formulae are similar in the gray-level case. In general terms, an opening will have a tendency to destroy the small, extruding and thin parts of objects; closing will tend to fill small holes and thin intruding parts of objects. This is illustrated in Figure 1.8.



(a) (b) (c)

(d) (e) (f)

Figure 1.8. *Openings and closings in the binary and gray-level cases, using the initial images from Figure 1.2 using a a* 5×5 *structuring element: (a) gray-level original; (b) opening; (c) closing; (d) binary original; (e) opening; and (f) closing*

These structuring element-based openings and closings are called *morphological* openings or closings. This is to distinguish them from the more general case of the operators that satisfy all the properties of the opening or closing, but are not necessarily the result of the composition of an erosion and a dilation.

Most importantly, we generally cannot combine any arbitrary erosion on the one hand and dilation on the other and call the result an opening or a closing. The two operators that compose a morphological opening or closing are called *adjunct* operators, by reference to the very specific duality that links the erosion and the dilation that are effectively used. This duality is generally not the same as taking the complement set and the symmetric structuring element. Much more detail about this is given in Chapter 2.

1.2.2. *Geodesy and reconstruction*

Let us now introduce the *conditional* dilation of a set X by a structuring element B, using a *reference* set R:

$$\delta^{(1)}_{R,B}(X) = (X \oplus B) \cap R. \tag{1.12}$$

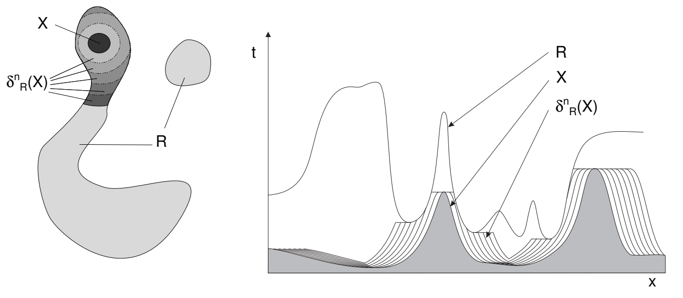
The result of this transform will always be included in the reference set R. Successive dilations are obtained by iteration of a (usually small) structuring element. Often the fundamental SE of the underlying grid is used (see section 1.2.2.2).

$$\delta_{R,B}^{(n+1)}(X) = (\delta_{R,B}^{(n)}(X) \oplus B) \cap R. \tag{1.13}$$

At convergence, we have

$$\delta_{R,B}^{\infty}(X) = \delta_{R,B}^{(n+1)}(X) = \delta_{R,B}^{(n)}(X). \tag{1.14}$$

This type of operator is illustrated in Figure 1.9.



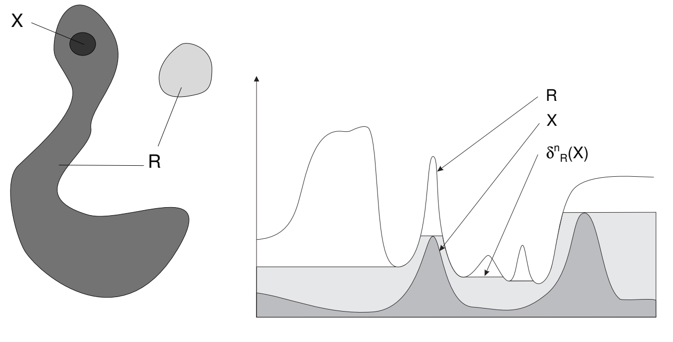
(a) (b)

Figure 1.9. *Geodesic dilation: (a) the binary case and (b) the gray-level case with a flat structuring element*

1.2.2.1. *Openings and closings by reconstruction*

One of the first applications of geodesic dilation is the *reconstruction* operator. We refer to the *reconstruction* of X under R by B as the set $\delta_{R,B}^{\infty}(X)$, i.e. what we obtain by iterating the geodesic dilation operator to infinity or equivalently to idempotence. Starting from 'markers' that designate the parts of an image we would like to *retain* in some way, a geodesic reconstruction allows us to regain the original shape of those parts even although they might have been damaged in order to obtain the markers. In gray level, a reconstruction operator will reconstruct the edges of the objects of interest. We illustrate this concept in Figure 1.10. For a given fixed set of markers, a geodesic reconstruction by dilation has all the properties of an opening.

As the name implies, the reconstruction operator is able to rebuild the shape of objects after they have been altered due to some other filtering operation. This



(a) (b)

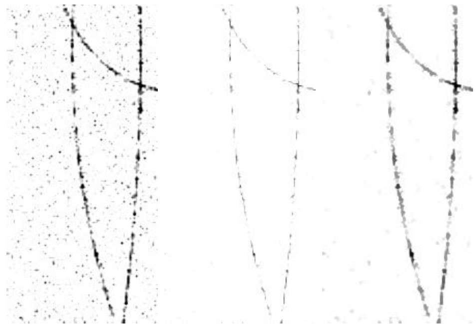
Figure 1.10. *The reconstruction operator: (a) the shape of the initial sets and (b) the 1D gray-level case*

operator is illustrated in Figure 1.11. The composition of an erosion followed by a reconstruction by dilations is a simple example of an algebraic opening, i.e. an opening which is not the composition of a single erosion followed by a single dilation. However, this kind of opening possesses all the properties of the opening. It is also a connected filter. Chapter 8 provides more information on this topic.

(a) (b)

Figure 1.11. *(a) Opening by reconstruction of the map of Australia, consisting of an erosion followed by reconstruction. Note that the initial erosion deletes the island of Tasmania such that (b) the reconstruction cannot recover. However, the shape of the Australian continent is preserved*

By complementation, it is also possible to define in the same way a geodesic reconstruction by erosion that will result in a closing. All these operators also work on gray-level images, as illustrated in Figure 1.12.



(a) (b) (c)

Figure 1.12. *Gray-level closing by reconstruction: (a) the original image of particle tracks in a detection chamber; (b) the dilation by a* 5×5 *SE; and (c) a reconstruction by erosions. Most of the scintillation noise has been deleted, while retaining the general shape of the tracks*

1.2.2.2. *Space structure, neighborhood*

Until now, we have not approached the subject of the spatial structure of E. The operators we have defined previously do not really depend on it. However, the conditional dilation example illustrates the fact that specifying a structuring element for the dilation also specifies a connectivity. We shall now express this more carefully in the $\mathbb{Z}^n$ case (but our discussion could also be carried out in a similar case in the continuous domain).

Let us begin with the notion of local neighborhood Γ on space E. In the discrete case, Γ is a binary relation on E, i.e. is reflexive ($(x,x) \in \Gamma$) and symmetric ($(x,y) \in \Gamma \leftrightarrow (y,x) \in \Gamma$). We say that (E,Γ) is a (non-oriented) graph. Γ denotes the transform from E to 2^E which associates $x \in E$ with $\Gamma(x) = \{y \in E | (x,y) \in \Gamma\}$, i.e. the set of neighbors of x.

If $y \in \Gamma(x)$, we say that x and y are *adjacent*. In image processing, the more classical relations are defined on a subset of $E \subset Z^2$. For instance, in the 4-connected case, for all $x = (x_1, x_2) \in E, \Gamma(x) = \{(x_1, x_2), (x_1+1, x_2), (x_1-1, x_2), (x_1, x_2+1), (x_1, x_2-1)\} \cap E$. We can define in the same way the 8- or 6- connectivity (see Figure 1.13). The transform Γ is really a dilation, and conversely, from every symmetric dilation defined on a discrete space, we can define a non-oriented graph. If a dilation is not symmetric, this is still true, but we need to involve oriented graphs.

1.2.2.3. *Paths and connectivity*

With the square grid, which is used most often in practice in 2D, it is not possible to use a single definition of neighborhood in all cases. Indeed, we would like to retain in the discrete case the Jordan property of the Euclidean case. This states that any

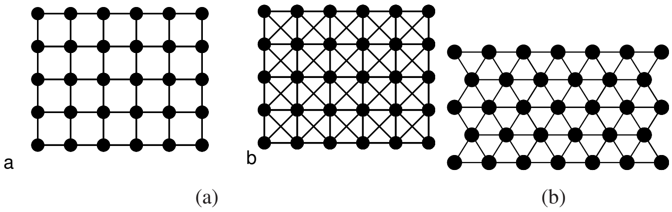


Figure 1.13. *The local grid. In the square grid case, we can specify that each point is connected to its four nearest neighbors as in (a), or its 8 neighbors including the diagonal pixels as in (b). In the case of the hexagonal grid in (c), each pixel has 6 neighbors*

simple closed curve (a closed curve that does not self-intersect) divides the plane into two distinct regions which are connected within themselves: one is of finite extent and the other not. In the discrete case, this property is not true by default. The Jordan problem is illustrated in Figure 1.14.

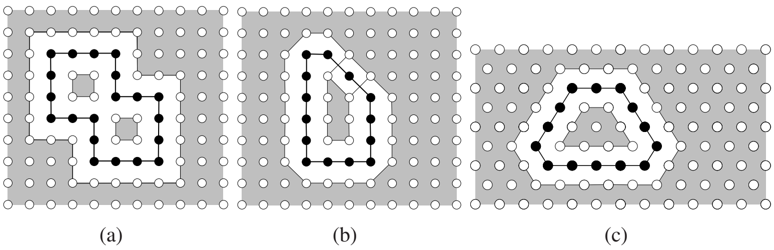


Figure 1.14. *The discrete Jordan property (not true in the square grid by default): (a) the non-degenerate, simple path separates the discrete plane into three connected components; (b) the path does not separate anything at all; and (c) the Jordan property is true (always the case with the hexagonal grid)*

If the grid in Figure 1.14a is 4-connected, the subset of the plane delimited by the path is not connected. If the grid in Figure 1.14b is 8-connected the path does not separate the inside of the curve from the outside. In contrast, with the hexagonal grid it is possible to show that these problems never occur.

In order to solve this problem in a pragmatic way, image analysts often consider two kinds of connectivity [ROS 73, ROS 75] concurrently: one for the foreground objects (inside the curves) and one for the background (outside). A more mathematically meaningful way of solving this problem is to consider a more complete topology for the discrete grid, e.g. following Khalimski [KHA 90].

1.2.3. *Connected filtering and levelings*

Combinations of openings and closings by reconstruction make it possible to define new operators which tend to extend flat zones in images. These combinations are called *levelings*. For more details, see Chapter 8 which is dedicated to this topic. From a more general point of view, levelings are part of a larger family of operators called *connected filters*.

An efficient image representation for connected filtering is the *component tree*. This is studied in detail in Chapter 7 with applications in biology and image compression in Chapters 13 and 16, respectively. A particular case of a connected operator is the area opening, which we present in the following section.

1.2.4. *Area openings and closings*

An opening or a closing using a particular structuring element (SE) modifies the filtered objects or image towards the shape of this SE. For instance, using a disk as an SE tends to round corners. Area openings or closings do not exhibit this drawback.

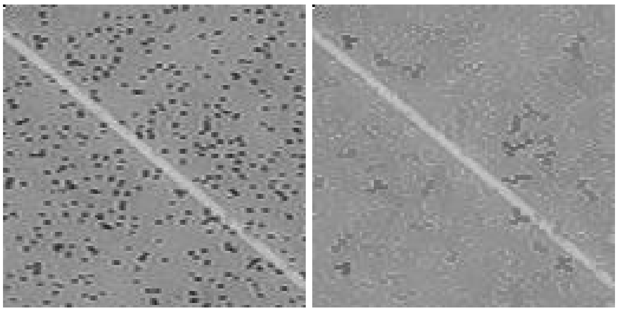
Let $X \subseteq E$, and $x_0, x_n \in X$. A *path* from x_0 to x_n in X is a sequence $\pi = (x_0, x_1, \ldots, x_n)$ of points of X such that $x_{i+1} \in \Gamma(x_i)$. In this case, n is the *length* of the path π. We say that X is *connected* if for all x and y in X there exists a path from x to y in X. We say that $Y \subseteq E$ is a connected component of X if $Y \subseteq X$, Y is connected and Y is maximal for this property (i.e. $Y = Z$ when $Y \subseteq Z \subseteq X$ and Z is connected).

In an informal fashion, an area opening will eliminate small connected components of arbitrary shape of area smaller than a given parameter λ. In a complementary manner, an area closing will fill small arbitrary holes of area smaller than λ.

It is easy to verify that an area opening has the three fundamental properties of an algebraic opening: it is anti-extensive (it eliminates small connected components but leaves the others untouched); it is increasing; and it is idempotent (the small components that are eliminated at the first iteration of the opening remain eliminated, and the large components remain untouched). Area closings are of course extensive instead of being anti-extensive. An area closing is illustrated in Figure 1.15.

1.2.5. *Algebraic filters*

Area filtering can be expressed in a different manner. Let us consider the case of the binary opening. A connected component C with area A will be preserved by any area opening of parameter $\lambda < A$. Clearly, there exists at least one morphological



(a) (b)

Figure 1.15. *Area closing using a parameter of 20 square pixels. Small minima in the image were filled adaptively. Maxima in the image are unaffected (e.g. the small fiber)*

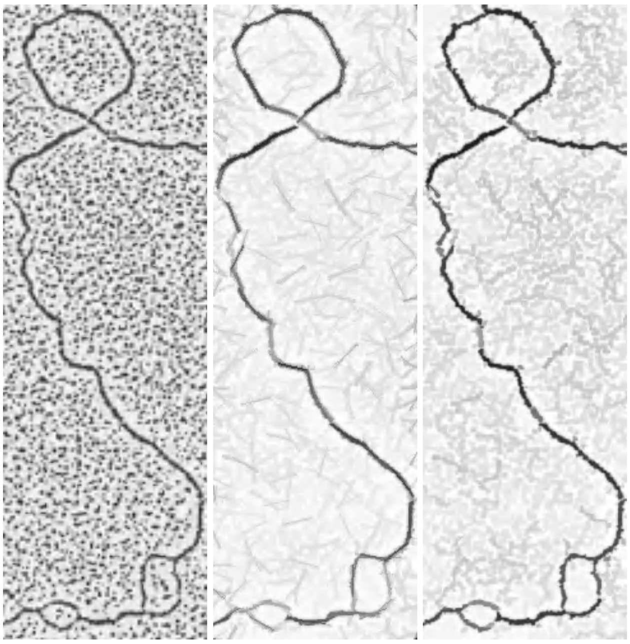
opening by a structuring element of area λ that preserves C, for instance the opening that uses C itself as a structuring element (or any subset of C with area λ).

Knowing that we should preserve all connected components with area at least λ, we deduce that we can consider (at least conceptually) all possible openings with all connected structuring elements of area λ. It is easy to show that the *supremum* of these openings, i.e. the operator that at each point preserves the maximum of all these openings, is itself an opening and that it preserves all connected components with area at least λ. The supremum is therefore the area opening with parameter λ.

It would be theoretically possible to implement the area opening operator by computing the result of all the possible openings using all connected structuring elements with area λ. However, this would be very inefficient as the size of the family of structuring elements increases exponentially with λ. However, the representation of an opening (or a closing) by such a morphological family is useful from the theoretical point of view. There exists a theorem by Matheron [MAT 75] that demonstrates the existence of a morphological decomposition for all openings and closings. For more details, see section 2.4.3.

From the practical point of view, it is useful to remember that a combination by a supremum of openings is itself an opening. Respectively, a composition by an infimum of closings is also a closing. These filters are called *algebraic* openings (respectively, closings).

In Chapter 12, we study how to implement some algebraic filters in practice. As an illustration, Figure 1.16 depicts an application of various algebraic filters to the denoising of thin objects. We used a closing by infimum combination of closings using various structuring elements families, either line segments or adaptive paths [HEI 05]. The objective here is to preserve the object of interest while filtering out the background. In this particular case, the object is not sufficiently locally straight and so paths are better suited to this problem.



(a) (b) (c)

Figure 1.16. *Algebraic closing by infimum composition: (a) an image of a strand of DNA seen in electron microscopy; (b) the infimum of closing by a sequence of segments spanning all orientations; and (c) the result of the infimum by a sequence of paths*

1.2.6. *Granulometric families*

The idea behind granulometries is inspired from sand sifting. When sifting sand through a screen (or sieve), particles that are larger than the dimension of the screen stay on top of the screen while smaller particles sift through. By using a family of screens of various sizes, we can sort the content of a sand pile by particle size.

In the same manner, we can use a family of sieves that are compatible in order to obtain reproducible results. In mathematical morphology, we must use particular families of openings and closings of increasing sizes. These families are indexed by a parameter λ (often an integer) such that:

$$\lambda \geq \mu \Rightarrow \gamma_\lambda \leq \gamma_\mu \text{ and } \varphi_\lambda \geq \varphi_\mu.$$

This property is called the *absorption* property. We often impose that $\gamma_0 = \varphi_0 = \mathrm{Id}$.

As an example of a granulometric family, it is possible to take a sequence of morphological openings or closings. For instance, in 8-connectivity in the square grid, we can use the family of squares B_n of size $(2n+1)\times(2n+1)$ as structuring elements. The resulting family of openings γ_{B_n} or closings φ_{B_n}, indexed by n, verifies the absorption property. We note here that B_1 is the structuring element that corresponds to the basic neighborhood of a pixel. For this reason we refer to it as the *unit ball* of the grid.

We can also use the corresponding openings or closings by reconstruction or take a family of area openings and closings, with increasing parameters.

We shall use granulometric families in section 1.4.2 in this chapter; more details are also given in Chapter 10.

1.2.7. *Alternating sequential filters*

Openings and closings are both increasing and idempotent; they only differ with respect to extensivity. This motivates us to study the class of operators that verify the former properties. We refer to these operators as *morphological filters*. This is both unfortunate and confusing because morphological openings and closings as well as algebraic openings and closings are morphological filters. However, this is to distinguish morphological filters from 'plain' filters which, in image processing, is often a generic term for an image operator.

The theory of morphological filtering allows morphological operators to be efficiently composed. In particular, we can introduce *alternating sequential filters* (ASF) which are, as the name indicates, a composition of openings and closings which form granulometric families of increasing sizes.

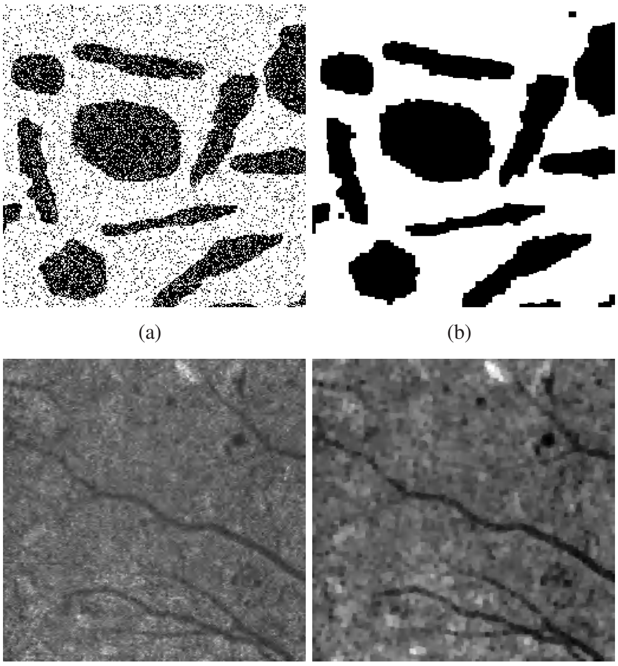
For instance, the *white* ASF, i.e. the ASF beginning with an opening, can be written:

$$\Phi_n(x_i) = \phi_n \gamma_n \phi_{n-1} \gamma_{n-1} \dots \phi_1 \gamma_1. \tag{1.15}$$

The *black* ASF (that begins with a closing) is defined:

$$\Psi_n(x_i) = \gamma_n \phi_n \gamma_{n-1} \phi_{n-1} \dots \gamma_1 \phi_1. \tag{1.16}$$

The theory of morphological filtering is relatively involved and cannot be described adequately here. We simply illustrate it with some elementary applications in Figure 1.17.



(a) (b)

(c) (d)

Figure 1.17. *Using alternating sequential filters: (a) a binary image; (b) the result of a size 2 white ASF; (c) an eye angiogram; and (d) the result of a black size 1 ASF*

Alternating sequential filters can be used to denoise both binary and grayscale images. The result is often easier to segment and analyze. In addition to the size parameter, the structuring element family used also has an impact and can be used to select shapes. Contrary to many filtering methods, these morphological filters allow practitioners to tune their denoising operator to the semantic content of the image and not be affected by the statistical properties of the noise. Morphological filters are therefore generally tailor-made to specific problems, depending on the content of the image under study.

Morphological filtering theory is further developed in Chapters 2, 7 and 8.

1.3. Residues

The operators we have seen until now are generally increasing, meaning that they preserve ordering. In contrast, the operators we present now do not.

What makes the morphological approach different from and complementary to many other approaches is the fact that morphological operators do not seek to preserve information present in an image. Indeed, since the basic operators of morphology are not invertible, we *expect* a reduction in information content after each operator application. The key to success with morphology is to realize this, and to use this defining characteristic to our advantage. We can achieve this by selectively destroying the undesirable content of the image: noise, background irregularities, etc. while preserving the desired content for as long as possible. Figures 1.15 and 1.16 are direct illustrations of this philosophy.

It is sometimes necessary to destroy undesirable content in an image, but not practical to do so. A complementary tactic is to effectively erase the *desirable* portion of an image, but to restore it through a difference with the original image. This gives rise to the idea of *residues*.

Simply put, residues are transforms that involve combinations of morphological operators with the differences (or subtractions). Top-hat transforms, morphological gradients [RIV 93] and other similar transforms that we present in the next section are all examples of residues.

Residues are generally well behaved in morphology, precisely because the basic properties of morphological operators are in our favour. For instance, because openings are anti-extensive, the difference between the original image and any opening derived from that image will always be positive.

1.3.1. *Gradients*

The gradient of an image is basically its first derivative. For a 2D or generally nD image ($n > 1$), the gradient operator produces an n*-vector* at each point, where each component corresponds to the slope along the n principal directions of the grid in the discrete space. With morphology, we most often only consider the magnitude of the gradient at each point, which is a scalar irrespective of the dimension of the image. Gradients show the amount of local variation in the image. Zones of high gradient typically correspond to object contours or texture.

Using erosions and dilations, we can define the morphological gradient as follows:

$$Grad(F) = \delta_B(F) - \varepsilon_B(F). \tag{1.17}$$

There are two other definitions:

$$\begin{aligned} Grad^+(F) &= \delta_B(F) - F \\ Grad^-(F) &= F - \varepsilon_B(F). \end{aligned} \tag{1.18}$$

These are the *external* and *internal* gradients, respectively. In general, B is taken to be the unit ball of the grid. We shall see an example of the use of the morphological gradient operator when we study the watershed line in section 1.5.

1.3.2. *Top-hat transforms*

So-called top-hat transforms are the pixel-wise difference between an original image and an opening of this image (for white top-hats) or between the closing of an image and its original (for black top-hats). Since top-hats essentially show what the opening or closing has deleted from the original image, the former makes it possible to detect peaks and bright small areas in the original image; the latter finds valleys and small troughs in the image. The black top-hat is the white top-hat of the complementary image. Figure 1.18 is an illustration of the principles of white top-hats, while Figure 1.19 depicts an application to cell fluorescence microscopy. There are as many top-hats as there are different openings and closings.

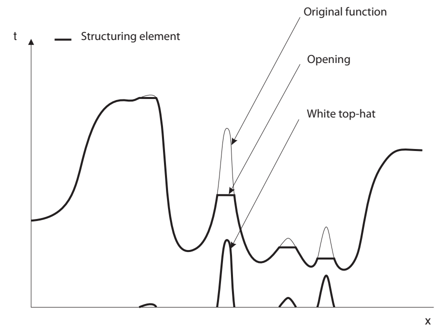
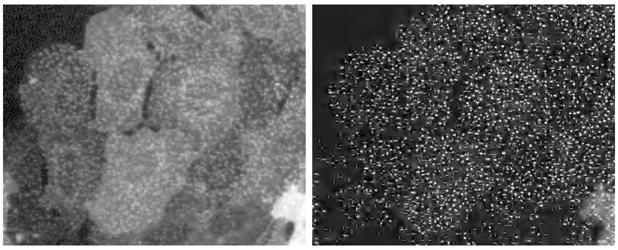


Figure 1.18. *White top-hat of a 1D signal*

1.4. Distance transform, skeletons and granulometric curves

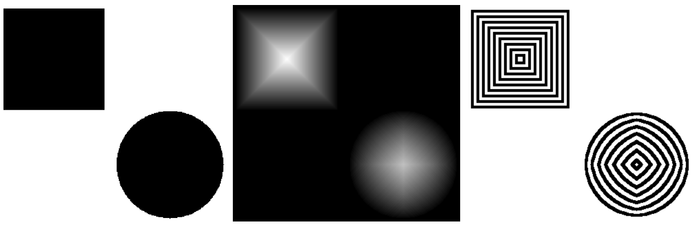
Let $X \subseteq E$. The distance from x to y in X is either the length of the smallest path from x to y within X, or $+\infty$ if there does not exist a path from x to y that stays within X.

The concept of distance makes it possible to introduce the related idea of *distance transform*. This associates each point x from a set X with the distance from this point to the nearest point in the complementary set of X (see Figure 1.20).



(a) (b)

Figure 1.19. *Top-hat on an image of cells: (a) original image (small bright spots are vesicles in the cells, made fluorescent through the use of a bio-molecular marker); (b) the result of a thresholding of the white top-hat overlaid over the original image*



(a) (b) (c)

Figure 1.20. *The distance transform of a set X is the application that from each point of X associates its distance to the complementary set. In this illustration we consider the 8-distance but this is by no means a rule: (a) set; (b) distance transform; (c) level sets*

1.4.1. *Maximal balls and skeletons*

A maximal ball with radius R is the set of points located at a distance less than or equal to R from a central point p. It is obvious this definition depends on the distance used. For instance, using the 8-distance in 2D, the ball of radius 3 is a 7×7 square. When the Euclidean distance is used, the ball is a discrete disk.

A maximal ball B relative to a set E is a ball such that there does not exist a ball B' such that $B \subset B' \subset E$ (see Figure 1.21). This apparently simple notion is useful to define some interesting residues, in particular the *skeleton*. This notion has been known since the 1930s [BOU 32, DUR 30, DUR 31], but was popularized in image processing contexts in the 1960s [BLU 61, CAL 68] under the name of *medial axis* of E. The medial axis is defined as the collection of all centers of maximal balls of E.

In Euclidean space, the medial axis is called the skeleton. The skeleton of a set O that is connected, open, non-empty and bounded has many interesting properties from

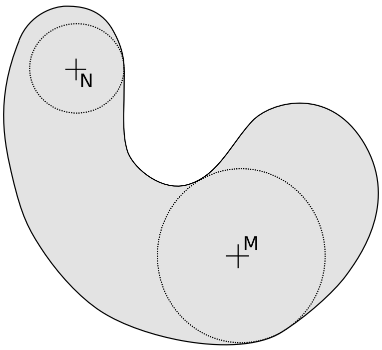
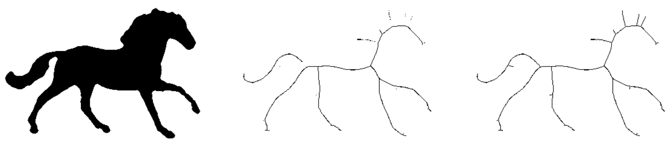


Figure 1.21. *Two balls included in a binary set. The ball centered at point M is maximal because no other ball from the same family (here a family of disks) can contain it and simultaneously be included in the binary set. The ball centered in N is not maximal*

the geometrical and topological point of view: it is connected, homotopic to O and negligible from the point of view of the Lebesgue measure (although it may be dense everywhere in O) [RIV 87].

In the discrete case, the centers of maximal balls in O are well defined but are not necessarily located on the grid; the medial axis may therefore not be connected. We then define informally the discrete skeleton $S(O)$ as a connected set, included in O, homotopic to O and as thin as possible. This notion is disjoint from the notion of medial axis, but it is possible to constrain the discrete skeleton to contain the medial axis. In practice, both discrete skeletons and medial axes are often noisy. To be able to use them in practice, we must be able to filter them.

In Chapter 10, these notions are more precisely defined and algorithms are given to compute them. An example of filtered medial axis and filtered skeleton are given in Figure 1.22.



(a) (b) (c)

Figure 1.22. *An example of a skeleton: (a) a binary image; (b) its filtered medial axis; and (c) a filtered skeleton of (a) that contains the medial axis*

1.4.2. *Granulometric curves*

Mathematical morphology, even within itself, is capable of providing information on the size of objects in images in several different ways. One of these methods relies on the notion of *granulometries*, which is directly derived from the notion of morphological filter (see section 1.2.6 and Chapter 10).

The granulometric curve of an image is a representation of the distribution of sizes in an image. This is based on the observation that intermediate residues of a granulometric family γ_n, indexed by n, are characteristic of the size of objects in images. More formally, the granulometric curve is the function $G_f(\lambda)$ defined on the interval of λ, such that:

$$G_I(\lambda) = \sum I - \sum \gamma_\lambda(I) \qquad \lambda \in [0, 1, 2, \ldots, R] \tag{1.19}$$

where I is the input image, $\sum I$ is the sum (or integral) of all the pixels in the image and R is the value of λ for which no further change occurs due to γ_R (since the image has become constant). For finite discrete images, the value R always exists. If we use closings instead of openings, the operands on either side of the subtraction sign in equation (1.19) are exchanged (since all closings are extensive, and so $\varphi(I) \geq I$).

Granulometric curves by openings make it possible to estimate the size of peaks in images, while granulometries by closings measure the size of troughs.

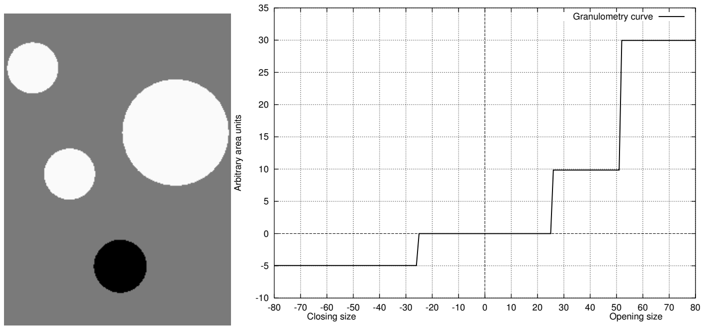
It is possible to build a granulometric curve using both openings and closings. In order to obtain a single continuous curve, convention states that the parameter of the closings is given in abscissa from the origin towards the negative value and the parameter of the openings is given from the origin towards the positive values. The same sign convention is used for both openings and closings; such a curve is depicted in Figure 1.23.

In this example, we illustrate the fact that the granulometric curve records the volume of image (i.e. the area of features times their gray level) that is erased beyond a certain size, both for openings and closings.

1.4.2.1. *Applications*

The granulometric curve summarizes the distribution of size of objects in an image, without necessitating a segmentation step. Consequently, the notion of object is not well defined in this context. We can only talk about volumes of gray levels. In addition, since the granulometric curve is one-dimensional, the information content is necessarily reduced from that of the whole image.

The interpretation of this content is not always easy. It can sometimes be interpreted in terms of texture energy, as in the example given in Figure 1.24



(a) (b)

Figure 1.23. *A granulometric curve by openings and closings of an artificial image, using a family of Euclidean disks as structuring elements (indexed by their radii). The increasing sizes of openings go from the origin towards $+\infty$ and the increasing size of closings go from the origin to $-\infty$: (a) image and (b) granulometric curve*

in the context of a study involving the aging of steam pipes used in electricity production. Many applications use granulometries to estimate size-related parameters for subsequent procedures; see [COM 07].

1.4.2.2. *Granulometries by erosions and dilations*

It is also possible to produce granulometries by using only erosion or dilations. There is a strong link between these and skeletons [MAT 92].

1.4.3. *Median set and morphological interpolation*

Another application of distances worth mentioning, also related to skeletons, is their capacity for computing a median set used as an interpolation algorithm. In the literature, median sets appeared in the work of Casas [CAS 96] and Meyer [MEY 96]. The equation of the underlying operation and its basic properties were given by Serra in [SER 98b]. Iwanowski has successfully developed it for various morphings on still images and video sequences, in black and white and in color [IWA 00a]. More recently, Vidal *et al.* used a recursive technique for improving the interpolations [VID 07].

Recall that the Hausdorff distance is the maximum distance of a set from the nearest point in the other set. It measures how far two subsets of a metric space are from each other. Informally, two sets are close in the Hausdorff distance if every point of either set is close to some point of the other set.

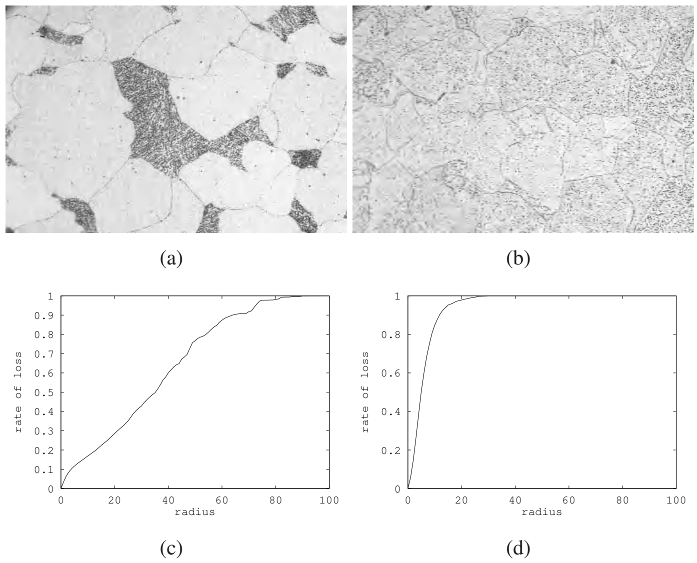


Figure 1.24. *Application of granulometries: granulometric curves of surface microscopy images of pig iron steam pipes, in the case of a young pipe and an old pipe. Age deteriorates grain boundaries, which results in a larger number of small grains compared with younger samples: (a) young sample; (b) old sample; (c) young sample curve; (d) old sample curve*

Consider an *ordered* pair of closed sets $\{X, Y\}$ with $X \subseteq Y$ and such that their Hausdorff distance is finite. Their median element is the closed set $M(X, Y)$, composed of X and Y and whose boundary points are equidistant from X and the complement Y^c to Y. In other words, the boundary of M is nothing but the skeleton by zone of influence (also known as the generalized Voronoï diagram) between X and Y^c.

The set M can depend on a parameter α which weights the relative importances of X and Y in the interpolation. The analytic expression of the weighted median set $M_\alpha(X, Y)$ is obtained from its two primitives X, Y by taking the union [SER 98b]:

$$M_\alpha(X, Y) = \cup_\lambda \left\{(X \oplus \alpha\lambda B) \cap (Y \ominus (1-\alpha)\lambda B)\right\}. \tag{1.20}$$

In the general case, for two sets A and B with non-empty intersection (i.e. $A \cap B \neq \emptyset$), we set $X = A \cap B$ and $Y = A \cup B$ so that $X \subset Y$ and apply equation (1.20). Figure 1.25, whose steps are described in more detail in [IWA 00b], depicts the progressive passage from a bird to a plane as α varies from 0 to 1. As the map defined by equation (1.20) is increasing for both X and Y, it extends directly to digital numerical functions by simply replacing union and intersection by supremum and infimum, respectively.

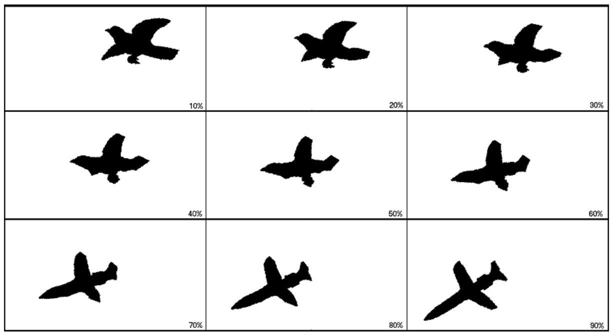


Figure 1.25. *Series of morphological interpolations from a bird to a plane, by means of equation (1.20)*

1.5. Hierarchies and the watershed transform

If we consider anew the analogy between grayscale images and a terrain topography, we can define an interesting transform called the *watershed line*. By analogy with hydrology, imagine a drop of water falling on the terrain represented by the image. Assuming sufficient regularity of the image, this drop will fall towards a local minimum in the image. With each local minimum M, we can refer to the set of points p such that a drop of water falling on p ends up in M. This set is call a *catchment basin*. The points located at the border of at least two such basins constitute a set of closed contours called the *watershed line*; see Figure 1.26 for an illustration of this.

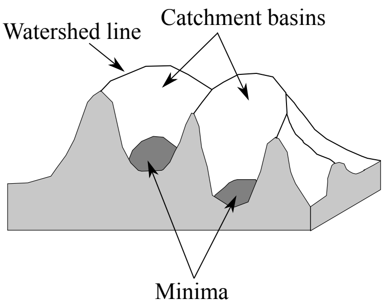


Figure 1.26. *The watershed line*

A different view of the watershed line consists of not considering the points p, but starting from the minima M. We imagine that the image is inundated starting from the bottom (as if every minimum in the image is hollow and the whole image was dipped in water from the bottom). In this case every local minimum gradually fills with water, and the watershed line is the locus of the points where at least two water bodies meet. Although both visions are equivalent in the continuous domain under sufficient regularity assumptions [NAJ 94b], they are not compatible in all discrete frameworks and, notably, not compatible in the pixel framework [COU 07c, NAJ 05].

Although the previous explanation may not appear to be very formal, the literature on the topic of watershed properties and algorithms is abundant; see [BEU 79b, COU 05, MEY 94b, ROE 01, VIN 91c]. The formalization of the various concepts derived from the watershed in the discrete case, as well as the mathematical properties of the objects so obtained, are presented in Chapter 3.

The watershed line transform forms the basis of a powerful and flexible *segmentation* methodology introduced at the Centre de Morphologie Mathematique in the 1970s [BEU 79b] and further developed in the 1990s [MEY 90b, VIN 91c]. This methodology was later unified using hierarchical approaches [BEU 94, NAJ 96].

The general idea is that we first need to produce internal *markers* of the objects under study. These are binary sets which can be labeled (i.e. given a distinct gray level for each connected component), that are included in the objects sought. The shape of these markers is unimportant; only their position and their extent matter. In a similar manner, we seek markers that are external to the objects, i.e. totally included in the background. A function which exhibits high values near contours and low values in a near-constant area of the image is used. Usually some regularized version of the gradient operator can be employed. This function is then reconstructed using the geodesic reconstruction operator of section 1.2.2 by imposing all markers (both internal and external) as minima in this function, and by eliminating the original minima present in the function. A single watershed line is then present. This separates internal and external markers, and tends to place itself on the contour of objects to be segmented.

Many chapters of the Applications part of this book (Part V, notably Chapters 14 and 17) use one of the many variations of the watershed. It is therefore useful to illustrate the above procedure on a simple example due to Gratin [GRA 93]. Here we seek to segment a 2D magnetic resonance image (MRI) of an egg. On this image (Figure 1.27), markers for the exterior of the egg, the white and the yolk are set manually, but it is of course possible to obtain these through an automated procedure. Contrary to expectations and despite the simple nature of the problem, a simple thresholding does not yield good results due to the high level of noise. In contrast, the watershed segmentation procedure result is almost perfect.

The general methodology for morphological segmentation is developed further in Chapter 9. It relies on defining some criterion that induces a hierarchy of segmentations, i.e. a nested sequence of connected partitions. Any hierarchy of segmentations is equivalent to a specific watershed referred to as a *saliency map* [NAJ 96] or

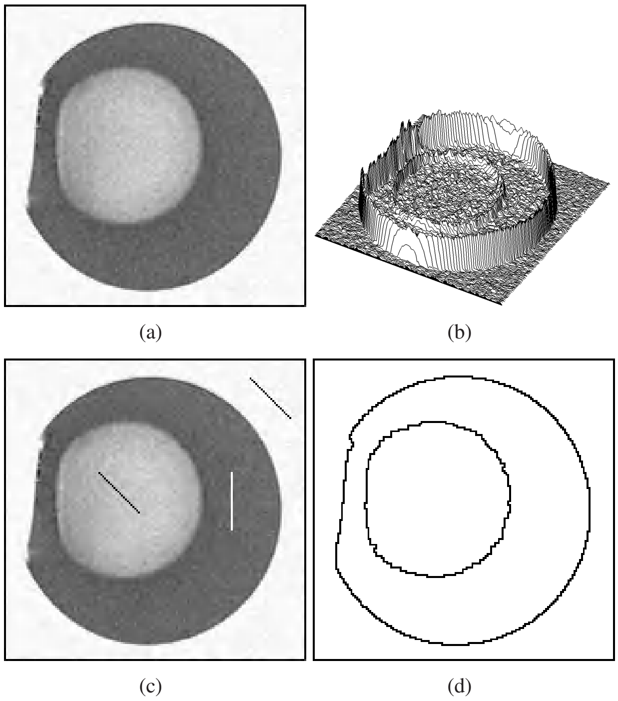
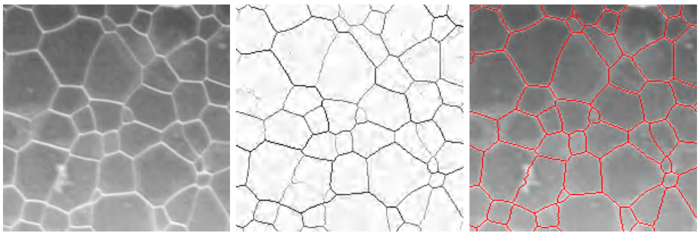


Figure 1.27. *MRI of an egg: (a) original image (courtesy of N. Roberts, University of Liverpool); (b) the gradient of this image (seen as a 3D terrain); (c) manually set markers; and (d) result of the segmentation*

ultrametric watershed [NAJ 09a, NAJ 09b]. Filtering such a watershed amounts to transforming the hierarchy into another watershed.

Figure 1.28 illustrates the principle on an image of uranium oxide. We want to extract the cells but, unfortunately, a brute-force watershed application gives an oversegmented image. Instead of trying to find some markers, we can filter the image to remove the background noise. Here the chosen filter depends on a depth criterion (see Chapters 7 and 9). Rather than setting a fixed level of noise reduction for the filtering, it is better to compute the whole hierarchy of segmentations that can be obtained by varying the parameter.

The resulting hierarchy is represented as a saliency map in Figure 1.28b. Any threshold of Figure 1.28b gives a segmentation. The more a contour is present in the hierarchy of segmentations, the more visible it is. It can be seen that there is a large difference between the noise contours and the ‘true’ contours; choosing the correct level of thresholding is therefore easy (Figure 1.28c). It is even possible to use granulometric curves on the hierarchy to automatically determine the correct thresholding parameter.



(a) (b) (c)

Figure 1.28. *A saliency map is more than a visual representation of a hierarchy of segmentations: (a) original image; (b) saliency map of (a); (c) threshold of (b)*

In fact, it can be proved that a saliency map is more than a visual representation of a hierarchy of segmentations, as any saliency map can be directly *computed* as an ultrametric watershed [NAJ 09b, NAJ 09a].

1.6. Some concluding thoughts

This introductory chapter only describes a small portion of existing morphological operators. Some parts that we have not developed in this chapter pertain, for instance, to the non-deterministic aspects of mathematical morphology; they are presented in Chapters 4, 5 and 6 with applications in Chapters 19 and 20. Other parts rely on the theory of partial differential equations; we recommend [GUI 04] (available online) to the interested reader.

We hope these words will inspire readers to delve in more detail into the theory and practice of mathematical morphology. Such a reader should, we hope, find enough in the rest of this book to satisfy his or her expectations.