The morphological approach to segmentation: an introduction

Luc VINCENT & Serge BEUCHER

Centre de Morphologie Mathématique
École Nationale Supérieure des Mines de Paris
35, rue Saint-Honoré
77305 Fontainebleau cedex, FRANCE

Abstract

These lecture notes present several segmentation tools provided by Mathematical Morphology. Together with that presentation, they describe the basic morphological transformations involved in the required algorithms. They then attempt to derive a general approach of segmentation problems using Mathematical Morphology, which is based on the two concepts of *markers* and *marking function*. This philosophy is illustrated with several examples.

Contents

1 Introduction

One of the most current problems in image analysis consists in extracting from a scene the objects or the regions to be analyzed. This *image segmentation* may well consist in a simple thresholding, but can also be a very complex operation, based, e.g., on the objects geometry [25], on *computational geometry* techniques [23], on texture analysis [11], on region growing approaches [13,22], etc...

In order to solve this kind of problems, mathematical morphology provides the user with a large set of tools, working on binary images as well as on grey-scale ones. The purpose of these lecture notes is to describe these tools and to illustrate with numerous examples the way to use them. However, before presenting them, we will briefly recall the basic transformations of mathematical morphology (MM) in the binary case (binary morphology) as well as in the grey-scale one (decimal morphology).

2 Morphological transformations

Mathematical Morphology refers to an image processing methodology which is based on neighborhood or *morphological* operations. These transformations act on sets in binary morphology and on functions in the decimal case. Since the result of any of these transformations is of the same kind as the object on which it is applied (a set is transformed into a set, a function, into a function), it is now easy to build more and more complex transformations either by iteration or by concatenation of any primitive transformation. We thus define operations of higher level which, in spite of the more complex algorithms they involve, can be easily understood and used, since their finality is known and well described. For example, the detection of isolated cells in a cytologic preparation may require hundreds of elementary transformations. Despite this apparent complexity, the process is rather comprehensible, and its result quite predictable.

Throughout these lecture notes, an example will be used as a guideline for illustrating the effects of the numerous morphological transformations that we are going to present: the image 1a, which represents coffee beans. This grey-scale image is in fact practically a binary one, since a simple thresholding yields the binary image 1b. However, one can easily observe on this picture that many coffee beans overlap. As soon as we will have presented enough morphological tools, we will be able to *segment* this image, i.e. to separate the different coffee beans.

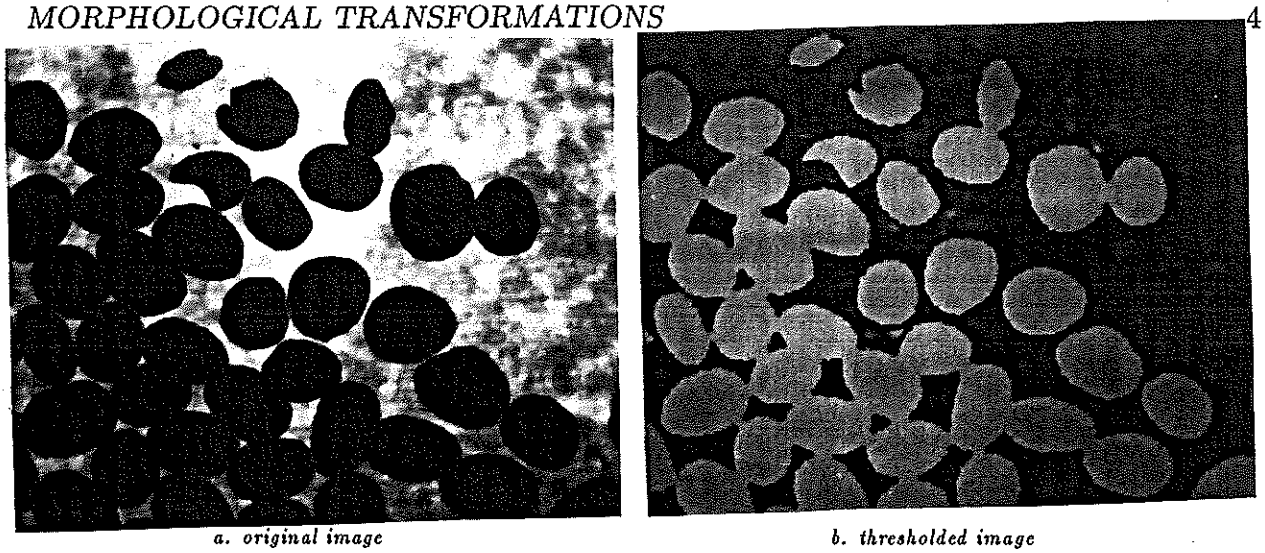
2.1 Some properties of morphological transformations

The transformations that we will describe are often equipped with some interesting properties that we will now review. Let Φ be a transformation which is first supposed to be a binary one. Φ therefore acts on elements of $\mathbb{R}^2$, i.e. on sets of $\mathbb{R}^2$ ($\Phi : \mathcal{P}(\mathbb{R}^2) \longrightarrow \mathcal{P}(\mathbb{R}^2)$). It is said to be an *extensive* transformation if and only if:

$$\forall X \in \mathcal{P}(\mathbb{R}^2), \quad X \subseteq \Phi(X) \tag{1}$$

and *anti-extensive* if and only if:

$$\forall X \in \mathcal{P}(\mathbb{R}^2), \quad \Phi(X) \subseteq X. \tag{2}$$



a. original image *b. thresholded image*

Figure 1: Coffee beans

Besides, Φ is *increasing* if and only if it preserves the order relations between the elements of $\mathbb{R}^2$ on which it acts, i.e.:

$$\forall (X,Y) \in \mathcal{P}(\mathbb{R}^2)^2, \quad X \subseteq Y \Longrightarrow \Phi(X) \subseteq \Phi(Y). \tag{3}$$

Φ will also be said to be *idempotent* when applying it several times successively comes down to applying it only once:

$$\forall X \in \mathcal{P}(\mathbb{R}^2), \quad \Phi(\Phi(X)) = \Phi(X). \tag{4}$$

Finally, two transformations Φ and Ψ are said to be *dual* if and only if applying the former to a set X is equivalent to applying the latter to the complement X^C of this set:

$$\forall X \in \mathcal{P}(\mathbb{R}^2), \quad \Phi(X)^C = \Psi(X^C). \tag{5}$$

These properties can be immediately transcribed into the decimal case where Φ acts on functions from $\mathbb{R}^2$ into $\mathbb{R}$, i.e. on elements of $\mathcal{F}(\mathbb{R}^2, \mathbb{R})$. In this case, the order relationship between functions is the following:

$$\forall (f,g) \in \mathcal{F}(\mathbb{R}^2, \mathbb{R})^2, \quad f \leq g \Longleftrightarrow \forall x \in \mathbb{R}^2, f(x) \leq g(x) \tag{6}$$

2.2 Erosion and dilation

These two operations are the classics of mathematical morphology [25]. They constitute the basic stone of a large number of more elaborated transformations. Before presenting them, a few notations that we shall use constantly have to be defined: let O be an arbitrary point of $\mathbb{R}^2$ which is referred to as the *origin*. With any point x in $\mathbb{R}^2$, we associate the vector $\overrightarrow{Ox}$. Hereafter, we use an abbreviated notation in which, according to the context, x denotes the point $x \in \mathbb{R}^2$ or the vector $\overrightarrow{Ox}$. Let now $B \in \mathcal{P}(\mathbb{R}^2)$ be an arbitrary set. We denote $\check{B}$ the transposed set (See Fig. 2):

$$\check{B} = \{-x, x \in B\}. \tag{7}$$

Besides, for any $a \in \mathbb{R}^2$, we denote B_a the translation of set B by vector a:

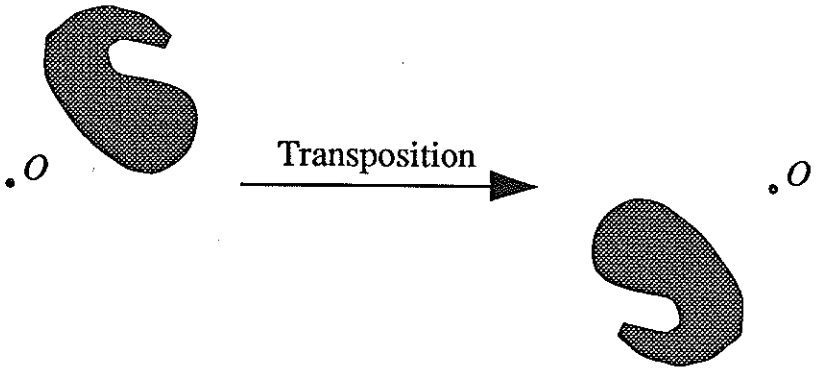


Figure 2: Transposition of a set

$$B_a = \{a + x, x \in B\}. \tag{8}$$

Consider now a set $X \subseteq \mathbb{R}^2$. The dilation of X by B is the set of the points x in $\mathbb{R}^2$ such that the intersection between X and B_x is not empty (see Fig. 3a). The resulting set is denoted $X \oplus \check{B}$:

$$X \oplus \check{B} = \{x \in \mathbb{R}^2, X \cap B_x \neq \emptyset\}. \tag{9}$$

The erosion is defined in a similar way: X eroded by B, denoted $X \ominus \check{B}$ is the set of the x in $\mathbb{R}^2$ such that B_x is totally included in X (see Fig. 3b):

$$X \ominus \check{B} = \{x \in \mathbb{R}^2, B_x \in X\}. \tag{10}$$

The set B, which plays a particular role in these transformations, is called *structuring*

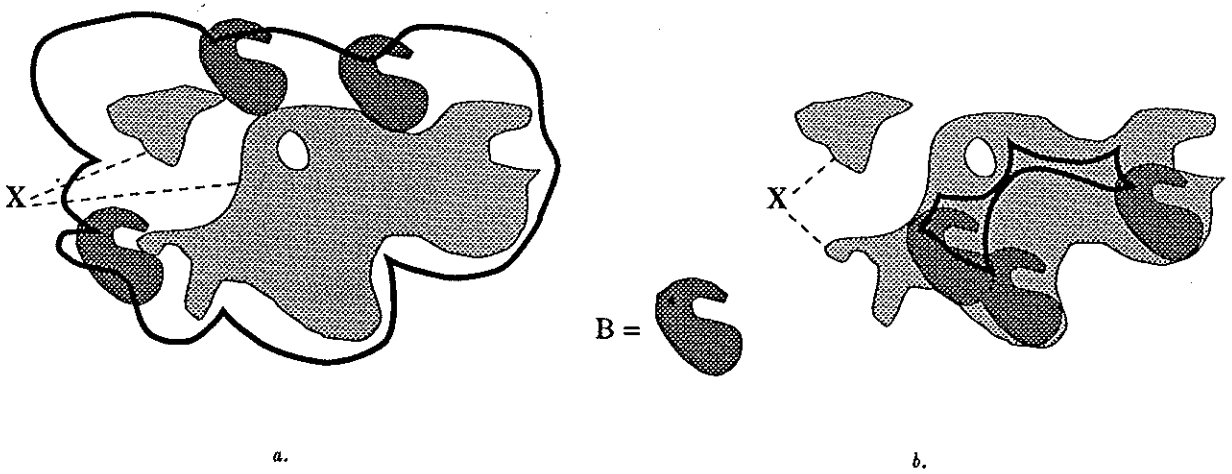
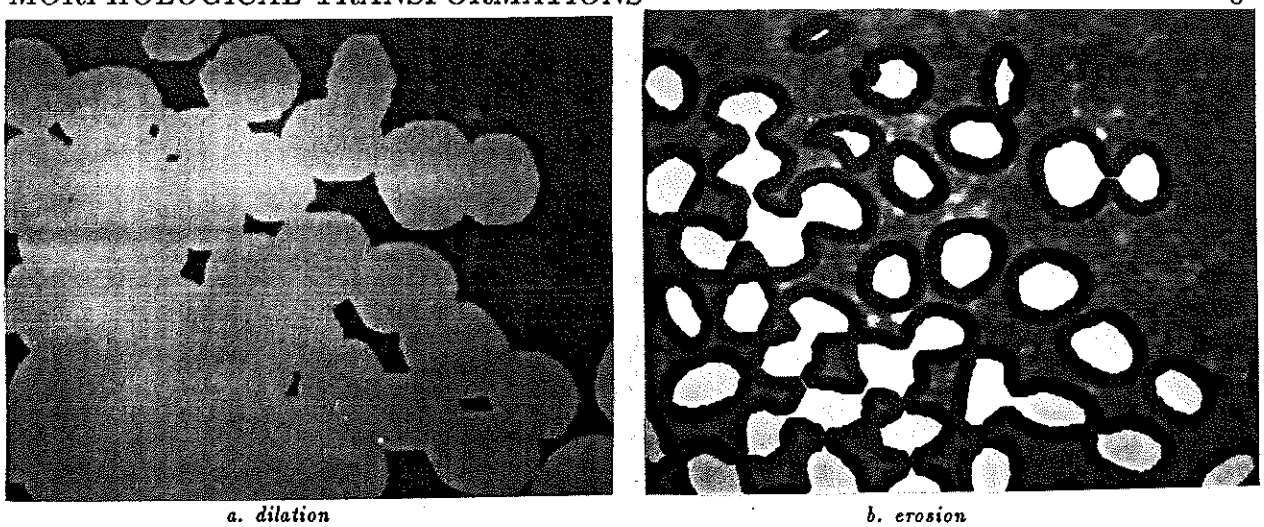


Figure 3: Dilation and erosion of X by B

element. Fig. 4 illustrates the effects of an erosion and of a dilation of our coffee beans by a hexagon.

Dilations and erosions are *increasing* transformations. Moreover, they are *dual,* which means that the dilation of a set X is identical to the complementary set of the erosion of X^C:

$$X \oplus \check{B} = (X^C \ominus \check{B})^C. \tag{11}$$

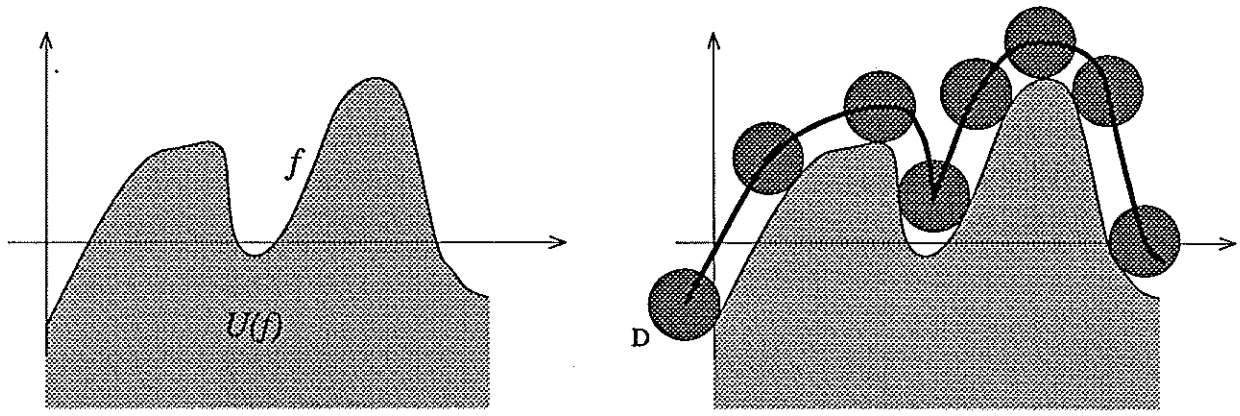


a. dilation *b. erosion*

Figure 4: Hexagonal binary dilation and erosion

One can also say that when eroding a set, its complement is automatically dilated. In this sense, the morphological transformations not only act on a given set X, but on the whole space. The notations $\oplus$ and $\ominus$ were introduced by Minkowski [21]. As a matter of fact, the dilation (resp.the erosion) of X by B is identical to the *Minkowski addition* (resp. *subtraction*) of X and B.

These two notions are easily extended to the decimal case [4]. To do so, it suffices to introduce the notion of *subgraph* or *umbra* of a function. Let f be a mapping from $\mathbb{R}^2$ into $\mathbb{R}$. Its (closed) umbra $U(f)$ is the set of the points (x, y) in $\mathbb{R}^2 \times \mathbb{R}$ such that $y \leq f(x)$. We can therefore dilate and erode this set by a three-dimensional structuring element B. One can show that the dilation (resp. the erosion) of the umbra of f by B is still the umbra of a function, which is called the dilation (resp. the erosion) of f by B and denoted $f \oplus \check{B}$ (resp. $f \ominus \check{B}$). Fig. 5 shows an example of a dilation of a decimal function in the one-dimensional case, i.e. when $f : \mathbf{R} \longrightarrow \mathbf{R}$.



a. function f and its umbra U(f) *b. dilation of f by disc D*

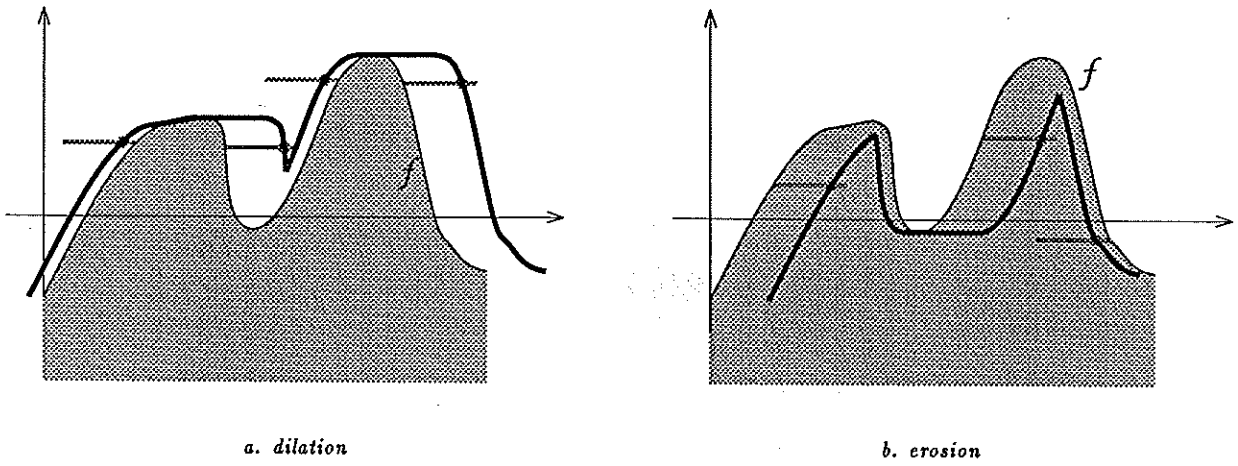
Figure 5: Grey-scale dilation

For the sake of algorithmic simplicity, morphologists most of the time restrict themselves to *planar* structuring elements, i.e. structuring elements which are included in $\mathbb{R}^2 \times \{0\}$. In

this particular case, one can prove that the dilation and the erosion of a function f by a planar structuring element B can be written:

$$\forall x \in \mathbb{R}^2 \quad \begin{cases} (f \oplus \check{B})(x) = \sup_{y \in B_x} f(y) \\ (f \ominus \check{B})(x) = \inf_{y \in B_x} f(y) \end{cases} \tag{12}$$

Fig. 6 shows—again in the one-dimensional case—how a function is transformed when it is dilated and eroded by a planar structuring element. In particular, the dilation widens



a. dilation *b. erosion*

Figure 6: Dilation and erosion of a function by a planar structuring element

peaks and fills valleys whereas the erosion has the opposite effect. On Fig. 7, the same transformations are applied to a real grey-scale image.

a. original image *b. dilation* *c. erosion*

Figure 7: Decimal dilation and erosion by a hexagon of size 2

2.3 Hit or miss transformation, thinning, thickening

Let us consider now a structuring element B made of two disjoined parts B_1 and B_2, with *the same origin.* Such an element allows us to define a new transformation, called the *hit or miss transformation (HMT).* This operation, denoted $\star$, is more general than the erosion or the dilation, and is defined as follows:

$$\forall X \in \mathbb{R}^2, \quad X \star B = (X \ominus \check{B}_1) \cap (X^C \ominus \check{B}_2). \tag{13}$$

The HMT is neither increasing, nor extensive or idempotent... It is at the basis of the definitions of the thinning $X \circ B$ of X by B and of the thickening $X \odot B$ of X by B:

$$X \circ B = X/(X \star B) \quad (14)$$
$$X \odot B = X \cup (X \star B) \quad (15)$$

The thinning is an *anti-extensive* operation whereas the thickening is an *extensive* one. Moreover, the dual operation of the thinning by $B = (B_1, B_2)$ is the thickening by $B' = (B_2, B_1)$:

$$X \circ B = (X^C \odot B')^C, \quad \text{with } B = (B_1, B_2) \text{ and } B' = (B_2, B_1). \quad (16)$$

Thinnings and thickenings can also be extended to the decimal case. They are mainly used in algorithms generating connected skeletons and skeletons by influence zones (See § 3.1).

2.4 Opening and closing

As was said above in § 2, morphological transformations and in particular erosions and dilations can be iterated. One can show that under certain conditions that will be discussed later (See § 2.4.1), dilating a set X (or a function f) n times successively is equivalent to dilating it only one time by the homothetic structuring element. By duality, this remark applies to the erosion. On the other hand, erosions and dilations can also be concatenated. Two transformations can thus be defined: the *opening* and the *closing*. In the binary case for instance, the opening of a set X by a structuring element B, denoted $(X)_B$, is defined by:

$$(X)_B = (X \ominus \check{B}) \oplus B. \quad (17)$$

Therefore, in order to open X by the structuring element B, we first erode X by B and then dilate the resulting set by the transposed set $\check{B}$ of B (indeed, $\check{\check{B}} = B$!...). The opening operation thus defined is *anti-extensive*. Fig. 8 displays an example of a binary opening by a disc.

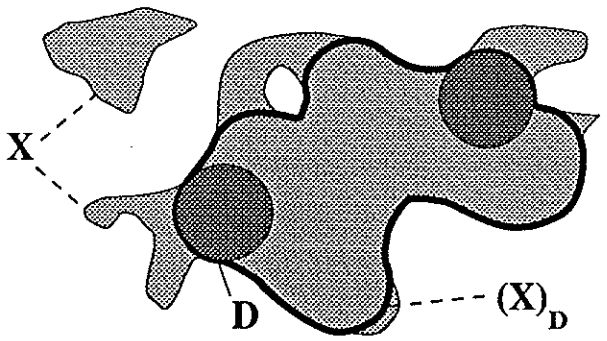


Figure 8: Opening of a set X by a disc D

Similarly, we define the closing of X by B, denoted $(X)^B$, as follows:

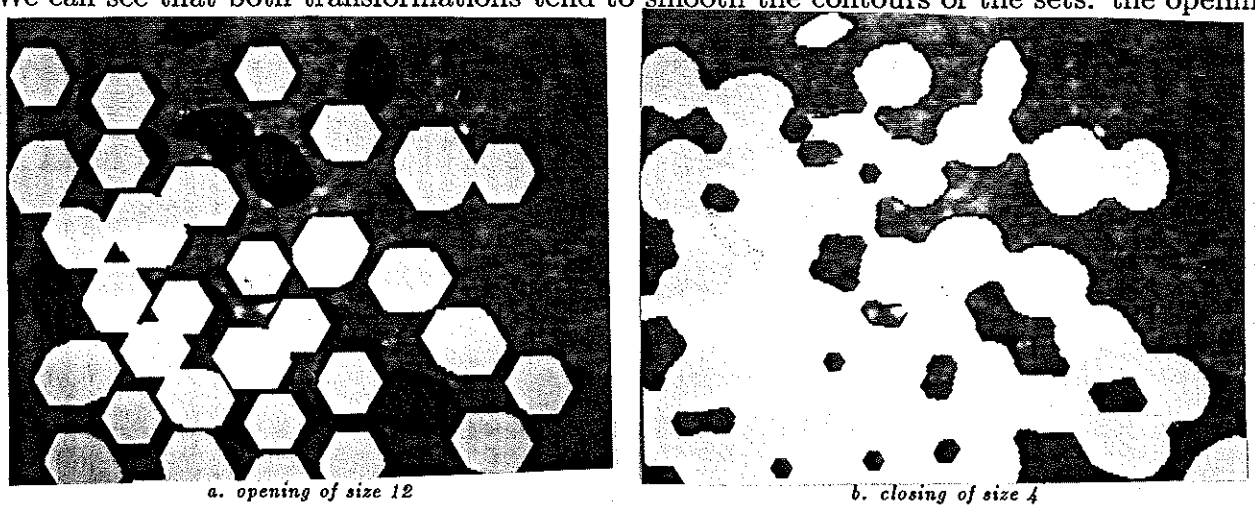
$$(X)^B = (X \oplus \check{B}) \ominus B. \quad (18)$$

This operation is *extensive*.

Like the pair erosion-dilation, opening and closing are dual operations:

$$(X)_B = ((X^C)^B)^C \tag{19}$$

Fig. 9 illustrates the effects of these transformations on our binary image of coffee beans. We can see that both transformations tend to smooth the contours of the sets: the opening

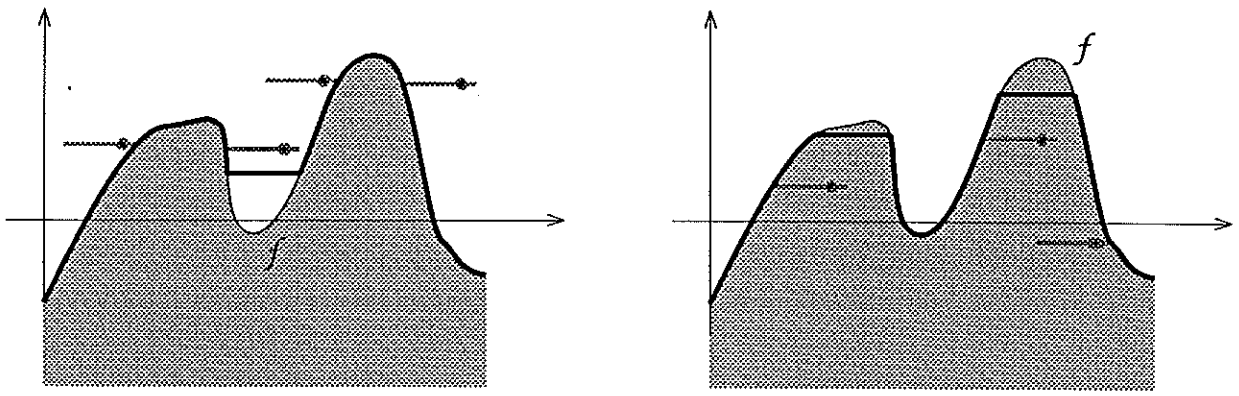


a. opening of size 12 *b. closing of size 4*

Figure 9: Hexagonal binary opening and closing

removes capes whereas the closing fills gulfs. Moreover, the opening may disconnect sets and suppress their small connected components whereas the closing tend to bind particles and to fill their holes. For these reasons, openings and closings serve two major purposes: *granulometries* and *filtering*.

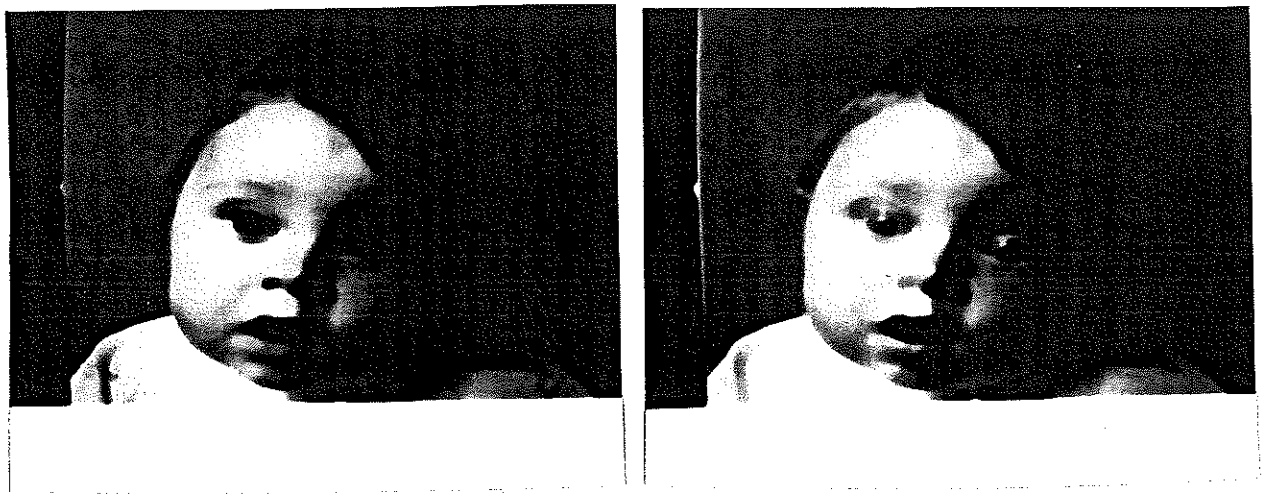
The above definitions of opening and closing easily extend to the decimal case. Fig. 10 shows—again in the one-dimensional case—how these transformations act on functions. One



a. closing *b. opening*

Figure 10: Closing and opening of a function by a planar structuring element

of the remarkable effects is again that of smoothing: the decimal opening tends to suppress the peaks whereas the closing tends to fill up the valleys. The results of a hexagonal opening and of a hexagonal closing on image 7a is displayed on Fig. 11.



a. opening *b. closing*

Figure 11: Hexagonal grey-scale opening and closing

2.4.1 Granulometries

In the binary examples presented earlier, we could notice that an opening suppresses the particles—i.e. the connected components—which are "smaller" than the structuring element (i.e. that cannot contain it). Therefore, it seems possible to sieve a set by using this operation with larger and larger structuring elements. In order to formalize this type of transformation, we introduce the notion of *granulometry* and of *granulometric transformation* [17]. Let $(\phi_\lambda)_{\lambda \in \mathbf{R}^+}$ be a family of transformations depending on a unique parameter λ. This family constitutes a granulometry if and only if the three following properties are satisfied:

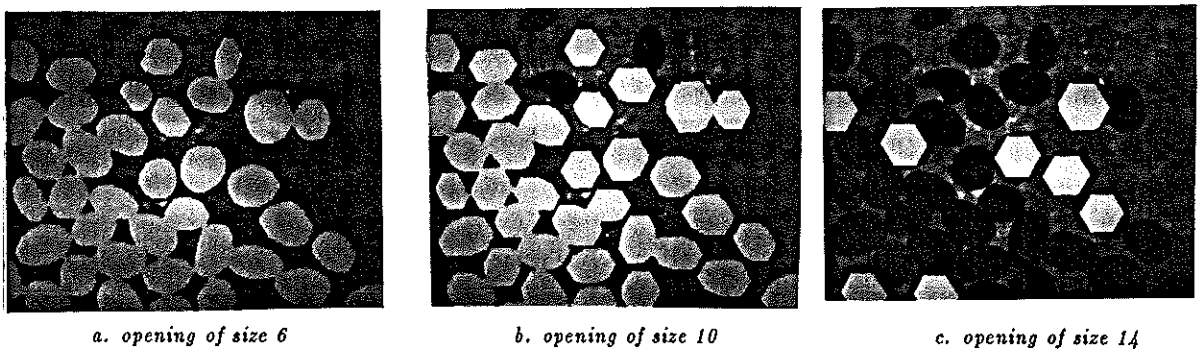
$$\forall \lambda \in \mathbf{R}^+, \quad \phi_\lambda \text{ is increasing,} \tag{20}$$

$$\forall \lambda \in \mathbf{R}^+, \quad \phi_\lambda \text{ is anti-extensive,} \tag{21}$$

$$\forall (\lambda, \mu) \in \mathbf{R}^{+2}, \quad \phi_\lambda \phi_\mu = \phi_\mu \phi_\lambda = \phi_{\max(\lambda,\mu)} \tag{22}$$

Notice that property 22 implies the idempotence of the ϕ_λ's.

One can prove that if B is a convex set, the family of the openings by the homothetics $(\lambda B)_{\lambda \in \mathbf{R}^+}$ of this convex is a granulometry (See Fig. 12).



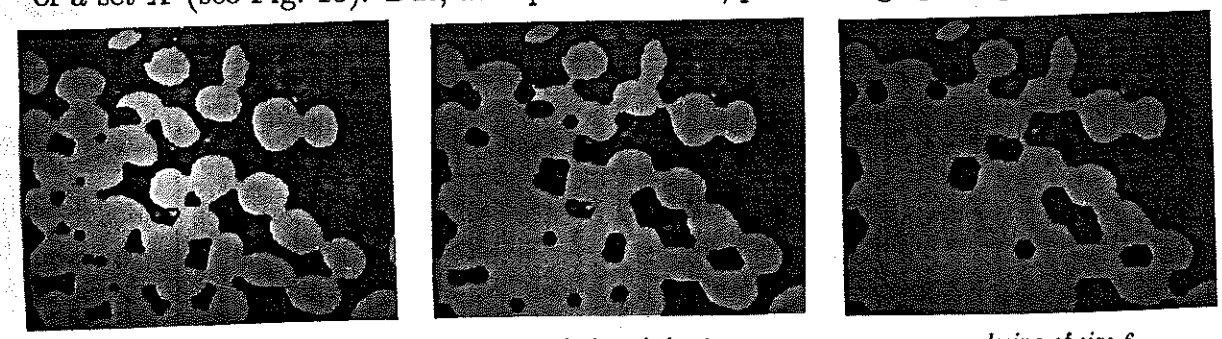
a. opening of size 6 *b. opening of size 10* *c. opening of size 14*

Figure 12: Granulometry by openings of the coffee beans image

Note that the use of a *convex* structuring element B is imperative in order to satisfy property 22, i.e. in order that the operation has a physical meaning. Moreover, it is not at all necessary to introduce the notion of particle for defining morphological granulometries:

contrary to usual sieving operations, granulometries are not related to the notion of connected components and can therefore apply to continuous media.

In particular, it is possible to perform the granulometry by openings of the complement X^C of a set X (see Fig. 13). But, as explained above, performing openings on X^C amounts



a. closing of size 2 *b. closing of size 4* *c. closing of size 6*

Figure 13: Granulometry by closings of the coffee beans image.

to performing closings on X. Therefore, rather than handling two different types of granulometric transformations—the openings by the λB's of X and of X^C—, it is more interesting to generalize the notion of granulometry to two-phase images. To do so, it suffices to transform property 21 into:

$$\begin{cases} (i)\ \Phi_0 = \text{identity} \\ (ii)\ \forall(\lambda,\mu) \in \mathbb{R}^2, \lambda \geq \mu \implies \forall X \subset \mathbb{R}^2,\ \Phi_\lambda(X) \subseteq \Phi_\mu(X) \end{cases} \tag{23}$$

In the case of granulometries by openings, $(\Phi_\lambda)_{\lambda \in \mathbb{R}}$ is made of two types of transformations and we have:

$$\begin{aligned} \Phi_\lambda(X) &= (X)^{-\lambda B} \quad \text{when } \lambda < 0 \\ \Phi_0(X) &= X \\ \Phi_\lambda(X) &= (X)_{\lambda B} \quad \text{when } \lambda > 0 \end{aligned} \tag{24}$$

It is then easy to associate a measure with each $\Phi_\lambda(X)$ and to plot granulometric curves [10].

Furthermore, there corresponds to every set X a function g_X called *granulometry function.* This function associates with every point x of X the size λ of the largest opening $(X)_{\lambda B}$ of X which still contains x. g_X synthesizes all the granulometric information (See Fig. 14). The notion of granulometry extends without any restriction to the decimal case. Finally, note that among all the convex sets B available for performing granulometries, balls and their digital approximations—hexagons (denoted H, $2H$,...,nH), squares, rhombododecahedrons, etc...—are preferred.

2.4.2 Filtering

We have seen that the opening and closing operations tend to smooth the objects contours. These smoothing properties are equally true in the grey-scale case, where they are particularly useful. In fact, openings and closings constitute a particular class of *morphological filters,* a morphological filter Φ being a transformation satisfying the two following properties:

(i) Φ is increasing

(ii) Φ is idempotent

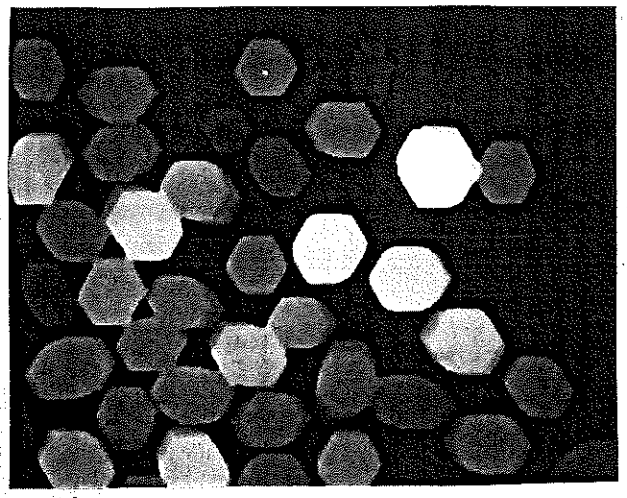


Figure 14: Granulometry function of the coffee beans.

The theoretical study of these filters allows to exhibit a certain number of transformations, mostly built as combinations of openings and closings of different sizes [26, chap. 5–10]. They are especially interesting in all problems concerning grey-scale image enhancement. However, we will not deal with these filters any longer. Their presentation would exceed the framework of these notes. We will just present an example of their use in § 4.4.

2.5 Maximal balls and skeletons

From what precedes, one can see that the opening of a set X by a disc of radius r is the set of the points of X that are scanned by this disc when it is forced to stay inside X. Moreover, when one performs a granulometry by openings (by balls $\{\lambda B\}_{\lambda \in \mathbb{R}^+}$), X is decomposed into a union of balls of increasing radii. Thus, the opened set of size λ of X—i.e. the opening of X by λB—is the part of X in which balls of radius λ can be included. Likewise, $(X)_{\mu B}$ is the part of X in which balls of radius μ can be included. However, if $\lambda < \mu$, the balls of radius λ which are also included in $(X)_{\mu B}$ are contained in at least one ball of radius μ. Hence, they are redundant for the description of X in terms of balls.

In order to better formulate this intuitive representation of X, the notion of *maximal ball* is introduced. A ball B included in X is said to be maximal if and only if there exists no other ball of X containing it:

$$\forall B' \text{ ball }, B' \subseteq X, \quad B \subseteq B' \Longrightarrow B' = B. \tag{25}$$

We can see that the redundant balls of radius λ mentioned in the preceding paragraph are precisely not maximal. It is now easy to prove that any set X can be expressed as the union of its maximal balls:

$$X = \bigcup_{B \text{ max. ball of } X} B \tag{26}$$

This representation can actually be simplified further: the centres of the maximal balls and their associated radii are sufficient. The locus of the centres of the maximal balls is called the *skeleton* of X [9]:

$$S(X) = \bigcup \{x \in X, \exists r \geq 0 \text{ such that } B(x,r) \text{ maximal in } X\}. \tag{27}$$

Fig. 15 shows a few examples of skeletons and of maximal balls. Now, to every point of the

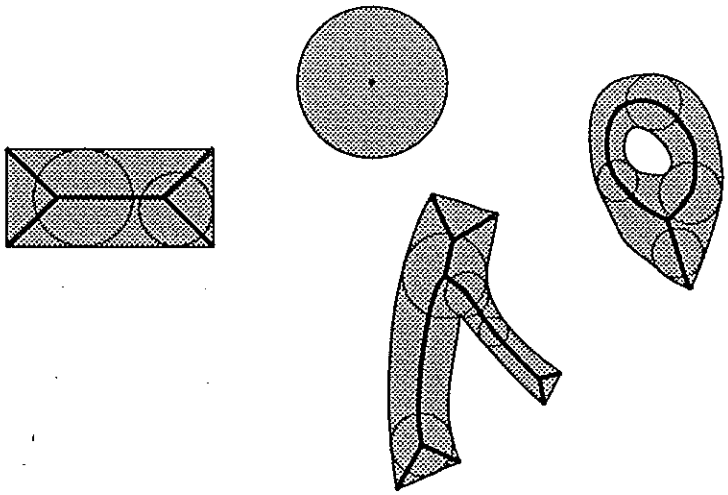


Figure 15: Some examples of skeletons and of maximal balls

skeleton, one can associate the radius of the corresponding maximal ball. A mapping q_X is thus defined on $S(X)$, which takes its values in $\mathbb{R}^+$. It is called *quench function* or *extinction function* (See Fig. 16b):

$$q_X \begin{cases} S(X) & \longrightarrow & \mathbb{R}^+ \\ x & \longmapsto & r, \quad B(x,r) \text{ maximal} \end{cases} \tag{28}$$

Moreover, an interesting result concerning the skeleton of a set X is that it can be expressed as the union of the residuals of the successive openings of X [18]. In the digital case, the formula is the following:

$$S(X) = \bigcup_{n=0}^{+\infty} [(X \ominus nB)/(X \ominus nB)_B] \tag{29}$$

Indeed, for a given $n \geq 0$, $X \ominus nB$ is the locus of the centres of the balls of size n included in X. Now, $(X \ominus nB)_B$ is exactly the locus of the centres of these balls of size n which are included in a ball of size $n+1$. Hence, they are not maximal ones. Therefore, $(X \ominus nB)/(X \ominus nB)_B$ is the locus of the centres of the maximal balls of size n of X and the union of these sets for all $n \geq 0$ yields the skeleton of X.

The quench function is particularly useful to compress binary images [19, page 49]. Besides, this representation of a set X by the pair $(S(X), q_X)$ allows us to define certain transformations by associating with q_X a new function $\gamma(q_X)$: starting from X given by its pair $(S(X), q_X)$, a new set $\gamma(X)$ can be reconstructed then, which is defined as follows:

$$\gamma(X) = \bigcup_{x \in S(X)} B(x, \gamma(q_X)(x)) \tag{30}$$

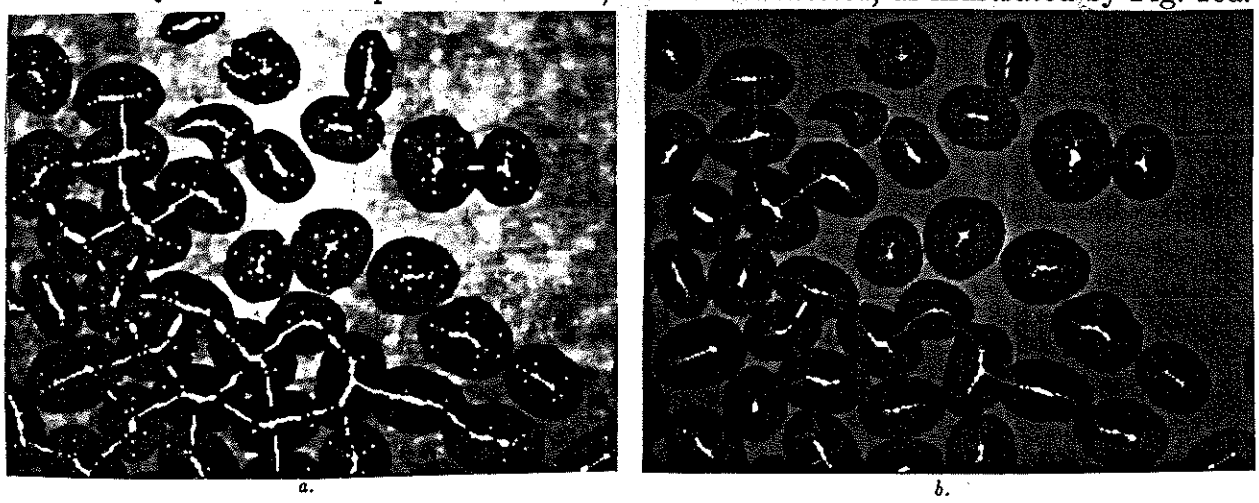
In particular, erosions, dilations and openings by balls may be obtained this way [19, pages 50–52] (See table 1). Note that the closing cannot be obtained by this means. Besides, the

Transformation	Associated function $\gamma(q_X)$
$X \ominus \lambda B$	$\gamma(q_X)(x)=-1$ if $x \leq \lambda$ $=q_X(x)-\lambda$ otherwise
$X \oplus \lambda B$	$\gamma(q_X)(x)=q_X(x)+\lambda$
$(X)_{\lambda B}$	$\gamma(q_X)(x)=-1$ if $x \leq \lambda$ $=q_X(x)$ otherwise

Table 1: How to obtain an erosion, a dilation or an opening (by balls) from the pair $(S(X), q_X)$—A ball of radius -1 stands here for $\emptyset$.

skeleton of a set $\gamma(X)$ deriving from this type of transformation has generally nothing in common with $S(X)$. Only the skeleton of the erosion of X is included in $S(X)$, but there is already no relation between $S(X)$ and $S(X \oplus \lambda B)$ any more.

Paradoxically, as efficient as this skeleton by maximal balls may be for representing sets, it is not very much used in practice. Indeed, it is *not connected,* as illustrated by Fig. 16a.



a. b.

Figure 16: Skeleton by maximal balls (a) and quench function (b) on the coffee beans

2.6 Gradient and "top-hat"

Other elementary transformations can be defined by concatenating erosions, dilations, openings and closings. In particular, in grey-scale morphology, the combination of morphological and arithmetic operators allows us to generate new transformations. Among them, two are widely used in practice: the *gradient* and the *top-hat* transformation.

2.6.1 Gradient

The morphological gradient of a function f is defined by:

$$grad(f) = \lim_{\varepsilon \to 0} \frac{(f \oplus \varepsilon B) - (f \ominus \varepsilon B)}{2\varepsilon}, \tag{31}$$

where εB designates a ball of radius ε. The preceding formula is digitized on a hexagonal grid in the following way:

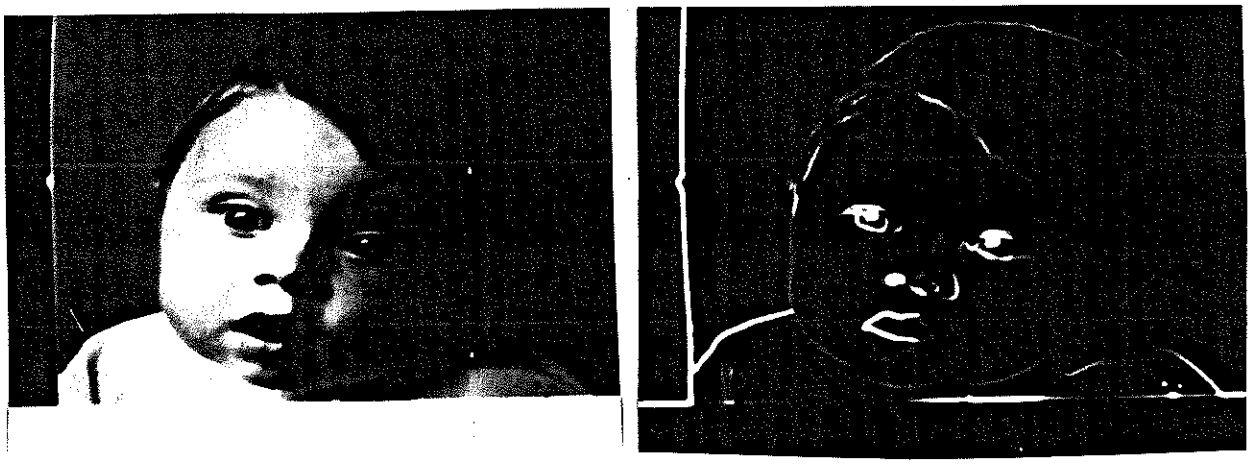
$$grad(f) = \frac{(f \oplus H) - (f \ominus H)}{2}, \tag{32}$$

with H standing for the elementary hexagon previously introduced [2]. One can show that the morphological gradient of f is equal to the modulus of the gradient of f—supposed to be continuously differentiable—classically defined as:

$$|\overrightarrow{grad}(f)| = \left[(\frac{\partial f}{\partial x})^2 + (\frac{\partial f}{\partial y})^2\right]^{\frac{1}{2}} \tag{33}$$

There exist numerous variations of this transformation. In particular, one can define in terms of morphological operations *directional gradients* and represent the azimuths of the gradient vectors.

Fig. 17 illustrates this transformation. Note that the high values of the gradient's modulus correspond to the highly contrasted areas of the image. This property will be used in the following for detecting contours of grey-tone images (See § 4.3).



a. original grey-scale image *b. gradient*

Figure 17: Morphological gradient

2.6.2 Top-hat transformation

The *top-hat* [19, page 101] is a morphological transformation which enables to extract the bright and narrow zones of a grey-tone image. These areas are those which disappear after a decimal opening. So, given a structuring element B, the top-hat transform $HdF_B(f)$ of a function f is defined by:

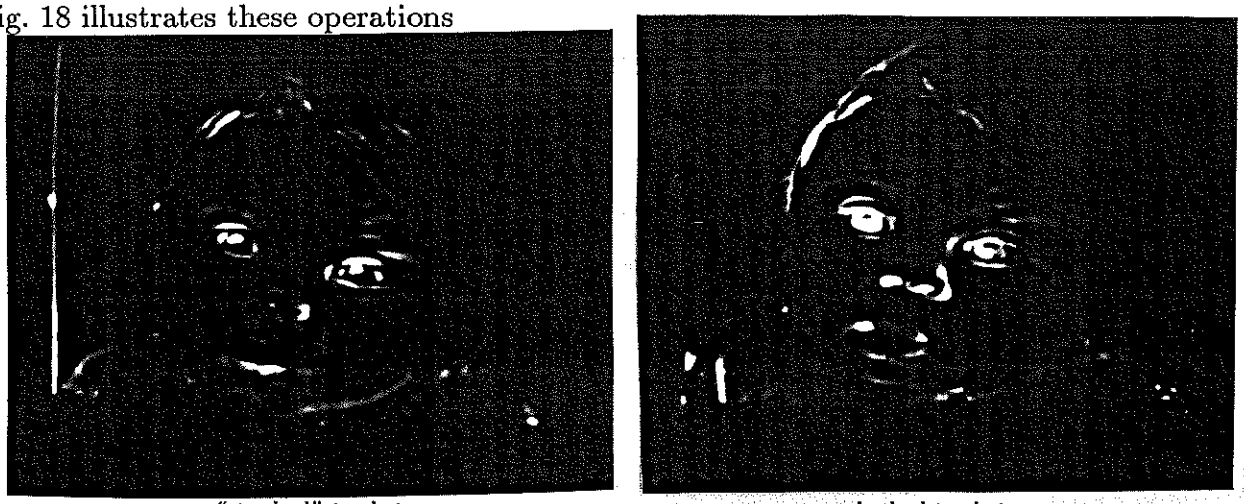
$$HdF_B(f) = f - (f)_B. \tag{34}$$

This transformation extracts the residuals of the opening of f, the "size" of these residuals being smaller than that of the structuring element B. Here again, B is preferably chosen to be a convex set (disc, segment,...). In particular, it is often judicious to combine different homothetics of the same convex set, or to use linear structuring elements with different orientations so as to precisely select the regions to be extracted.

Similarly, the dark and narrow areas of an image can be extracted by means of the dual top-hat transformation HdF^*_B, which is defined by:

$$HdF^*_B(f) = (f)^B - f \tag{35}$$

Fig. 18 illustrates these operations



a. "standard" top-hat *b. dual top-hat*

Figure 18: Hexagonal top-hat transforms of image 17a

Quite often, the regions extracted by these transformations are also filtered according to their contrast by thresholding the residuals. For example, thresholding at level h a top-hat performed with a ball of radius λ allows us to extract from an image its crests higher than h and less wide than 2λ (See Fig. 19).

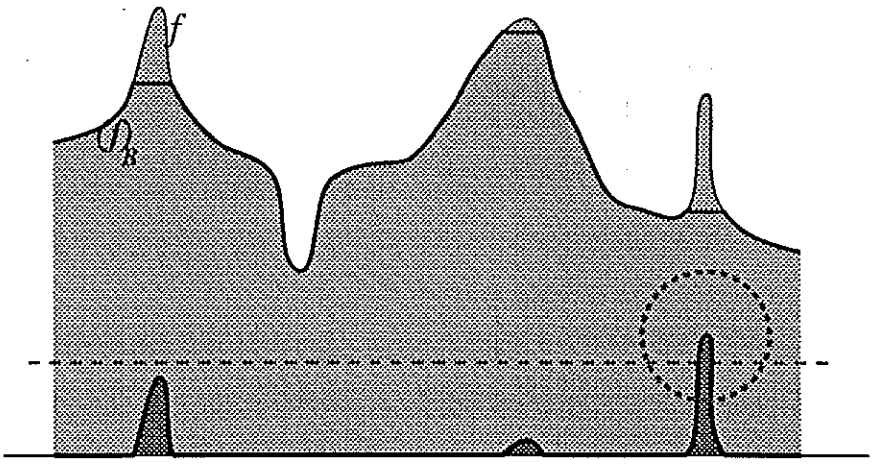


Figure 19: "Crests" extracted by a top-hat followed by a thresholding.

2.7 Extrema of a function

The notion of the *extrema* of a function will soon be very useful and should be introduced right now. Let f: $\mathbb{R}^2 \longrightarrow \mathbb{R}$ be a function. The graph of f may be considered as a topographic

surface. A *maximum* of f is then a summit of this surface, i.e. an area (connected, but not necessarily reduced to a single point) from which it is impossible to join a higher point by a never descending path (See figure 20).

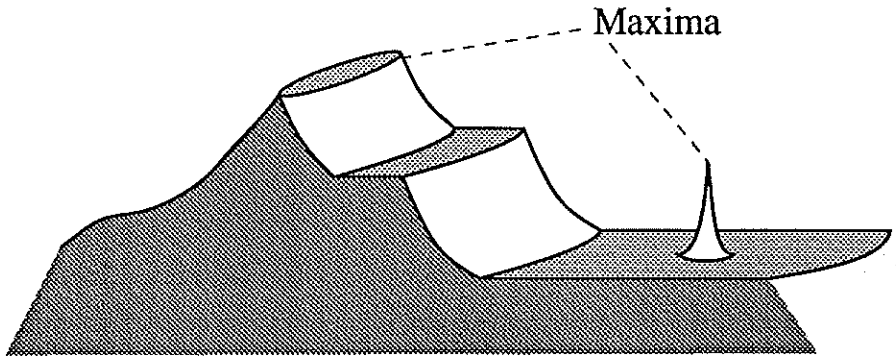


Figure 20: Maxima of a function

Let us consider in particular the different thresholds of f. A maximum of this function at the altitude h is a connected component of the threshold $S_h(f) = \{x \in \mathbb{R}^2, f(x) \geq h\}$ which contains no connected component of any threshold $S_{h'}$ where $h' > h$.

2.8 First steps toward segmentation: marking the objets

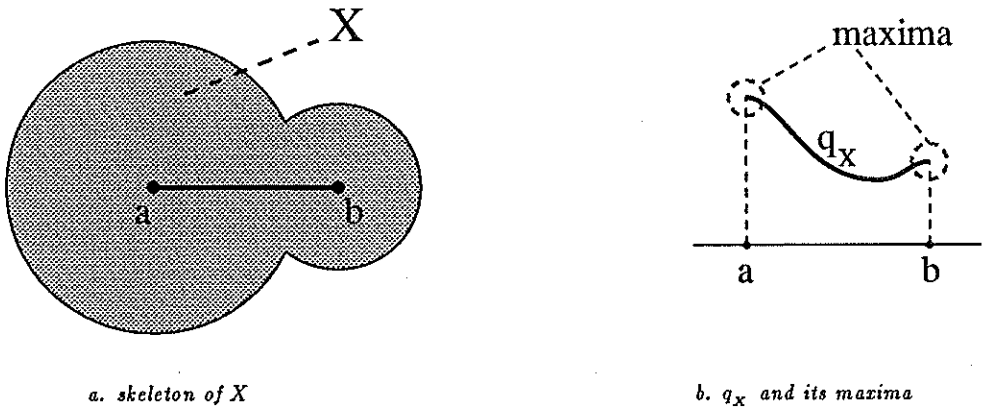
The notions we have just introduced will now help us solving a typical segmentation problem in the binary case: the separation of balls or objects that partially overlap. This kind of situation is frequently encountered in numerous fields of application of image processing, like the distribution of marbles in agglomerate materials [10], the separation of cells in a smear [19], the counting of particles,... not to mention our coffee beans image. Solving this problem first requires a correct marking of the components to be separated. This approach is in fact quite natural: someone who would be asked to point with his finger at the different objects present in the image would do exactly the same.

Let us first consider the very simple case where the set X under study is made of two circular components (See Fig. 21a). As seen above, X can be represented by its skeleton $S(X)$ and its quench function q_X. In order to bring its two components to the fore, it seems quite natural to look for the regions of X containing the largest maximal balls. We thus search on $S(X)$ the points whose associated maximal balls are the "largest" ones: these points are the *maxima* of the quench function q_X (See Fig. 21b).

Unfortunately, it is rather delicate in practice to determine the maxima of a function which is defined on a non-connected support. This is the reason why we need introducing a new function $dist_X$, defined on X and called *distance function:*

$$dist_X \begin{cases} X & \longrightarrow & \mathbb{R}^+ \\ x & \longmapsto & d(x, X^C) \end{cases} \tag{36}$$

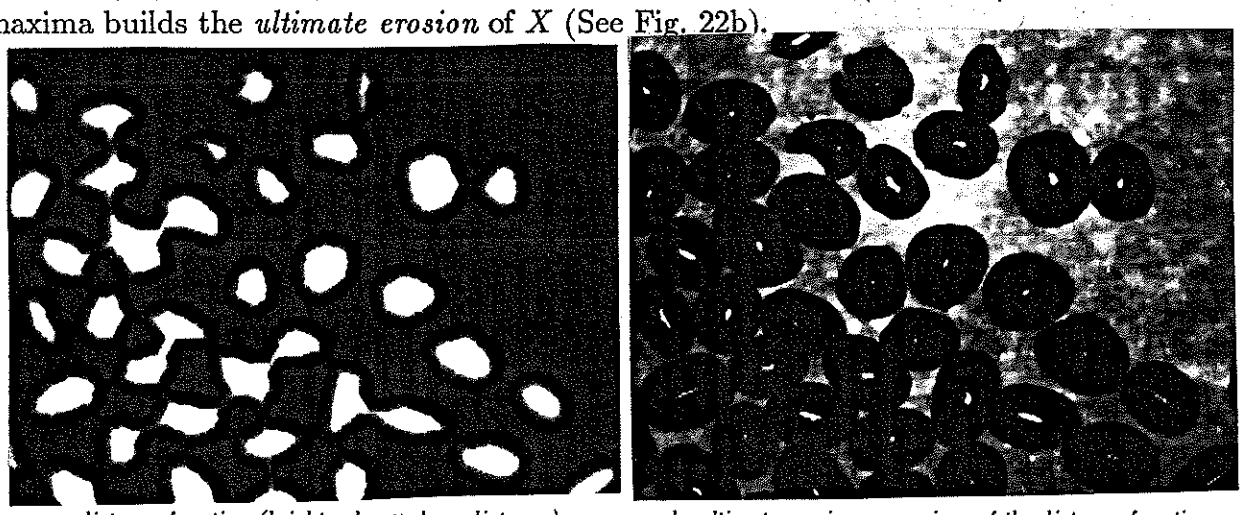
With every point of X, $dist_X$ associates its distance to the nearest point in the complement X^C (See Fig. 22). The different contour lines of this function correspond to the boundaries of



a. skeleton of X *b. q_X and its maxima*

Figure 21: Skeleton and maxima of the quench function

the sucessive erosions of X. One can show that the maxima of $dist_X$ are exactly the same as those of the quench function q_X. Besides, in the discrete case, these maxima can be detected as follows: let X_n designate the erosion of size n of X. X_n is the threshold of the distance function between levels $n+1$ and $+\infty$. Let us now perform an elementary erosion—of size one—of this eroded set. Any connected component C of X_n which disappears after this erosion is, by definition, a maximum of the distance function (See § 2.7). The set of these maxima builds the *ultimate erosion* of X (See Fig. 22b).



a. distance function (bright color ⇔ long distance) *b. ultimate erosion = maxima of the distance function*

Figure 22: Distance function and ultimate erosion

The example of the coffee beans presented on Fig. 22 shows that the ultimate erosion is indeed a marker in the sense indicated above. Each component of the initial set—the coffee beans—is marked through this procedure. Moreover, this marking is a one-to-one correspondance: up to the parity errors due to digitization, to each coffee bean corresponds one and only one marker. Besides, we have seen that these markers can be defined as the extrema of a function, namely the distance function.

In the following, we will try to show that this approach is quite general in mathematical morphology. Any segmentation obeys to one—or several—marking function(s). The detection of the extrema of this (these) function(s) constitutes the first step of the segmentation process: the generation of markers of the objects to be segmented. We shall see that marking functions are also used in the second step of the segmentation process (See § 4).

3 Homotopic and geodesic transformations

Now that we have the markers, let us proceed to the actual segmentation step, that is to the extraction of the marked objects. In so doing, the use of extensive operators will be required. Moreover, to preserve the marking, these operators should satisfy the two following criteria:

1. Not to connect markers, preserve their number and the neighborhood relationships between them. Such morphological operators do exist and are designated as *homotopic transformations.*

2. To work only on the set X to be segmented. Indeed, it is useless that the operator extends beyond X, since the points of X^C are of no interest. This category of morphological operators is formalized through the notion of *geodesy* and of *geodesic transformations.*

3.1 Homotopy and skeletons

Let $X \subset \mathbb{R}^2$ be a connected set and x and y two points of X. Consider two different paths C_1 and C_2 joining x and y. C_1 and C_2 are said to be *homotopic* if and only if it is possible to go from C_1 to C_2 by a continuous sequence of continuous and reversible deformations. Take for example Fig. 23: the paths C_2 and C_3 are homotopic whereas this is not the case for C_1 and C_2, since the hole in the set makes it impossible to map C_2 onto C_1 by continuous deformations. One can show that the homotopy relationship between paths is a relation of equivalence.

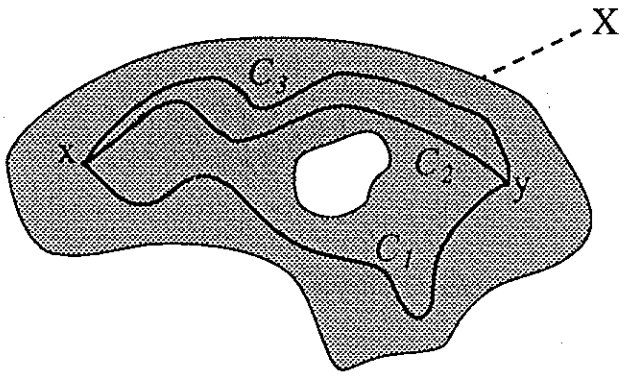


Figure 23: The pair of paths (C_2, C_3) is homotopic whereas the pair (C_1, C_2) is not

S is said to be *simply connected* if and only if there exists one and only one class of path equivalence between any two points of X. This comes down to saying that X is connected and without any hole.

A transformation Φ: $\mathcal{P}(\mathbb{R}^2) \longrightarrow \mathcal{P}(\mathbb{R}^2)$ is said to be homotopic if and only if:

$$\forall X \subset \mathbb{R}^2, \quad \exists \Psi : \mathbb{R}^2 \times [O;1] \longrightarrow \mathbb{R}^2, \text{ continuous mapping,}$$
$$\text{such that} \quad \begin{cases} \Psi(X,0) = X, \\ \Psi(X,1) = \Phi(X). \end{cases} \tag{37}$$

To use an image, an homotopic transformation transforms a set X into a set Y through a continuous sequence of continuous transformations. Thus, a simply connected set is transformed into a simply connected set, a set with one hole into a set with one hole, etc... More generally, if $X \subset \mathbb{R}^2$ is made of several connected components $(X_i)_{i\in[1,n]}$, the number and the relative positions of these X_i are preserved (an homotopic transformation preserves the path classes for both X and X^C. See Fig. 24).

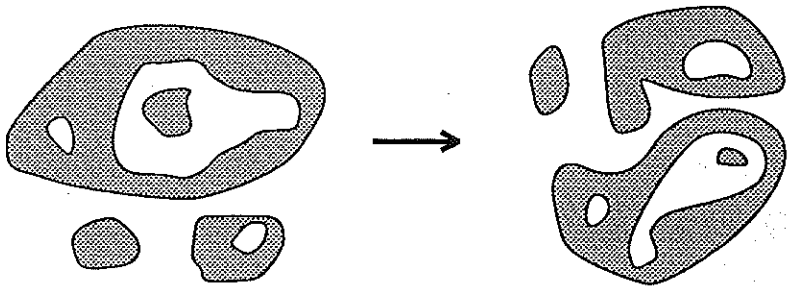


Figure 24: An example of a homotopic transformation

Classical examples of homotopic transformations in MM are given by the skeletons which are said to be *connected* [19, page 75]. They are obtained by a sequence of thinnings (for the classical skeleton) or a sequence of thickenings (for the *exoskeleton*, i.e. the skeleton of the background) using structuring elements defined on the "elementary ball"—hexagon on the hexagonal grid, square on the square grid. These structuring elements are selected in order to produce homotopic thinnings and thickenings. Thus, on the hexagonal grid, we can show that there exist three structuring elements (up to the six rotations) having this property. These elements are called L, M and D [12] (See Fig. 25). When using them, we see that

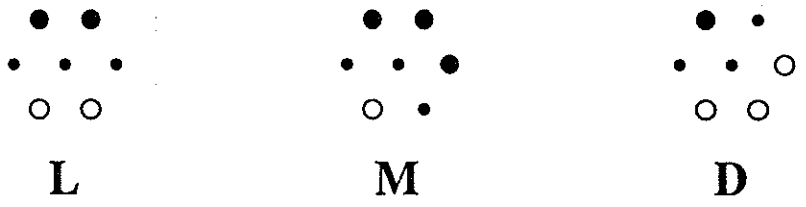
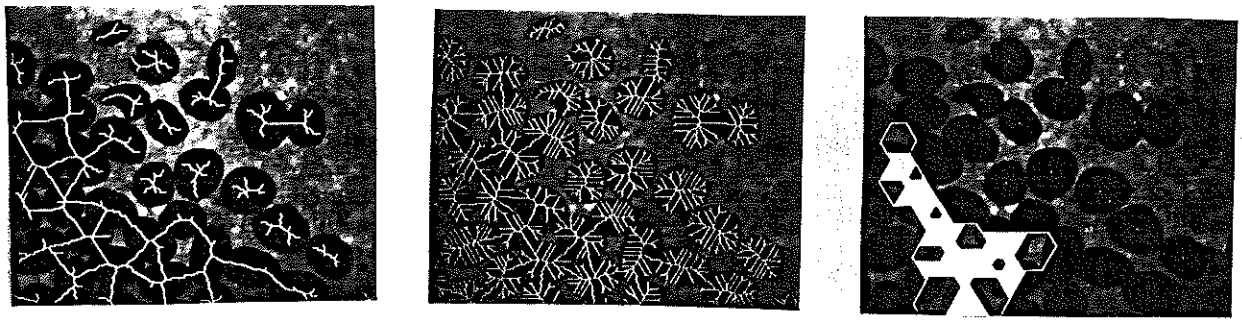


Figure 25: Structuring elements used for constructing connected skeletons

only L and M allow to derive skeleton-like transformations, i.e. a sort of medial axis of the objects (See Fig. 26). As a matter of fact, it is possible to show that there exist very close relationships between these skeletons of type L and M and the skeleton by maximal balls described in § 2.5 [8].

One can also show that the connected skeleton equals the *crest lines* of the distance function and that it can be obtained as such [26, chap. 13] (See Fig. 27). Moreover, this approach allows to extend the notion of skeleton to the case of grey-scale images [20]. Fig. 28 displays an image of chromosomes and the corresponding decimal skeleton (For this image, we actually performed the skeleton of the inverted image, i.e. we followed the "bottom of the valleys".).



a. L-skeleton *b. M-skeleton* *c. D-skeleton*

Figure 26: Skeletons resulting from successive thinnings with the structuring elements L, M and D

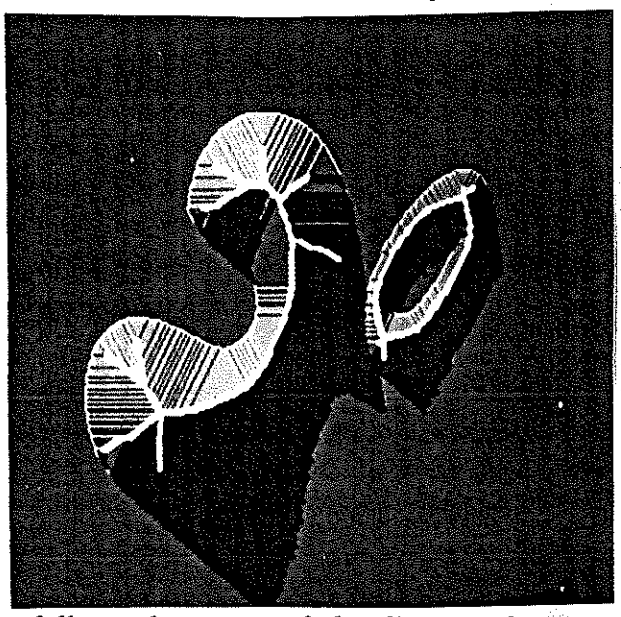
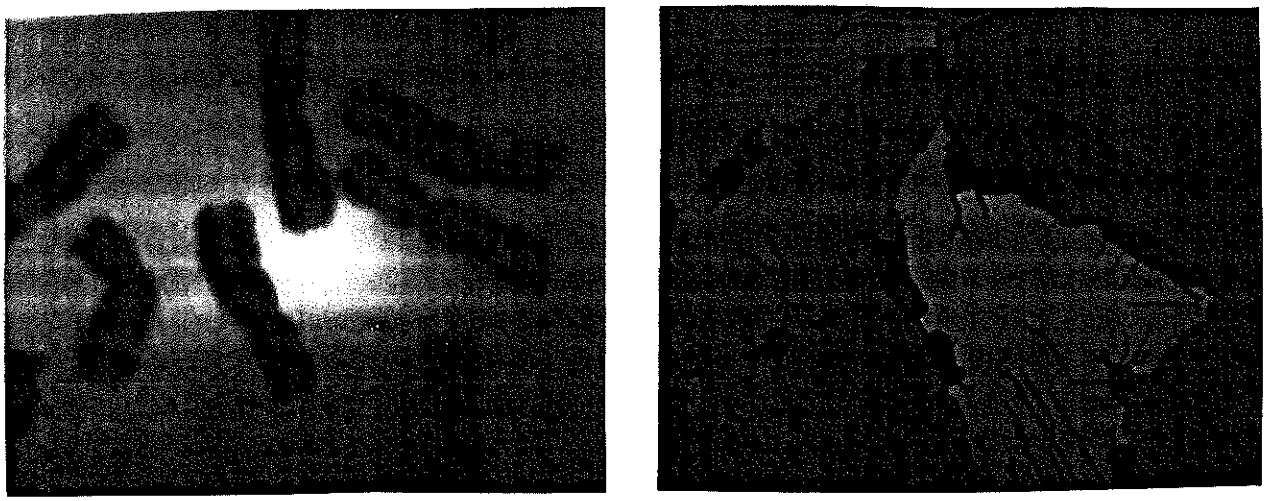


Figure 27: The skeleton follows the crests of the distance function (artificially shaded here [30])



a. image of chromosomes *b. decimal skeleton*

Figure 28: decimal skeleton

As concerns the skeleton by thickenings, it is mainly used to produce the *SKeleton by Influence Zones* or *SKIZ*. The SKIZ is defined as follows: let $X \subset \mathbb{R}^2$ be a set made of n connected components $\{X_i\}_{i\in[1,n]}$. The *influence zone* $z(X_i)$ of X_i is defined as the set of the points of $\mathbb{R}^2$ that are closer to X_i than to any other connected component of X:

$$z(X_i) = \{x \in \mathbb{R}^2, \forall j \neq i, d(x, X_i) < d(x, X_j)\} \tag{38}$$

Remark that the influence zone of a connected component X_i of X has generally not the same homotopy as X_i itself (e.g. on Fig. 29, we see that X_i may have holes whereas $z(X_i)$ does not.). The set of the boundaries of the different influence zones is by definition the skeleton by influence zone of X [14]. Fig. 29 shows an example of SKIZ.

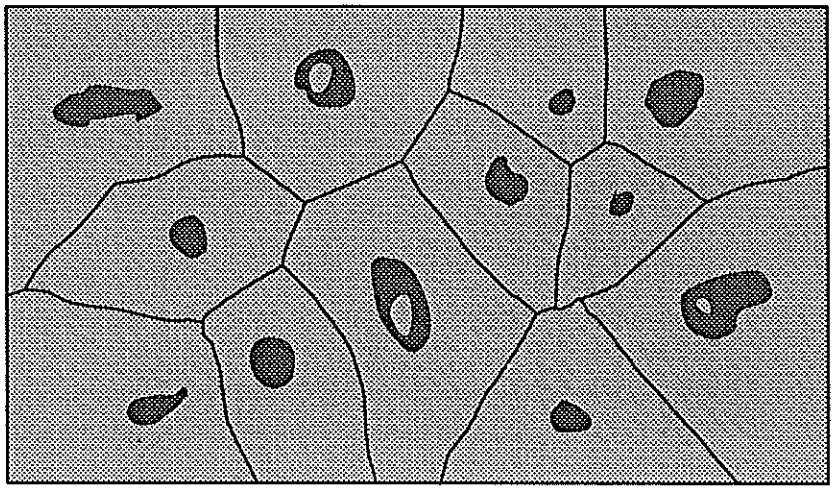


Figure 29: An example of a skeleton by influence zones

The SKIZ is usually obtained by keeping only the closed arcs of the skeleton by thickening of X (i.e. its exo-skeleton). In order to suppress the non-closed arcs of $S(X^C)$, a particular kind of thinning is used, which is called a *pruning*. Let us stress that the set resulting from this step may, in some special cases, still be a subset of the true SKIZ (See Fig. 30).

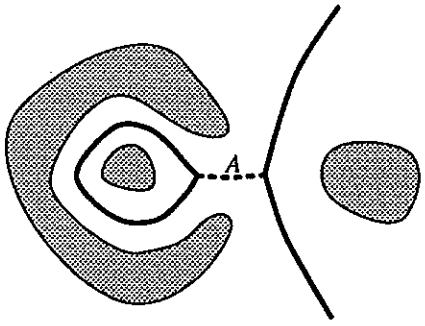


Figure 30: A trap in the construction of the SKIZ: arc A does not belong to the true SKIZ

3.2 Geodesics

3.2.1 Geodesic distance

The concept of path has already been used for defining the homotopic transformations. It will again be useful here to define a distance associated with a set X of $\mathbb{R}^2$. Let $(x, y) \in X^2$. The geodesic distance $d_X(x, y)$ between x and y is the lower bound of the length of the paths between x and y in X (if any):

$$d_X(x, y) = \inf\{l(C_{x,y}), C_{x,y} \text{ path linking } x \text{ and } y \text{ in } X\}. \tag{39}$$

When x and y belong to different connected components of X, we conventionally put $d_X(x, y) = +\infty$. According to this remark, d_X is a "generalized" distance (See Fig. 31).

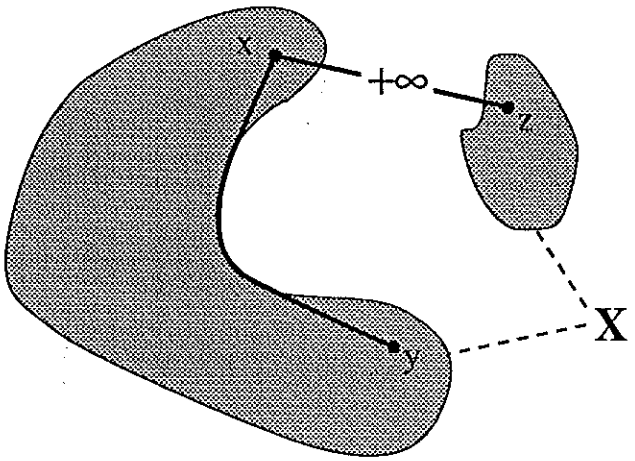


Figure 31: Geodesic distance

We call *geodesic ball* of radius r centered in $x \in X$ the set $B_X(x, r)$ defined by:

$$B_X(x, r) = \{y \in X, d_X(x, y) < r\}. \tag{40}$$

3.2.2 Geodesic transformations

Let now Y be a subset of X equipped with the geodesic distance d_X induced by X. The erosion $E_X^{(r)}(Y)$ and the dilation $D_X^{(r)}(Y)$ of Y by geodesic balls of a given radius r are defined this way [15,16]:

$$E_X^{(r)}(Y) = \{y \in X, \ B_X(y, r) \subseteq Y\} \tag{41}$$

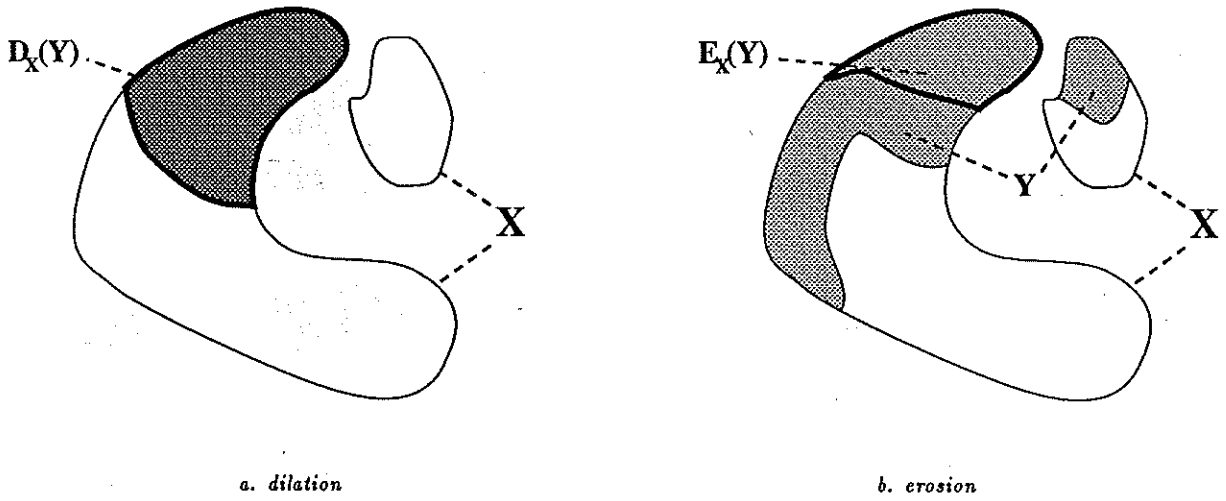
$$D_X^{(r)}(Y) = \{y \in X, \ B_X(y, r) \cap Y \neq \emptyset\} \tag{42}$$

These definitions are quite similar to the classical Euclidean definitions of the erosion and the dilation by balls. The differences come from two main factors:

- X has become the working space (instead of $\mathbb{R}^2$). In other words, the result of these transformations is always contained in X.

- The structuring elements that are used are defined by means of a distance and do not act as "rigid" structuring elements as in the Euclidean case. The range of available structuring elements is here very poor.

Fig. 32 shows an example of geodesic erosion and dilation.



a. dilation *b. erosion*

Figure 32: Basic geodesic operations

In practice, the geodesic dilation of size n is performed on a digital image by iterating n geodesic dilations of size 1. The latter is obtained by simply intersecting the Euclidean dilation of size 1 of Y with the "space-set" X:

$$D_X^{(1)}(Y) = (Y \oplus H) \cap X \tag{43}$$

$$D_X^{(n)}(Y) = \underbrace{D_X^{(1)}(D_X^{(1)}(\ldots D_X^{(1)}}_{n \text{ times}}(Y))) \tag{44}$$

The erosion $E_X^{(n)}(Y)$ is then easily obtained by duality. By analogy, the same formulae immediately extend to functions. A function $g : \mathbb{R}^2 \longrightarrow \mathbb{R}$ can thus be dilated geodesically with respect to a majoring function $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ as follows:

$$D_f^{(1)}(g) = \inf(g \oplus H, f) \tag{45}$$

$$D_f^{(n)}(g) = \underbrace{D_f^{(1)}(D_f^{(1)}(\ldots D_f^{(1)}}_{n \text{ times}}(g))) \tag{46}$$

Remark that it is difficult here to exhibit the underlying geodesic distance. The geodesic digital erosion of f with respect to g (minoring function) is defined in a similar way:

$$E_g^{(1)}(f) = \sup(f \ominus H, g) \tag{47}$$

$$E_g^{(n)}(f) = \underbrace{E_g^{(1)}(E_g^{(1)}(\ldots E_g^{(1)}}_{n \text{ times}}(f))) \tag{48}$$

Consider now a set Y included in the connected component X_i of X. If Y is dilated geodesically in X, we are sure not to go beyond X_i, whatever the size of the dilation. Moreover, if Y is dilated enough, we obtain the entire X_i, i.e. the connected component of X which is marked by Y. Besides, this extends to the case where Y marks several connected components of X. The transformation we have just brought to the fore is called *reconstruction*. It is denoted R_X and is defined in a more formal manner as follows:

$$R_X(Y) = D_X^{(+\infty)}(Y) = \lim_{r \to +\infty} D_X^{(r)}(Y). \tag{49}$$

Fig.33 displays an example of a binary reconstruction. Obviously, the reconstruction will

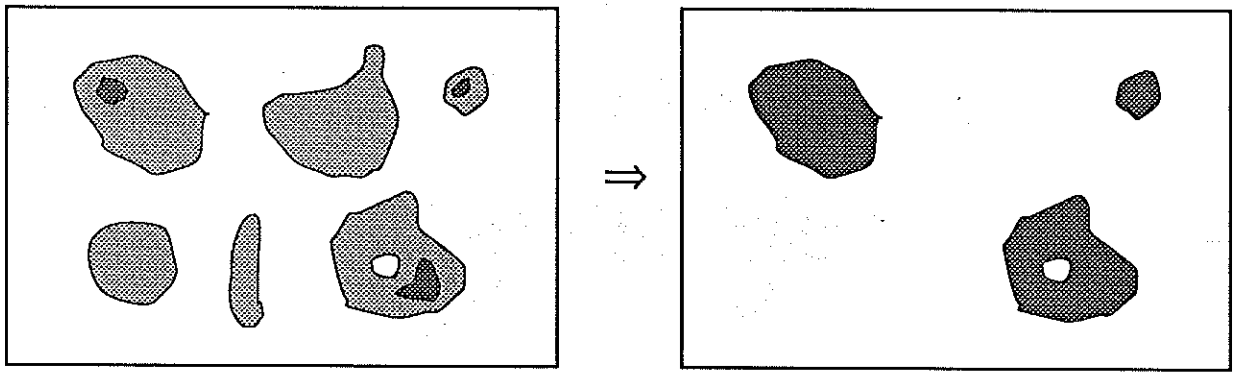


Figure 33: Binary reconstruction from markers

be very useful for solving segmentation problems, since, as we have seen, it allows us to extract all the connected components marked by a *marking set* Y. This tool is however insufficient: the result of the reconstruction is indeed unchanged whatever the number of markers per connected component.

In the decimal case, a function f can also be reconstructed from a "marking-function" g, majoring or minoring f. When e.g. $g \leq f$, we set:

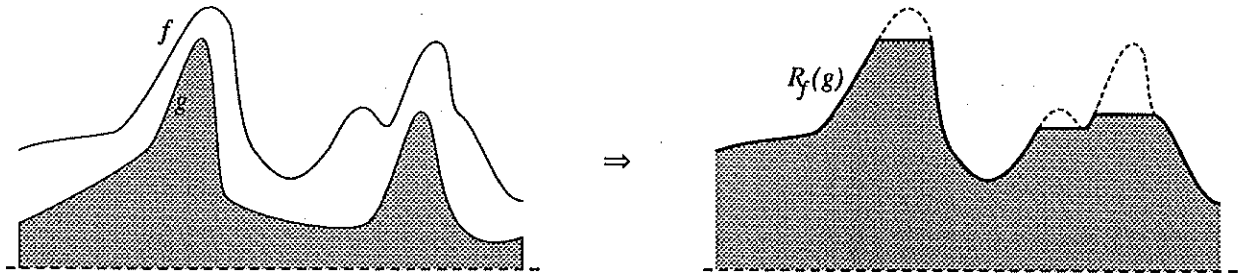
$$R_f(g) = D_f^{(+\infty)}(g). \tag{50}$$

Through this operation, only the "domes" of f marked by g can be (partially or totally) reconstructed (See Fig. 34). On the other hand, when $g \geq f$, we set:

$$R_f^*(g) = E_f^{(+\infty)}(g) \tag{51}$$

and g is now a marker of the "basins" of f which are left after this reconstruction operation.

In order to exhibit powerful separation tools, we have now to add the previous notion of homotopy to the geodesic operations. This is done quite simply: we can for example define geodesic skeletons by a similar approach to that which is used in the Euclidean case [19, page 43]. The notion of maximal ball can thus be transposed to the geodesic case. Likewise, in the definition of the skeleton by influence zones, the Euclidean distance can perfectly be replaced by a geodesic one. Thus, we define the *geodesic SKIZ* of a set Y in a set X (see Fig. 35). Notice in particular that, if Y is made of n connected components, X will be segmented in n parts.



a. function f and marking function g *b. $R_f(g)$ = reconstructed function f*

Figure 34: The decimal reconstruction R_f

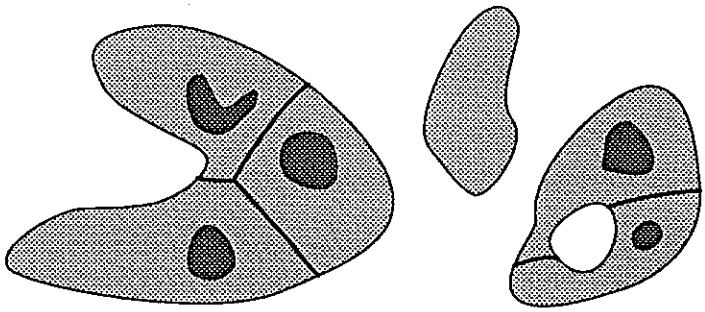
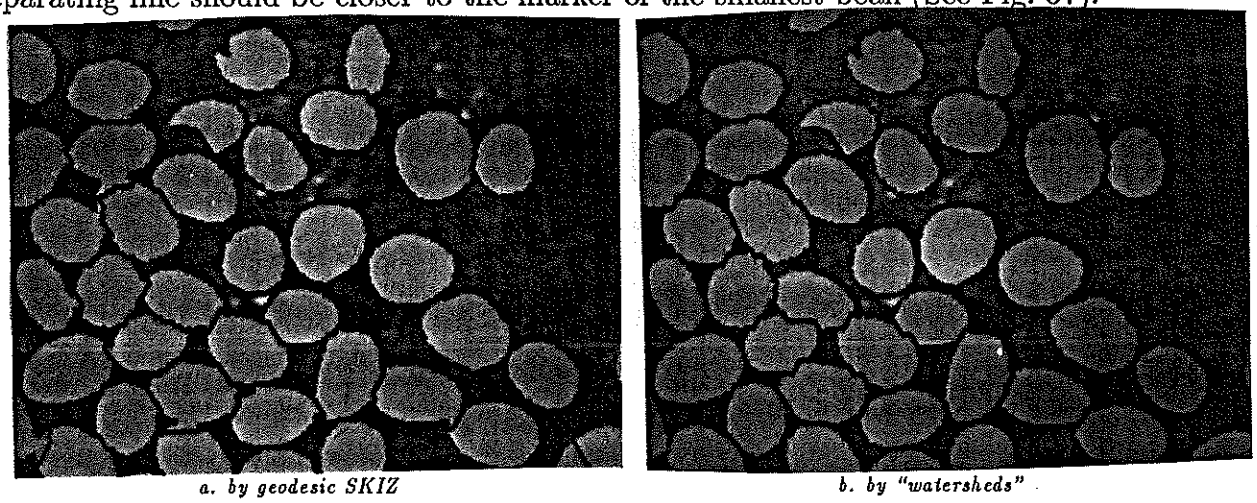


Figure 35: geodesic SKIZ

4 From marking to segmentation

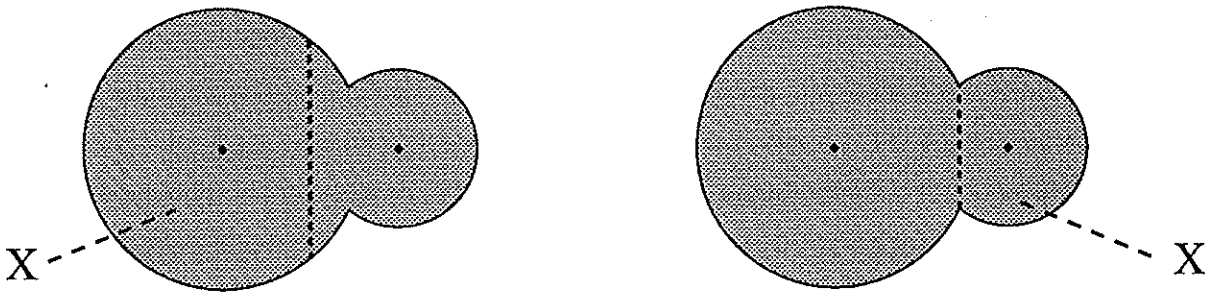
4.1 A binary segmentation algorithm

Let us come back to our problem of beans segmentation. In a first step, they have been marked by ultimate erosion. Now, we can consider using the geodesic skeleton by influence zones described in § 3.2 to separate the beans. Each of them would be defined as the geodesic influence zone of its marker (See Fig. 36a). However, the result is not satisfactory, because the separating lines exhibited between the beans are not well positioned. This is due to the fact that, by construction, each separating line tends to be placed mid-way from the markers. Yet, we know that these markers do not correspond to the same erosion size (See § 2.8) in the ultimate erosion algorithm. When two adjacent beans are of very different size, their separating line should be closer to the marker of the smallest bean (See Fig. 37).



a. by geodesic SKIZ *b. by "watersheds"*

Figure 36: Segmentation of coffee beans



a. by geodesic SKIZ: wrong result *b. desired result*

Figure 37: Separation of overlapping balls

To solve this problem, one has to take into account the values associated with the connected components of the ultimate erosion by the quench function q_X. This can be done by using the following procedure, presented here in the digital case: let $n_{\max}$ be the size of the largest non empty erosion of X:

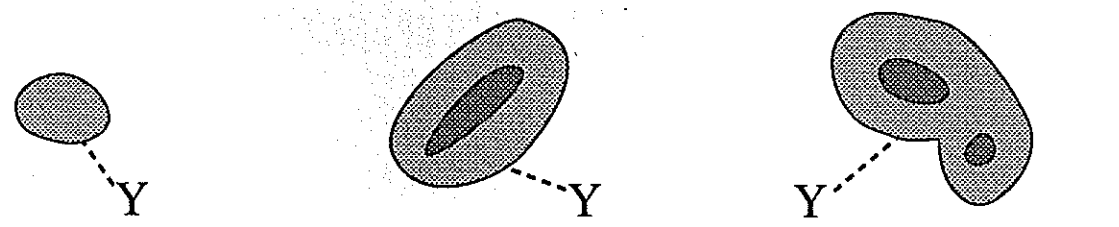
$$X \ominus n_{\max} H \neq \emptyset \quad \text{and} \quad X \ominus (n_{\max} + 1)H = \emptyset. \tag{52}$$

$X \ominus n_{\max} H$ is necessarily a subset of the ultimate eroded set. Let $X_{n_{\max}}$ designates this set. Consider now the erosion of the set X of a size immediately inferior to the preceding, that is $X \ominus (n_{\max} - 1)H$. The following inclusion relation holds:

$$X_{n_{\max}} \subseteq X \ominus (n_{\max} - 1)H. \tag{53}$$

Let Y be a connected component of $X \ominus (n_{\max} - 1)H$. There are several possible relations of inclusion between Y and $Y \cap X_{n_{\max}}$ (See Fig. 38):

1. $Y \cap X_{n_{\max}} = \emptyset$: in this case, Y is a new connected component of the ultimate erosion of X.
2. $Y \cap X_{n_{\max}} \neq \emptyset$ and is connected: Y can then be used as a new marker.
3. $Y \cap X_{n_{\max}} \neq \emptyset$ and is not connected: the new markers that are used in this case are the geodesic influence zones in Y of the connected components of $Y \cap X_{n_{\max}}$.



a. $Y \cap X_{n_{\max}} = \emptyset$ *b.* $Y \cap X_{n_{\max}} \neq \emptyset$, *connected* *c.* $Y \cap X_{n_{\max}} \neq \emptyset$, *not connected*

Figure 38: The three possible relations of inclusion between Y and $Y \cap X_{n_{\max}}$

Denote $X_{n_{\max}-1}$ the set of markers exhibited at level $n_{\max} - 1$. What applies to one connected component applies to all, hence $X_{n_{\max}-1}$ is made up of the geodesic influence zones of $X_{n_{\max}}$ in $X \ominus (n_{\max} - 1)H$ to which are added the connected components of the ultimate erosion at level $n_{\max} - 1$ (See Fig. 39).

This reconstruction-based procedure can be iterated at levels $n_{\max} - 2$, $n_{\max} - 3$, etc... down to level 0. In a more formal manner, let $n \in]0; n_{\max}]$. For any $X \subset \mathbb{R}^2$, we introduce the following notations:

- $u_n(X)$ = set of the connected components of the ultimate erosion of X at level n:

$$x \in u_n(X) \iff x \in S(X) \text{ and } q_X(x) = n. \tag{54}$$

- $\forall\, Y \subseteq X$, $z_X(Y)$ = set of the geodesic influence zones of the connected components of Y in X.

The recursion formula between levels n and $n-1$ is then written:

$$X_{n-1} = z_{X \ominus (n-1)H}(X_n) \cup u_{n-1}(X) \tag{55}$$

as illustrated by Fig. 39. The set X_0 derived from this procedure provides a much better segmentation than the rough geodesic SKIZ (See Fig. 36b). Indeed, the different connected components of the ultimate eroded set intervene here in an order which depends on the value they are associated with by the quench function.

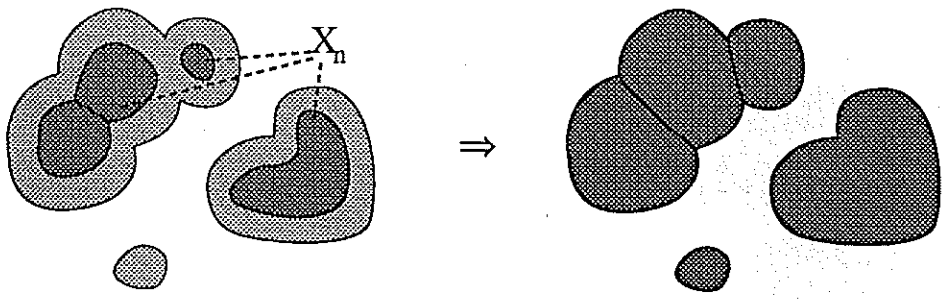


Figure 39: Recursion from X_n to X_{n-1}

4.2 Watersheds

The preceding algorithm may seem a bit complex. Therefore, we are now going to give it a more intuitive interpretation by using once again the distance function $dist_X$ associated with X. For the sake of consistency, we consider the function $f = -dist_X$ and regard it as a topographic surface. The connected components of the ultimate erosion of X exactly equal the minima of f.

Suppose that we pearce this topographic surface at the location of each minimum. Then, plunge it slowly into a lake. The water first gets through the holes located at the deepest minima and gradually floods the whole surface. To complete the experiment, an observer is assigned the mission of erecting a dam in any point where waters coming from two disjoined minima could melt. At the end of the plunging procedure, when the surface is completely immerged, the finally built dam constitutes the *dividing lines* of the waters coming from the different minima (See Fig. 40). These dividing lines are called the *watersheds* of f. The different pools separated by the watersheds cover particular zones of f designated as the *catchment basins* associated with each minimum.

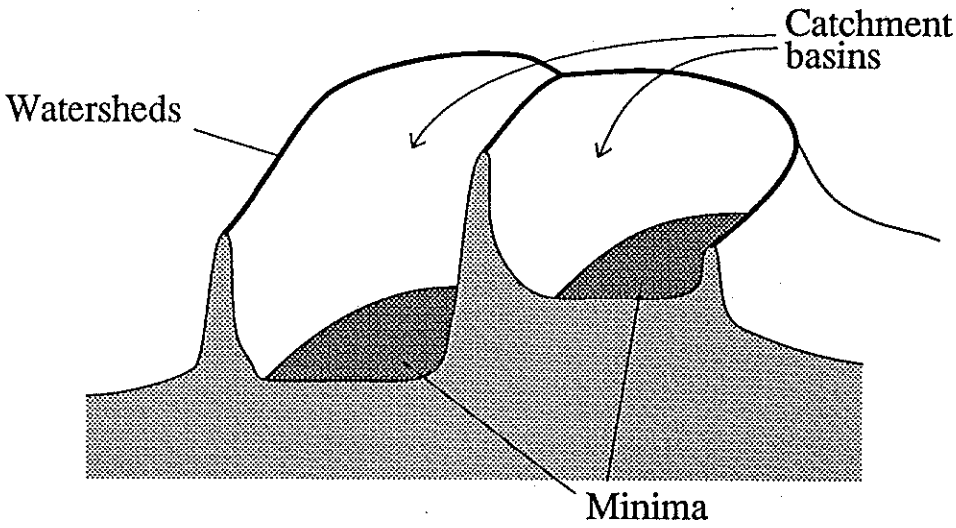


Figure 40: Watersheds and catchment basins

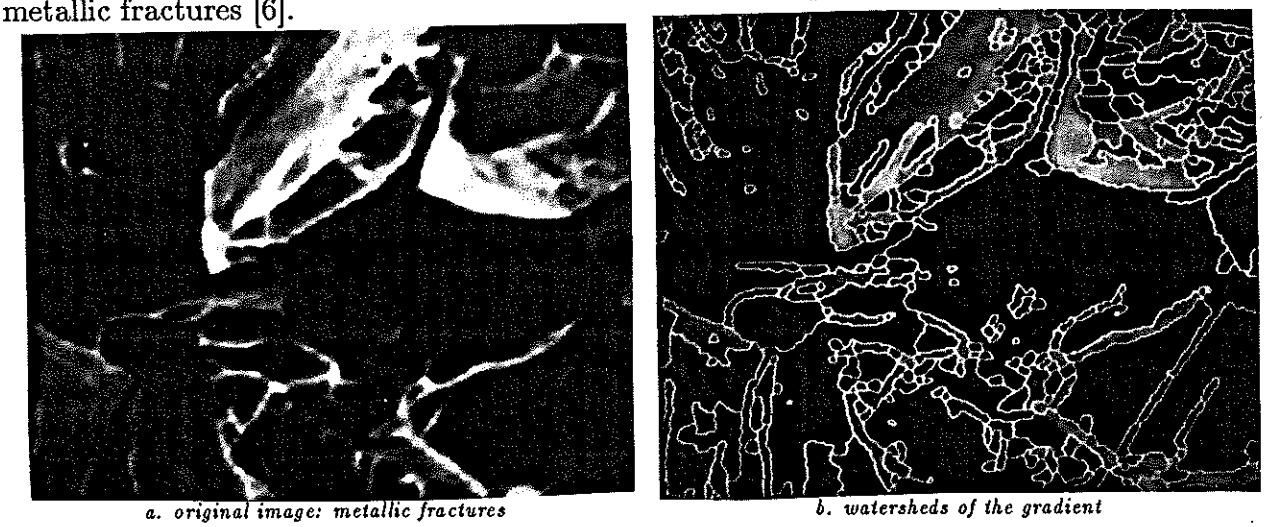
In examining the pools obtained at each step of the immersion, we can see at once that their set exactly corresponds to one of the X_n previously used (See § 4.1). In other words, the plunging procedure is identical to the separation algorithm designed above, and the watersheds equal the separating lines defined by this algorithm.

More generally, the watersheds can be defined for any decimal function $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ (Refer to [3,7]). If we consider the graph of f as a topographic surface and detect its minima, nothing prevents us from applying the same plunging procedure. Hence, there corresponds one catchment basin to each minimum of f. Besides, the used procedure can be transposed for computing the watersheds of f. To do so, it suffices to replace the erosions $X \ominus nH$ by the successive thresholds of f.

4.3 Watersheds and segmentation of decimal images

The notion of watersheds will now serve as a guideline for the segmentation of grey-scale images. In this case, the segmentation is most of the time related to a contour detection problem. However, the notion of contour is not always clearly defined and here again, when one is asked to specify which objects to segment, he may be able to point them out without being able to draw their precise contours. In other words, in a grey-scale image, a shape is most of the time characterized by the fact that, to it, corresponds a homogeneous grey-tone area rather than a clearly contrasted and delimited contour. Therefore, the homogeneous zones are once again the markers of the objects to extract.

To say that an area of a function is rather homogeneous means that there are only slight variations of the grey levels inside this area, or else that it has a *low gradient.* Our objects markers may then be taken as the *minima* of the gradient's modulus of f (See § 2.7). Hence, delimiting the objects consists in computing the watersheds of the gradient. Each object represent indeed a "hollow zone", i.e. a catchment basin of the gradient function. Fig. 41 shows an example of contour detection by watersheds of the gradient on a picture representing metallic fractures [6].



a. original image: metallic fractures *b. watersheds of the gradient*

Figure 41: Contour detection by watersheds of the gradient

Once again, we see that the minima of a function—the gradient—serve as the markers of the zones to be segmented. The watersheds then allow us to extract the actual contours of these zones. This approach is quite similar to that which is used in the binary case and

consists in defining a so-called *marking function.* In the segmentation process, this function provides both the markers and the contours. The only problem is thus to define a *good* marking function, i.e. a function that synthesizes the physical characteristics of the objects to segment. The distance function was chosen for the binary segmentation of beans since it is closely related to the notion of maximal balls included in a set. Similarly, the gradient is used as a marking function for the segmentation of grey-scale images, since its modulus corresponds to the physical notion of contrast. However, we shall see that in practice, the marking function is rarely easy to exhibit.

4.4 How to make a good use of the segmentation tools

We are going to illustrate what has just been said by applying our segmentation tools to two-dimensional electrophoresis spots [5] (See Fig. 43 *(a)*). The problem is here clearly defined: we want to obtain the contours of the dark spots present in this grey-scale image in order to compute their coordinates on the one hand and their sizes and densities on the other hand.

If, by using again the method described above, we simply perform the watersheds of the gradient of this image (See § 4.3), the result is obviously disappointing (See Fig. 44 *(a)-(b)*). Indeed, the gradient exhibits a large number of minima, mainly due to the presence of noise in the initial image. However, one can notice that *all* the spots are marked by these minima and hence all the contours that seem pertinent to the eye are indeed present (See Fig. 44 *(b)*). The watersheds are only *highly over-segmented.*

When facing this result, we have the following alternative:

- To suppress over-segmentation by eliminating the non-significant arc elements of the watersheds.

- To avoid the over-segmentation by questioning the choice of the markers and of the marking function.

We will not dwell at length on the first possibility, which is most of the time a last resort measure—with a few exceptions—, due to the fact that one cannot always be exact about the objects to extract. The over-segmentation mainly comes from the fact that the markers are not perfectly appropriate to the objects to be contoured. In short, the quality of the segmentation is directly linked to the marking function.

The second solution for avoiding over-segmentation involves the elaboration of an appropriate marking function. The extrema of this function have to mark exactly the objects we want to extract. Several possibilities exist for constructing this marking function, e.g. filtering of the image's gradient (See § 2.4.2). This possibility, which constitutes an improvement w.r. to an existing marking function rather than the real construction of a new one, is still sufficient in many cases. Unfortunately, it is not true for the electrophoresis image we are dealing with. Here, we actually need to construct a new marking function. To do so, the first step—as described above—consists in marking the regions we want to segment, i.e. the spots, *but also the background.* To each region must correspond one and only one marker.

Determining these markers involves the use of morphological tools other than the gradient, which is not appropriate in this case. These tools shall be adapted to the problem under study. Here, the spots constituting the dark part of the image, they should be interpreted as the image minima. Yet, the mere extraction of the initial image minima still produces many

unsignificant minima (See Fig. 43 *(b)*). However, contrary to the gradient image, filtering the initial image (with a morphological filter called *alternating sequential filter (ASF)* [26, chap. 10]) suffices here to produce an image whose minima correctly mark the spots (See Fig. 43 *(c)-(d)*).

As to the background marker, it is provided by the watersheds of the filtered image. Indeed, these dividing lines separate the different spots and are located on the lightest zones of the image (See Fig. 43 *(e)*).

We have now the desired markers (See Fig. 43 *(f)*), but this is not enough: we also need a marking function whose minima exactly correspond to these markers. Moreover, this function should be "close" enough to the gradient image so that its watersheds be situated on the contours of the spots. This is why we shall start from the morphological gradient (See Fig. 44 *(a)*) and perform two operations:

1. Impose as minima the previously selected ones (in the following, M will denote the set of the points of $\mathbb{R}^2$ that belong to these markers-minima).

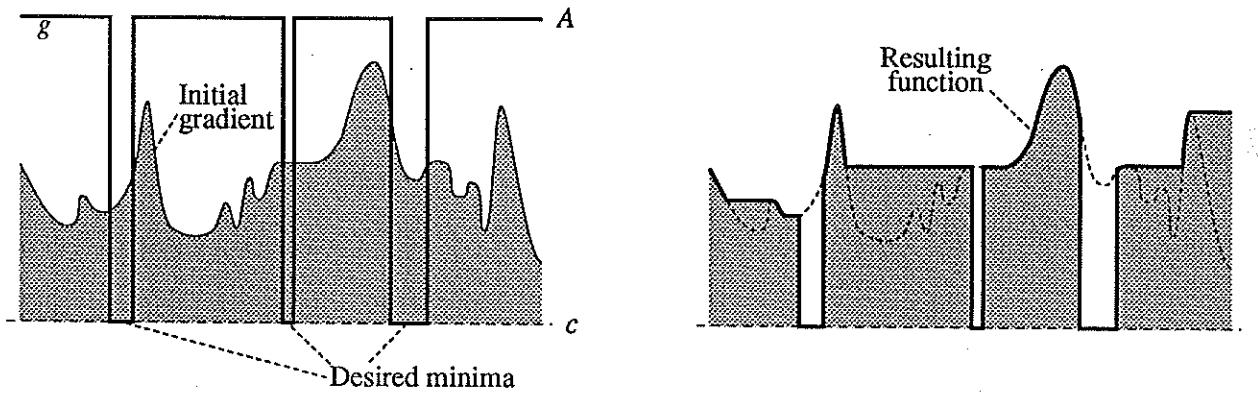
2. Suppress the undesirable minima of the gradient.

In order to do so, we first assign to the points corresponding to the required minima an arbitrary negative value c. Next, we eliminate the undesirable minima by filling in the catchment basins that are associated with them. The morphological technique for achieving the latter operation resorts to the decimal reconstruction R^* introduced in § 3.2. The marking function g which is used here is the following:

$$\forall x \in \mathbb{R}^2, \begin{cases} x \in M & \Longrightarrow \quad g(x) = c \\ x \notin M & \Longrightarrow \quad g(x) = A \end{cases} \tag{56}$$

with R being an arbitrary constant which majorates the gradient function. As concerns the function f to be reconstructed, it is merely the inf of g and of the gradient grad(f) (See Fig. 42a). Note on Fig. 42b that the reconstructed function exactly corresponds to what was looked for. In particular, its watersheds build a subset of the watersheds of the initial gradient grad(f). Moreover, the preserved dividing lines are located on the zones with the highest gradient's modulus.

Fig. 44-*c* to 44-*e* illustrate this segmentation procedure as well as the result we obtain. Fig. 44-*f* shows the markers of the spots and the corresponding contours. Considering the selected markers and the used segmentation algorithm, the result is in accordance with our expectation: each spot has a unique contour which is located on the inflexion points of the initial luminance function—i.e. the original image. It is even possible to improve the contouring by using more sophisticated procedures [5]. For example, we can note that the contours obtained with the above method delimit for each spot a measuring mask which is a bit too small. However, by taking into account the values of the gradient along a given contour, it is possible to extend this mask to the periphery of the spots. The used algorithm consists in dilating each point of the contour in inverse proportion to the corresponding value of the gradient's modulus.



a. function to reconstruct and desired minima *b. filling in of the undesirable catchment basins*

Figure 42: "Geodesic" watersheds

4.5 Choice of the marking functions: variations and recent developments

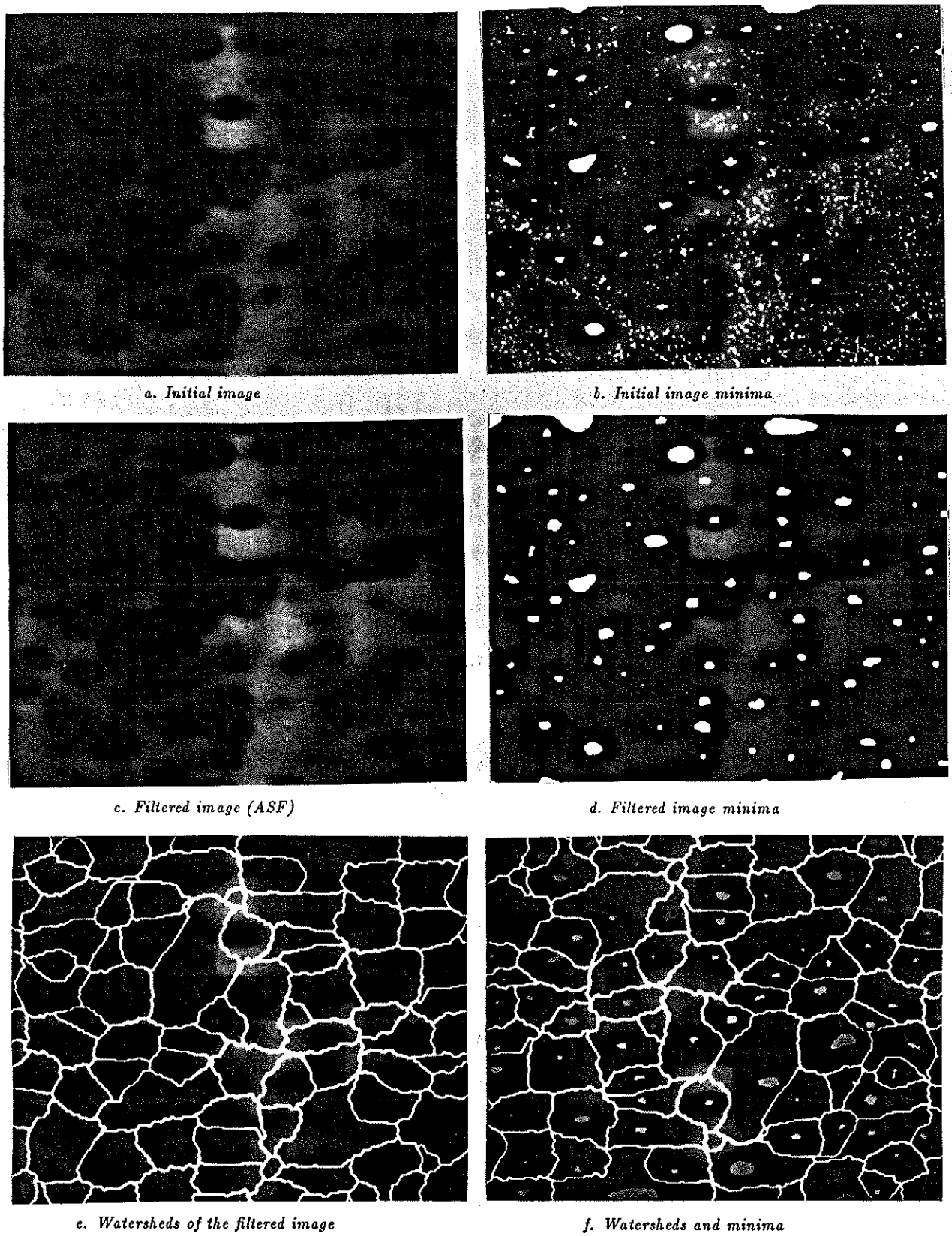
4.5.1 The difficulties of a good marking

The description of the algorithm used to extract the contours of electrophoresis spots have showed that it is not so simple to define markers. It is often cursory to say that the object markers correspond to the gradient function minima. Likewise, the marking by ultimate erosion in the binary segmentation has its limits [1]. The latter is efficient only if the components of the set X under study are "relatively circular" on the one hand, and if they do not overlap too much on the other hand. For instance, when X is composed of two overlapping discs D_1 and D_2, they are both marked by ultimate erosion if and only if their centres are situated on either sides of the radical axis (See Fig.45). Fig. 46 illustrates a case where the ultimate erosion and the watersheds of the opposite of the distance function proved insufficient to separate salt grains.

However, there exists more refined marking tools. Among them, let us mention a transformation called *conditional bissector* [19, page 55], in which the marking is no longer related to the maxima of the quench function but to the extrema of its derivative. Similarly, when the set to be segmented is made up of long particles, it may be interesting to mark their extremities. This can be achieved either by techniques based on geodesic ultimate erosions or by using the maxima of the *propagation function* [24]. The above methods are more complex to use for segmentation problems, since it is not always easy to exhibit the marking functions associated with such markers!...

4.6 Recent developments

In binary morphology, many segmentation problems have not found an appropriate solution yet. Among them, we can mention those which consist in segmenting any heap of convex objects, or intersecting fibers, etc... In such cases, the marking of the objects often requires an a priori knowledge of their geometry. We will not try to solve such complex problems here, nevertheless we can give a few hints as to the possible improvements in the already

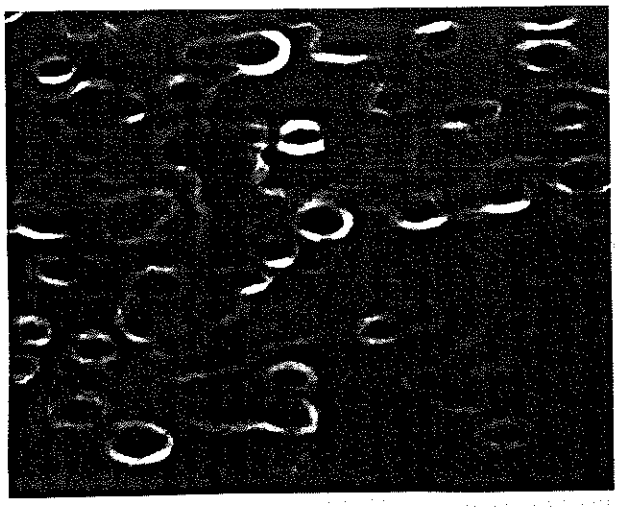


a. Initial image *b. Initial image minima*

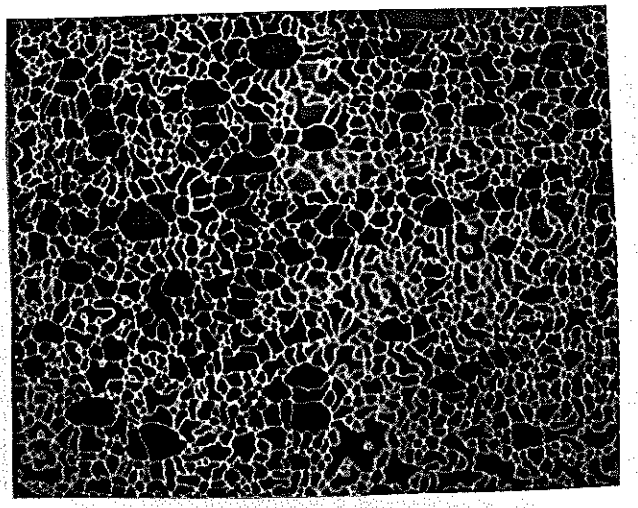
c. Filtered image (ASF) *d. Filtered image minima*

e. Watersheds of the filtered image *f. Watersheds and minima*

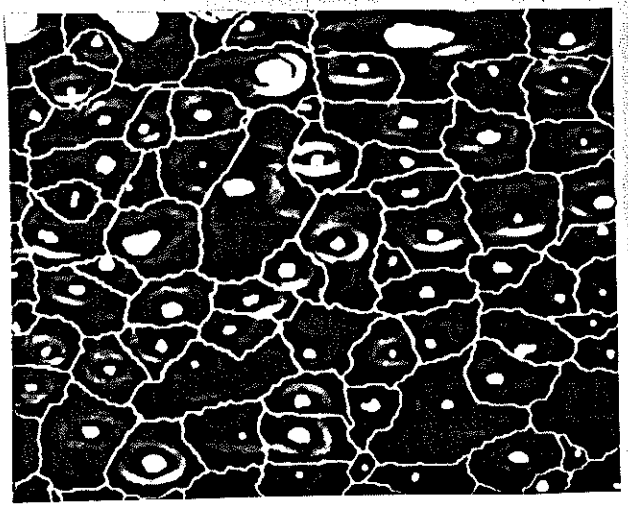
Figure 43: Segmentation of electrophoresis spots



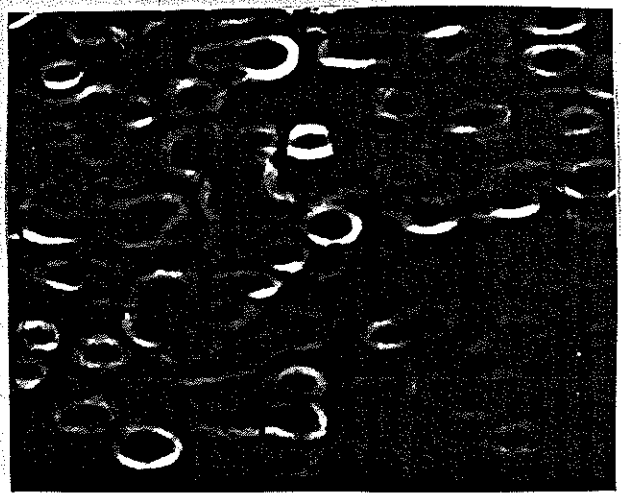
a. Morphological gradient



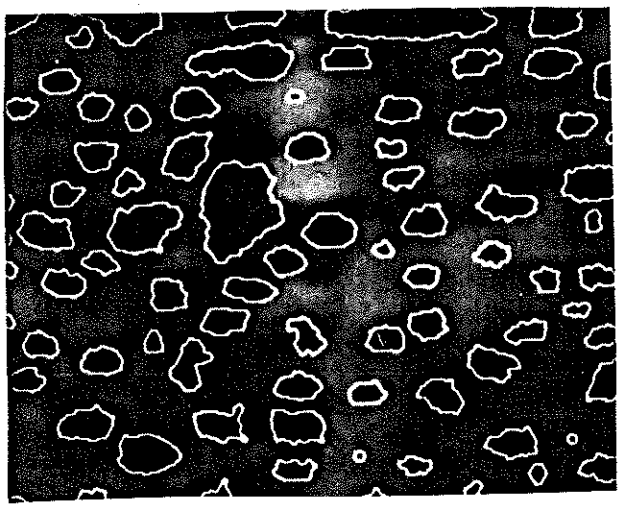
b. Gradient watersheds



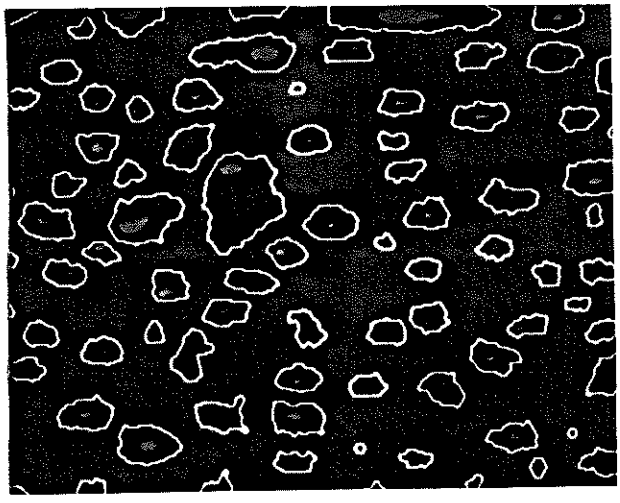
c. Gradient + required minima



d. Modified gradient



e. Watersheds of the modified gradient



f. Markers and contours of the spots

Figure 44: Segmentation of electrophoresis spots (cont.)

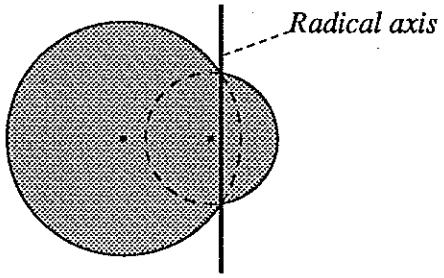
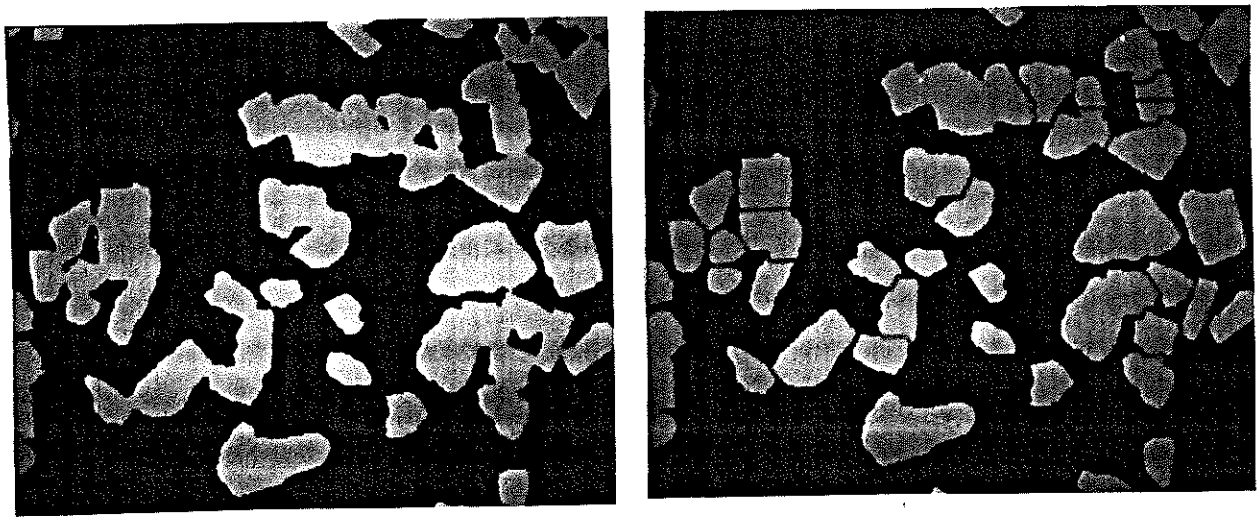


Figure 45: These two discs overlap too much to be separated by an ultimate erosion



a. initial image (grains of salt) *b. "separation" by watersheds of the opposite of the dist. fct.*

Figure 46: Inadequacy of the segmentation by ultimate erosion

described procedures, when dealing with heaps of highly intricated convex particles. We can for example introduce the notions of *critical balls* and *critical function.* The critical balls form a minimal subset of maximal balls necessary for reconstructing the initial set. When used together with the critical function, which associates with each point of a set the density of the maximal balls covering it (See Fig. 47), they provide a more efficient separation than the one based on ultimate erosion (See Fig. 48).

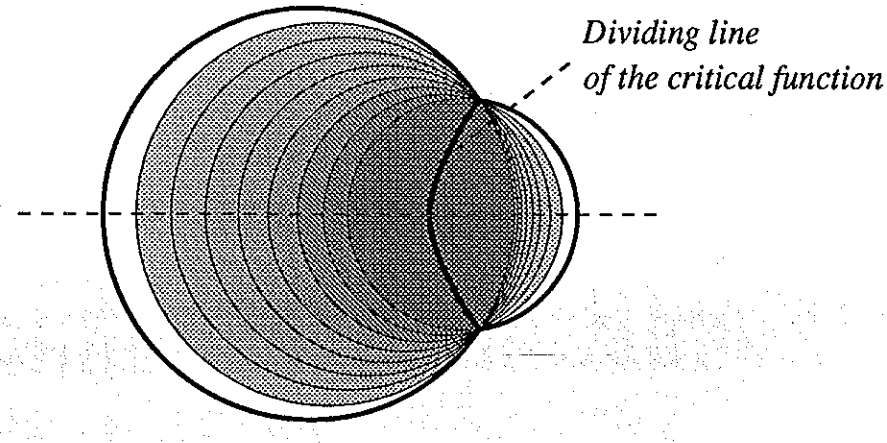
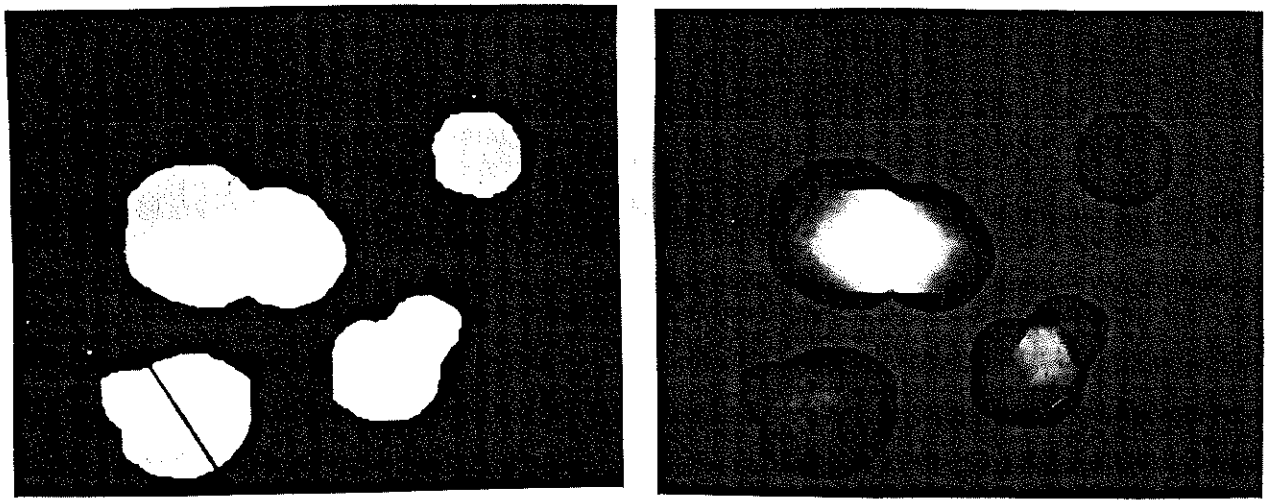


Figure 47: Separation of highly overlapping balls by watersheds of the critical function



a. "segmentation" by watersheds of the opposite of the dist. fct. *b. critical function*

Figure 48: The segmentation of this binary image by the watersheds of the opposite of its distance function is not satisfactory, whereas the crest lines of its critical function correctly separate its components.

As concerns the segmentation of grey-scale images, the major problem lies in the over-segmentation of the gradient watersheds. We have seen—and illustrated on an example—that sophisticated techniques for selecting minima can prevent this over-segmentation. But it is also possible to suppress it afterwards by techniques introducing a hierarchy among the contours and the catchment basins: the watersheds allow us to simplify the image by reducing it into a mosaic of uniform areas (See Fig. 49), which can be interpreted then as a graph. It is

a. initial image *b. mosaic of the catchment basins*

Figure 49: Watersheds and image simplification

then possible to define and perform morphological transformations on this graph [27,28,29], and in so doing, to determine the most significant contours.

5 Conclusion

Throughout these lecture notes, we have attempted to draw a general philosophy of the segmentation in mathematical morphology. Markers and marking functions are the key concepts of this approach. Because of their didactical aspect, the examples we have presented do not allow an exhaustive illustration of the infinite possible variations around these notions. Markers and marking functions are a quite appropriate formalization of the two steps we must necessarily go through when solving any segmentation problem:

1. *What is to be segmented?*
2. *How to do it?*

List of Figures

References

[1] M. Benali, *Du choix des mesures dans les procédures de reconnaissance des formes et d'analyse de texture,* Ph.D. Thesis, School of Mines, Paris, 1986.

[2] S. Beucher & F. Meyer, "Méthodes d'analyse de contrastes à l'analyseur de textures", *Proc. AFCET-IRIA conference on pattern recognition and image processing,* Chatenay-Malabry, France, Fevrier 1978.

[3] S. Beucher & C. Lantuejoul, "Use of watersheds in contour detection", *Proc. Int. Workshop on image processing, real-time edge and motion detection/estimation,* Rennes, France, September 17–21, 1979.

[4] S. Beucher & J. Serra, "Shapes and patterns of microstructures considered as grey-tone functions" *Proc. 3rd European Symposium on stereology,* Ljubljana, Yugoslavia, June 22–26, 1981.

[5] S. Beucher, "Analyse automatique des gels d'électrophorèse bi-dimensionnels et Morphologie Mathématique", *Centre de Morphologie Mathématique, School of Mines,* Paris, 1982.

[6] S. Beucher, M. Blanc & T. Hersant, "Analyse quantitative de clichés stéréoscopiques", *Journal de Microscopie et de Spectroscopie Electronique,* Vol. 7, 1982, pp. 105–106.

[7] S. Beucher, "Watersheds of functions and picture segmentation", *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing,* Paris, May 1982, pp. 1928–1931.

[8] S. Beucher, "Squelettes connexes et non connexes", *Centre de Morphologie Mathématique, School of Mines,* Paris, March 1989.

[9] L. Calabi & W.E. Harnett, "Shape recognition, prairie fires, convex deficiencies and skeletons", *Scientific report 1, Parke Mathematical Laboratories Inc.,* Carlisle, MA, USA, February 1966.

[10] M. Coster & J-L. Chermant, *Précis d'analyse d'images,* CNRS editions, 1985.

[11] A. Gagalowicz, *Vers un modèle de texture,* Higher Doctoral Thesis, Paris VI University, 1983.

[12] M. Golay, "Hexagonal pattern transforms", *Proc. IEEE trans. on Comp.,* Vol. C18, No. 8, August 1969.

[13] R. Haralick & L. Shapiro, "Survey: image segmentation techniques", *Computer Vision, Graphics and Image Processing,* Vol. 29, 1985, pp. 100–132.

[14] C. Lantuejoul, *La squelettisation et son application aux mesures topologiques des mosaïques polycristallines,* Ph.D. Thesis, School of Mines, Paris, 1978.

[15] C. Lantuejoul & S. Beucher, "On the use of geodesic metric in image analysis", *Journal of Microscopy,* Vol. 121, 1981, pp. 39–49.

[16] C. Lantuejoul & F. Maisonneuve, "Geodesic methods in image analysis", *Pattern Recognition,* Vol. 17, 1984, pp. 117–187.

[17] G. Matheron, *Eléments pour une théorie des milieux poreux,* Masson, Paris, 1967.

[18] G. Matheron, "Quelques propriétés topologiques du squelette", *Centre de Morphologie Mathématique, School of Mines,* Paris, 1978.

[19] F. Meyer, *Cytologie quantitative et Morphologie Mathématique,* Ph.D. Thesis, School of Mines, Paris, 1979.

[20] F. Meyer, "Skeletons and perceptual graphs", *Signal Processing,* Vol. 16, No. 4, April 1989, pp. 335–363.

[21] H. Minkowski, "Allgemein Lehrsätze über konvexe Polyeder", *Nach. Ges. Wiss. Göttingen,* 1897, pp. 198–219.

[22] O. Monga, *Segmentation d'images par croissance hiérarchique de régions,* Ph.D. Thesis, Paris XI University, 1988.

[23] F.P. Preparata & M.I. Shamos, *Computational Geometry, an introduction,* Springer Verlag, 1985.

[24] M. Schmitt, *Des algorithmes morphologiques à l'intelligence artificielle,* Ph.D. Thesis, School of Mines, Paris, February 1989.

[25] J. Serra, *Image Analysis and Mathematical Morphology,* Academic Press, London, 1982.

[26] J. Serra, *Image Analysis and Mathematical Morphology, part II: Theoretical Advances,* J. Serra ed., Academic Press, London, 1988.

[27] L. Vincent,"Mathematical Morphology on Graphs", *Proc. SPIE Int. Congress Visual Communications and Image Processing 88, 3^{rd} in a series,* Cambridge, MA, USA, November 1988, pp. 95–105.

[28] L. Vincent, "Graphs and Mathematical Morphology", *Signal Processing,* Vol. 16, No. 4, April 1989, pp. 365–388.

[29] L. Vincent,"Mathematical morphology for graphs applied to image description and segmentation", *Proc. Electronic Imaging West,* Pasedena, CA, USA, April 1989, pp. 313–318.

[30] L. Vincent, "Ombrages et ombres morphologiques", *Proc. PIXIM conference l'image numérique à Paris,* Paris, September 1989.