Journal of Microscopy, Vol. 121, Pt 1, January 1981, pp. 39–49.
Revised paper accepted 7 January 1980

On the use of the geodesic metric in image analysis

by C. LANTUEJOUL *and* S. BEUCHER, *Centre de Morphologie Mathématique, Ecole Nationale Supérieure des Mines de Paris*, 35 *rue Saint-Honoré*, 77305 *Fontainebleau, France*

SUMMARY

Let X be a phase in a specimen. Given two arbitrary points x and y of X, let us define the number $d_X(x, y)$ as follows: $d_X(x, y)$ is the greatest lower bound of the lengths of the arcs in X ending at points x and y, if such arcs exist, and $+\infty$ if not. The function dX is a distance function, called 'geodesic distance'. Note that if x and y belong to two disjoint connected components of X, $d_X(x, y) = +\infty$. In other words, d_X seems to be an appropriate distance function to deal with connectivity problems.

In the metric space (X, d_X), all the classical morphological transformations (dilation, erosion, skeletonizations, etc.) can be defined. The geodesic distance d_X also provides rigourous definitions of topological transformations, which can be performed by automatic image analysers with the help of iterative algorithms.

All these notions are illustrated with several examples (definition of the length of a fibre; automatic detection of cells having at least one nucleus, or having exactly a single nucleus; definitions of the geodesic centre and of the ends of a particle without holes, etc.). As an application, a general problem of segmentation is treated (automatic separation of balls in a polished section).

1. INTRODUCTION

Many specimens in image analysis have two phases. The first phase contains the objects under study (biological cells, stringers in steel, etc.). The second phase is the background.

Two successive steps are usually required in order to obtain measurements from an image: firstly, an image transformation in order to emphasize the salient features of the initial image; secondly, the measurements on the transformed image. Most of the time, the transformation does not affect the whole image, but only a single phase, which depends on the type of measurements to be carried out. For instance, in order to determine the size distribution of stringers in a sheet of steel, it is sufficient to analyse only the stringers. On the other hand, the determination of their spatial distribution requires only a study of the background.

It is thus possible to limit the field of study to only one of the phases. This can be very useful in certain cases to increase the speed of measurements.

Let us return to the particular case of the computation of the stringer size distribution. It is usually obtained by analysing the stringers individually. In this procedure, all the stringers are studied one after the other and independently of each other. This procedure is of course very time consuming, and we may wonder whether it would be possible to study all the stringers at the same time. In such a case, all of them should nevertheless be analysed independently of each other. One way to maintain this independence is to introduce a distance function on the stringers, such that two distinct stringers are considered to be at an infinite distance from each other.

It may thus be helpful to introduce on the field of study a metric which is not the natural Euclidian one.

In this paper we are especially concerned with a particular metric: the geodesic distance function. Experience has shown that this metric is well suited to deal with connectivity problems, and provides rigorous definitions of topological transformations performed by automatic image analysers capable of manipulating iterative algorithms (e.g. reconstruction of particles from masks, etc.). It should be noted that in a field of study with its geodesic metric, a simply connected particle behaves as if it were convex. This fact is very helpful in extending to particles of any shape the transformations and measurements which are usually defined only for convex particles (e.g. length of a particle, etc.).

In this paper we have tried to present all these different considerations in an intuitive way. Some of them raise several mathematical difficulties which are either not presented here or avoided by the simple assumption that all of the objects under study contain their own boundary. For more information, see Lantuejoul & Beucher (1979).

2. HOW LONG IS A FIBRE?

In image analysis, many specimens contain long narrow particles (e.g. glass or asbestos fibres, eutectic alloy lamellae, neuron arms, etc.). Workers concerned with such specimens usually want to know the distribution function of the length of these particles. But what is the length of a particle?

Let us consider such a particle. In order to compute its length, we can measure the distance between its two ends (see Fig. 1-1). This method unfortunately does not work for round or twisted fibres (see Fig. 1-2).

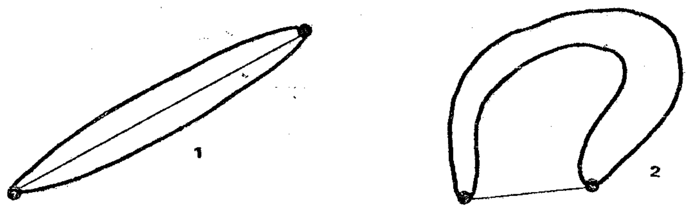

Fig. 1. The definition of the length of a particle as the length of the segment of line between it two ends is not always effective.

Another method consists in measuring the length of the skeleton of the particle (Blum, 1973). This method is also ineffective in the case where the particle has round ends or a rough boundary (see Fig. 2).

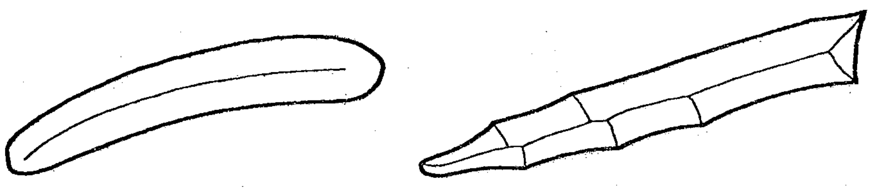

Fig. 2. The definition of the length of a particle as the length of its skeleton is not always effective.

It is our aim to define the length of a particle precisely. The two definitions given above were both rejected because they were not sufficiently general. They are applicable only to a limited class of particles. We thus arrive at the question of how the class of long narrow particles can be defined. Unfortunately, it is impossible to answer such a question, because the description of a particle as long and narrow is quite subjective. At what length/width ratio can a rectangle be considered as being elongated? Without information on the shape of the particles under study, length must be defined regardless of shape, so as to remain applicable to particles of any shape.

Consider an arbitrarily shaped particle X and let x and y be two points within X. There exist several paths in X linking x and y (see Fig. 3-1). The shortest one is called 'geodesic arc' (see Fig. 3-2), and its length is denoted by $d_X(x, y)$.

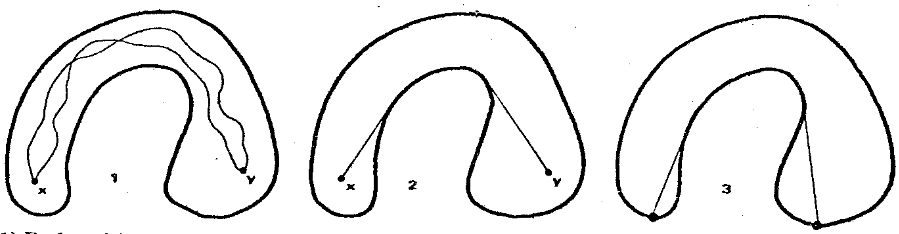

Fig. 3. (1) Paths within the particle. (2) Geodesic arc. (3) The length of the particle is defined as the length of its longest geodesic arc.

We define the length of the particle as the length of the longest geodesic arc within the particle (see Fig. 3-3).

$$L(X) = \sup_{x,\, y \in X} d_X(x,y)$$

This definition may seem somewhat arbitrary to the reader. However, it has three big advantages: (i) this definition is general, as it is applicable to particles of any shape; (ii) this definition is robust, in the sense that a slight deformation of a particle slightly modifies its length; (iii) this definition is operational, in so far as a technology exists or can be devised to use it.

Sometimes particles intersect themselves. A discussion of this is presented in Appendix 1.

3. GEODESIC DISTANCE FUNCTION AND PARTICLE RECONSTRUCTION

Now, let us assume that X is not a single particle, but a population of particles. If the two points x and y belong to two distinct particles, there is no path in X linking x and y, and we can write $d_X(x, y) = +\infty$.

It can easily be shown that the function d_X satisfies all the properties of a distance function:

(i) $d_X(x, y) \geqslant 0$ and $d_X(x, y) = 0$ if and only if $x = y$

(ii) $d_X(x, y) = d_X(y, x)$

(iii) $d_X(x, z) \leqslant d_X(x, y) + d_X(y, z)$

In the following, the function d_X is termed 'geodesic distance function'. The reader can compare on Figs. 4-1 and 4-2 the discs $B_X(x, \lambda)$ and $B(z, \lambda)$ with centre x and radius λ, with the geodesic metric d_X and with the natural Euclidian metric d of the space $\mathbb{R}^2$ in which X is embedded. Obviously, $d \leqslant d_X$.

In order to have a better understanding of Fig. 4-1, imagine that the particles are a string of ponds, and that a stone is thrown into one of them. A front of ripples appears, and we observe them at successive moments.

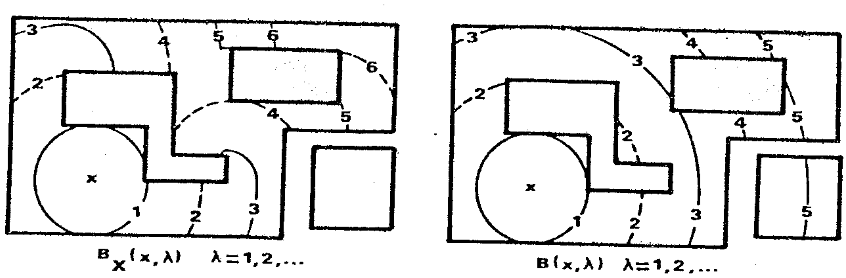

Fig. 4. (1) Discs with geodesic metric. (2) Discs with Euclidian metric.

From the metric d_X, we can define the geodesic distance between a point x of X and a subset Y of X. $d_X(x, Y)$ is the smallest geodesic distance between x and any point y of Y:

$$d_X(x, Y) = \inf_{y \in Y} d_X(x, y)$$

The main interest in the geodesic distance function lies in that it is perfectly suited to deal with connectivity problems. An illustration of this is provided by the following example. Consider two biological images X and Y. X is a population of cells with parts of broken cells, artefacts, etc. Y is the population of the nuclei. X and Y are obtained using a double staining technique (see Fig. 5). The only cells that must be studied are the complete cells containing a nucleus. The other ones are just artefacts and must be disregarded. How can we detect cells of X having a nucleus?

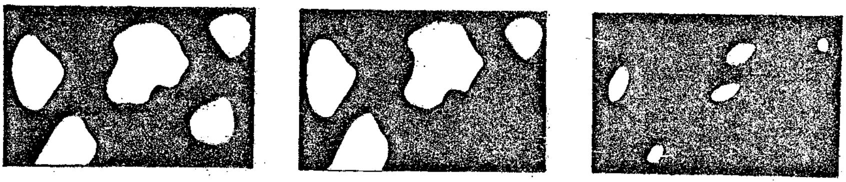

Fig. 5. Detection of cells from a population of cells and artefacts, cells being marked by their nuclei.

Let x be a point of a cell that contains a nucleus. There exists a path in the cell linking x and a point y of the nucleus. In other words, the geodesic distance between x and the nuclei is finite. Mathematically speaking, the population of cells with a nucleus is:

$$\{x \in X \mid d_X(x, Y) < +\infty\}$$

4. MORPHOLOGICAL TRANSFORMATIONS AND GEODESIC DISTANCE FUNCTION

In the space X with the geodesic metric d_X, it is possible to generalize the transformations commonly used in mathematical morphology (Serra, 1980).

(i) if $Y \subset X$, points at a geodesic distance less than λ from Y constitute a set called 'λ-dilated set from Y in X' and denoted $D_\lambda(Y; X)$ (see Fig. 6).

$$D_\lambda(Y; X) = \{x \in X \mid d_X(x, Y) \leqslant \lambda\}$$

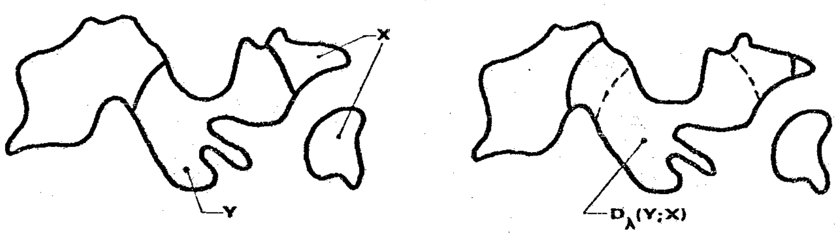

Fig. 6. Geodesic dilation. Y=initial set. $D_\lambda(Y; X)$=λ-dilated set from X in X.

(ii) the points x of X such that $B_X(x, \lambda)$ is totally included within Y, constitute a set called 'λ-eroded set from Y in X' and denoted $E_\lambda(Y; X)$ (see Fig. 7).

$$E_\lambda(Y; X) = \{x \in X \mid B_X(x, \lambda) \subset Y\}$$

It should be noted that X is invariant under dilations and erosions in X:

$$D_\lambda(X; X) = E_\lambda(X; X) = X$$

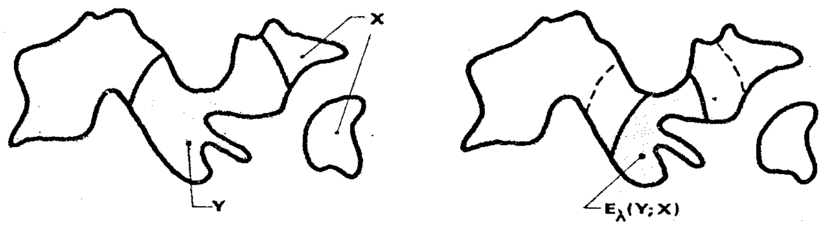

Fig. 7. Geodesic erosion. Y=initial set. $E_\lambda(Y; X)$=λ-eroded set from Y in X.

On the other hand, dilation and erosion in X are dual transformations:

$$E_\lambda(X-Y;\, X)=X-D_\lambda(Y;\, X)$$

(iii) if Y is made up of n distinct particles $K_1, K_2 \ldots K_n$:

$$\begin{cases} Y=\bigcup\limits_{p=1}^{n} K_p \\ p\neq q \Rightarrow K_p \cap K_q=\phi \end{cases}$$

A point x of X is said to belong to the skeleton by zone of influence of Y with respect to X if and only if there exist two subscripts p and q such that

$$d_X(x, Y)=d_X(x,\, K_p)=d_X(x,\, K_q)<+\infty$$

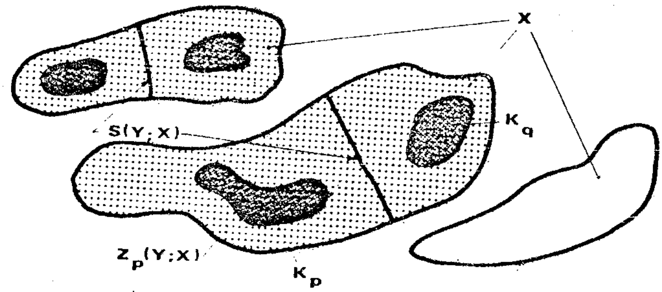

Fig. 8. Geodesic skeletonization by zone of influence. Y=initial set. $Y= \cup K_p$. $Z_p(Y; X)$=zone of influence of K_p in X. $S(Y; X)$=skeleton by zone of influence of Y in X.

The skeleton $S(Y; X)$ bounds the zones of influence. By definition, the zone of influence of K_p is a set denoted $Z_p(Y; X)$, and made up of points of X at a finite geodesic distance from K, and geodesically closer to K_p than to any other K_q:

$$x \in Z_p(Y;\, X) \Leftrightarrow \begin{cases} d_X(x, Y)<+\infty \\ \forall\, q\leqslant n,\, p\neq q \Rightarrow d_X(x,\, K_p)<d_X(x,\, K_q) \end{cases}$$

It should be noted that the zones of influence and the skeleton do not necessarily partition X. The reason is that X can have points at an infinite distance from Y.

As an example, let us return to the two populations of cells (X) and nuclei (Y). Cells having more than one nucleus are overlapping cells and must be disregarded. How can we detect cells with only one nucleus?

We proceed as follows:

(i) construction of the zones of influence of the nuclei with respect to the cells (see Fig. 9-1)

$$Z=\{x \in X \,|\, d_X(x,\, Y)<+\infty\}=S(Y;\, X)$$

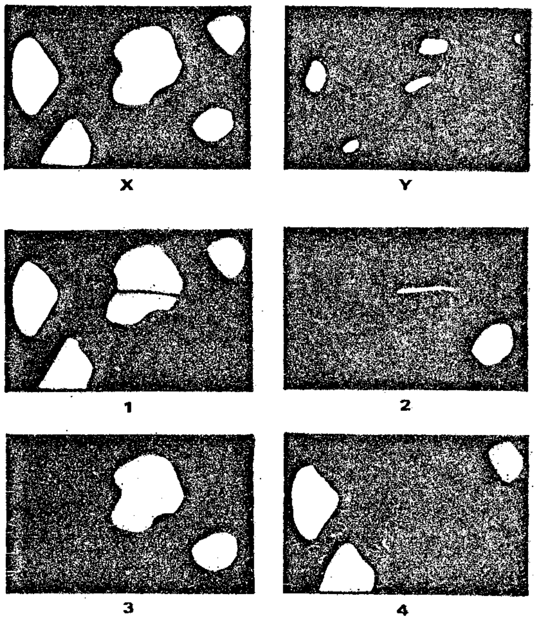

Fig. 9. Detection of cells having only one nucleus.

(ii) points of X which do not belong to any zone of influence, belong either to cells without nucleus or to the skeleton in cells with strictly more than one nucleus (see Fig. 9-2)

$$Z = X - Z$$

(iii) construction of the cells which do not have one nucleus (see Fig. 9-3)

$$Z = \{x \in X \mid d_X(x, Z) < +\infty\}$$

(iv) obtaining cells with exactly one nucleus by substraction from the initial population (see Fig. 9-4)

$$Z = X - Z$$

5. GEODESIC CENTRE AND ENDS OF A PARTICLE WITHOUT HOLES

It is often necessary to mark a particle by a point. In order to increase the speed of subsequent processing, the point must not be arbitrarily located within the particle. In the following, we shall associate with each simply connected particle, a point which is the centre of the circumscribed circle when the particle is a triangle with non-obtuse angles.

Let us again imagine that X is a pond. When a stone is thrown into the pond at point x, we can measure the first time $\lambda(x)$ at which all the shores have been reached by the ripples:

$$\lambda(x) = \sup_{y \in X} d_X(x, y)$$

We thus introduce a function λ which is represented by the time-level lines on Fig. 10.

It can be shown (see Appendix 2) that the function λ is continuous on X and has a single minimum. The point x_0 such that $\lambda(x_0)$ is a minimum is called 'geodesic centre of the particle'. Clearly, we have $B_X[x, \lambda(x_0)] = X$ if and only if $x = x_0$.

If the particle is not simply connected, the function λ always has several maxima (see Fig. 10). The points x such that $\lambda(x)$ is a maximum are located on the boundary of X. They are called the ends of the particle X. In the case where X is a triangle with acute angles, its ends are the three vertices.

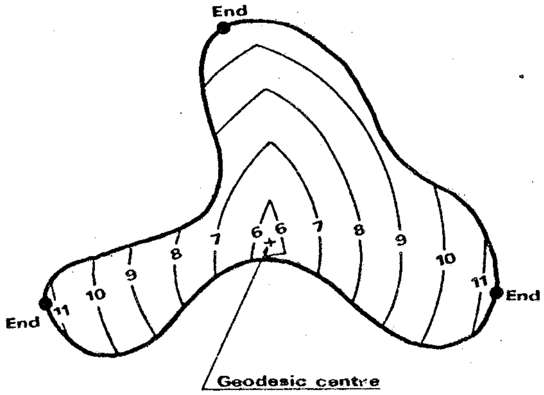

Fig. 10. Construction of the geodesic centre and of the ends of a particle.

6. AUTOMATIC SEPARATION OF BALLS IN A POLISHED SECTION

In this part we are concerned with a structure which is a polished section of a population of balls. They may be metallic or ceramic balls at the beginning of a sintering process, spherical grains of a powder, the granulometry of which has to be computed, etc. Quantitative image analysis on such a structure is not very easy. Indeed, for mechanical and thermodynamical reasons, balls are not necessarily isolated. They sometimes touch, or even cake. In order to perform certain types of measurements, it is essential to separate the balls in sections by drawing contact lines between them. But . . .

6-1. *What is a contact line?*

At a first approximation, the structure under study can be modelled by a subset X of $\mathbb{R}^2$, which is a finite union of discs:

$$X = \bigcup_{p=1}^{n} B(x_p, \lambda_p)$$

Let us consider at first two discs $B(x_p, \lambda_p)$ and $B(x_q, \lambda_q)$ of the population, and let us denote as L_{pq} the radical axis of the two discs, that is, the line which is the set of the points x such that

$$d^2(x, x_p) - \lambda_p{}^2 = d^2(x, x_q) - \lambda_q{}^2$$

When the two discs are touching, L_{pq} is just the line passing through the two points of intersection of the circumferences (see Fig. 11).

If $d^2(x_p, x_q) > |\lambda_p{}^2 - \lambda_p{}^2|$, L_{pq} splits $B(x_p, \lambda_p) \cup B(x_q, \lambda_q)$ into two parts U_{pq} and U_{qp} containing respectively x_p and x_q. If the two discs are touching, a contact line is generated; by definition this contact line is the part of L_{pq} contained within $B(x_p, \lambda_p) \cup B(x_q, \lambda_q)$.

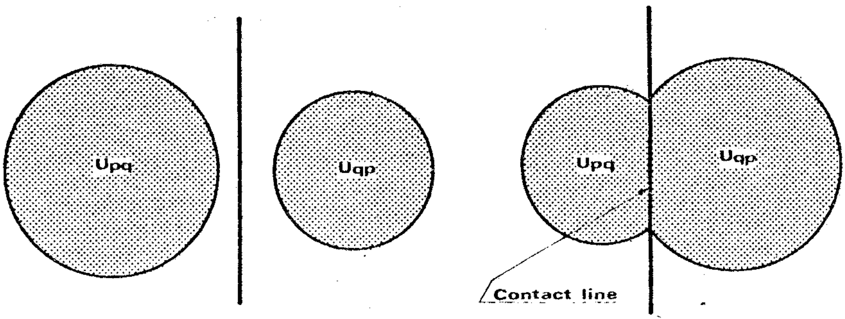

Fig. 11. Definition of a contact line between two discs.

Let us now consider the whole structure X, and let us assume that the discs do not overlap too much, in the sense that

$$\forall p, q \quad p \neq q d^2(x_p, x_q) > |\lambda_p{}^2 - \lambda_q{}^2|$$

Thus, every pair of distinct discs $B(x_p, \lambda_p)$ and $B(x_q, \lambda_q)$ can be split into two parts U_{pq} and U_{qp} containing respectively x_p and x_q. Then, with each disc $B(x_p, \lambda_p)$ of X, we can associate the set $U_p = \bigcap_{q \neq p} U_{pq}$. U_p is non empty, since x_p belongs to all the U_{pq}, and is equal to $B(x_p, \lambda_p)$ if and only if this disc is non-connected with any other disc of X. When X is made up of non disjoint discs, the union of all the U_p's is not equal to X. By definition, the set of all the contact lines between the discs of X is $X - \bigcup_{p=1}^{n} U_p$ (see Fig. 12).

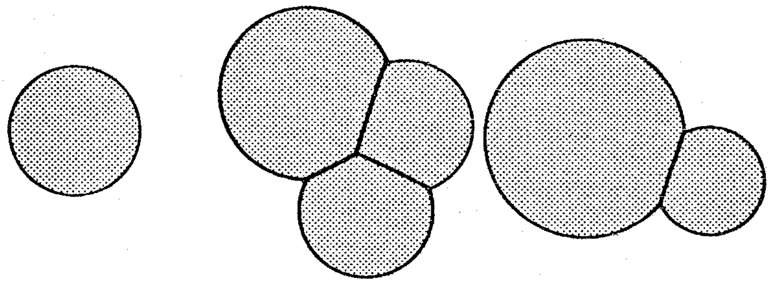

Fig. 12. Definition of the contact lines between the discs of the population.

Now, this definition can be used if and only if the discs of the population are known. Unfortunately, they are not. So, this study entails a preliminary problem which is the identification of the discs of the population.

6-2. *Identification of the discs of the population*

Let $\mathscr{B}$ be a finite family of discs enclosed within X. The discs of $\mathscr{B}$ are said to span X if their union is equal to X. They are said to be independent if no disc of $\mathscr{B}$ is included within the union of the other discs of $\mathscr{B}$. If $\mathscr{B}$ is made up of independent discs spanning X, $\mathscr{B}$ is called a base for X. Clearly, a population of discs can have several bases (see Fig. 13), and if $\mathscr{B}$ spans X, there exists a sub-family $\mathscr{B}'$ of $\mathscr{B}$ which is a base for X.

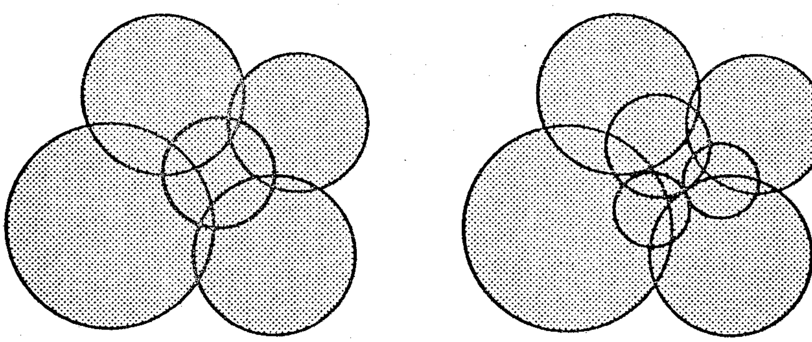

Fig. 13. A population of discs can admit several bases.

Now, let us make the basic assumption that three balls in contact always generate a hole (see Fig. 14). This assumption is realistic for many applications (in particular, at the beginning of a sintering process).

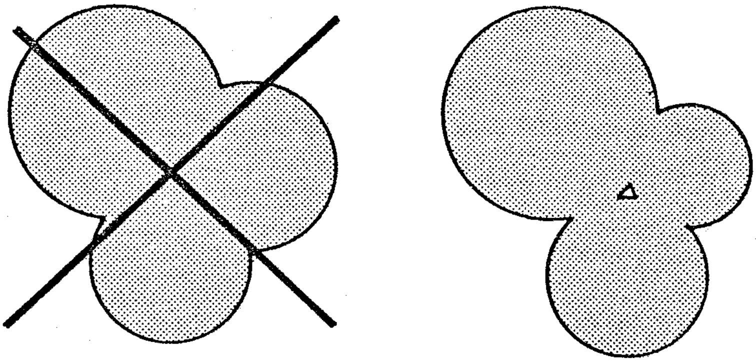

Fig. 14. The basic assumption is made that three discs in contact always generate a hole.

With this assumption, some geometric arguments can be used to show that X has one single base and, furthermore, that the following property holds: any disc within X can be enclosed within two discs of the base. The uniqueness property make it possible to define the balls in section as the discs of the base. We denote them as $B(x_1, \lambda_1) \ldots, B(x_n, \lambda_n)$ or as $B_1, \ldots, B_n$ for brevity. We have thus:

$$\forall\ B(x, \lambda) \subset X, p, q \leqslant n \quad B(x, \lambda) \subset B_p \cup B_q$$

As an immediate consequence, eroding the whole population in $\mathbb{R}^2$ is exactly the same as taking the union of all the eroded couples of discs of the base:

$$E_\lambda(X; \mathbb{R}^2) = \bigcup_{p,q} E_\lambda(B_p \cup B_q; \mathbb{R}^2)$$

Assuming again that balls in section do not overlap too much, which was expressed as follows:

$$\forall\ p, q \quad p \neq q d^2(x_p, x_q) > |\lambda_p{}^2 - \lambda_p{}^2|$$

it can be easily shown that $\{x_p\}$ is a connected component of $E_{\lambda_p}(X; \mathbb{R}^2)$; and conversely, if$\{x\}$ is a connected component of $E_q(X; \mathbb{R}^2)$, the disc $B(x, \lambda)$ is a disc of the base. Thus, the centres of the balls in section are the points which appear as connected components during the successive erosions of the population.

If x_p is the centre of a ball in section, the corresponding radius satisfies

$$\lambda_p = \sup\{\lambda > 0 \mid x \in E_\lambda(X; \mathbb{R}^2)\}$$

6-3. *Construction of the contact lines*

Let us consider first the case of two balls in section $B(x_p, \lambda_p) = B_p$ and $B(x_q, \lambda_q) = B_q$, which are not necessarily in contact. Since $d^2(x_p, x_q) > |\lambda_p{}^2 - \lambda_q{}^2|$, there exists a positive number of two disjoint connected components. Furthermore, the line of contact between B_p and B_q

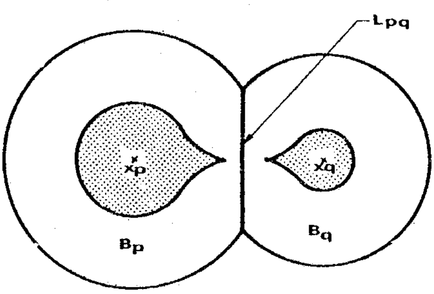

Fig. 15. Construction of a contact line between two discs.

(which is possibly empty) is just the skeleton by zone of influence of these two components in the union $B_p \cup B_q$:

$$L_{pq} \cup (B_p \cup B_q) = S[E_\lambda(B_p \cup B_q; \mathbb{R}^2); B_p \cup B_q]$$

Unfortunately, this formula cannot be immediately generalized to build all the lines of contact of the whole population at the same time.

Let us denote by L_λ the points of the contact line which are farther than λ from the boundary ∂X of the population:

$$L_\lambda = \{x \in L \mid d(x, \partial X) \geq \lambda\}$$

Obviously, we have $L_\mu \subset L_\lambda$ if and only if $\lambda \leq \mu$, with the two extremal properties $L_{+\infty} = \phi$ and $L_0 = L$. It can be proved by induction on the λ-values (Lantuejoul & Beucher, 1979) that the following formula holds:

$$L_\lambda = S[\bigcup_{\mu > \lambda} (E_\mu(X; \mathbb{R}^2) - L_\mu); E_\lambda(X; \mathbb{R}^2)]$$

This formula can be implemented on an image analyser. It has been used on a texture analyser to study the coalescence of bronze balls at the beginning of a sintering process (Chermant *et al.*, 1981).

It should be noted also that this formula is very general. The fact that the objects under study are balls does not appear in it. The only property of the objects which is actually used is that they must remain connected after erosion. Thus, this formula is surely effective in separating general convex objects in section.

ACKNOWLEDGMENTS

The authors wish to thank John Piret and Margaret Reeves for their kind help during the preparation of this paper.

REFERENCES

Blum, H. (1973) Biological shape and visual sciences (Part I). *J. theoret. Biol.* **38**, 2.

Chermant, J.L., Coster, M., Jernot, J.P. & Dupain, J.L. (1981) Morphological analysis of sintering. *J. Microsc.* **121**, 89.

Lantuejoul, C. & Beucher, S. (1979) Quelques considérations sur la distance géodésique. Note interne du Centre de Morphologie mathématique.

Serra, J. (1980) *Image Analysis and Mathematical Morphology*. Academic Press, London (to be published).

APPENDIX 1: *On the length of particles with self-intersections*

It sometimes happens that particles present self-intersection (see Fig. 16-1). In such cases, the general definition of length is applicable, but does not provide results comformable to intuition (see Fig. 16-2).

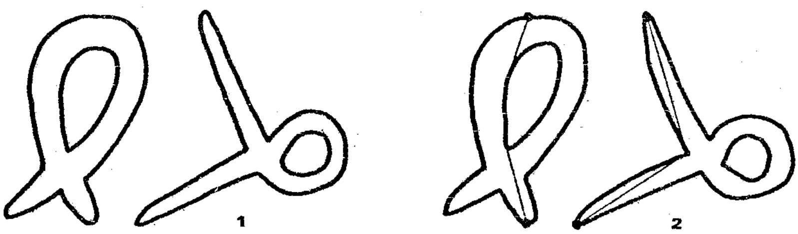

Fig. 16. Self-intersecting particles.

When meeting such particles, the human mind is not content with taking them into account. It also gives a three-dimensional interpretation of them (see Fig. 17).

Of course, several interpretations are possible. For a given one, all the arcs within a particle are not permitted (see Fig. 18). The interpretation of a particle is just the specification of the

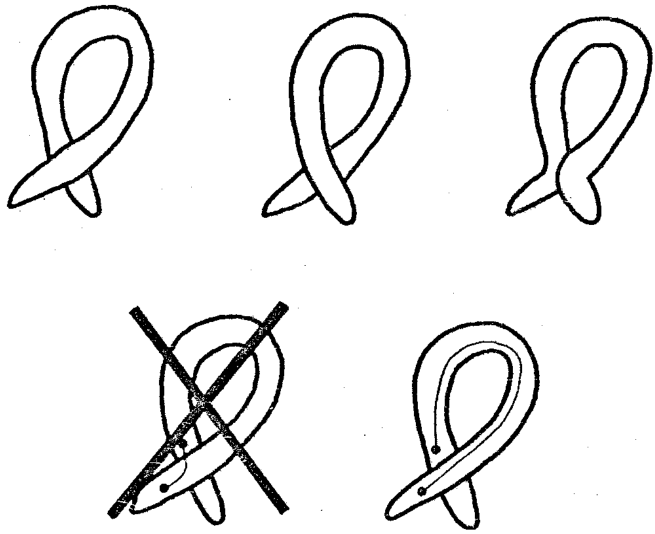

Fig. 18. Prohibited and permitted arcs.

family of the permitted arcs of this particle. Starting with this family, we can define exactly, as previously, the geodesic arcs and then the length of the particle.

APPENDIX 2: *Proof that the centre of the circumscribed disc of a simply connected particle exists and is unique*

In order to establish the proof, we make the following assumptions: (i) X is a simply connected compact set for the metric d_X; (ii) for every $x \in X$, $\lambda(x)$ is finite.

These two assumptions are realistic in practice, and avoid pathological sets which could occur mathematically. For instance, let X be the hyperbolic spiral of polar equation $\rho = 1/\theta$, with $\theta \in [1, +\infty]$; the point whose polar coordinates are $(+\infty, 0)$ is at an infinite distance from any other point of X. X is a compact set in the Euclidian sense, but not in the geodesic sense. Moreover, X is not geodesically connected.

So, let us assume $\lambda < +\infty$. λ is continuous, for $|\lambda(x) - \lambda(y)| \leqslant d_X(x, y)$. Since X is compact, λ attains its minimum value at least one point x_0 of X. It now remains to see that x_0 is unique.

Let x, y and z be three points of X. Suppose that the domain within X that is bounded by the three geodesic arcs $B(x,y)$, $G(y,z)$ and $G(z,x)$ is simpl connected (which always occurs if X is itself simply connected). Then for any point $t \in G(y,z)$ different from y and z, the following convexity inequality holds:

$$d_X(t, x) < \sup\,[d_X(y, x), d_X(z, x)]$$

Let us now suppose that the function λ has two minimal points x_0 and x_1. Let x be a point of the geodesic $G(x_0, x_1)$ different from x_0 and x_1. There exists a point $y \in Y$ such that $\lambda(x) = d_X(x, y)$. Using the convexity inequality, we obtain:

$$\lambda(x) = d_X(x, y) < \sup\,[d_X(x_0, y), d_X(x_1, y)] \leqslant \lambda(x_0)$$

which is a contradiction.