N-11/90/MM

MORPHOLOGICAL SEGMENTATION

F. MEYER & S. BEUCHER

Centre de Morphologie Mathématique
Paris School of Mines
35, rue St Honoré
77305 - Fontainebleau (France)

FONTAINEBLEAU Avril 1990

MORPHOLOGICAL SEGMENTATION

F. MEYER & S. BEUCHER

Centre de Morphologie Mathématique
Paris School of Mines
35, rue St Honoré
77305 - Fontainebleau (France)

Title : MORPHOLOGICAL SEGMENTATION

Send proofs to :

Serge BEUCHER
Centre de Morphologie Mathématique
35, Rue Saint Honoré
77305 FONTAINEBLEAU CEDEX
FRANCE
Telephone : (33) 1.64.22.48.21
Telefax : (33) 1.64.22.39.03
Telex : MINEFON 694736

Abstract :

Methods for image segmentation using mathematical morphology are presented. These methods are based on two main tools : the watershed transform and the homotopy modification which solve the problem of the over-segmentation and introduce the notion of markers of the objects to be segmented in the image.

Some examples in various domains (biology, medecine, scene analysis, 3D images, detection of moving objects, color images) are given. We tried in these examples to emphasize the encountered problems and to explain shortly the proposed solutions.

The algorithms are postponed in the appendix.

List of symbols

$\oplus$: circled plus (dilation)

$\ominus$: circled minus (erosion)

λ , γ : lambda, gamma (greek lowercase)

Φ : phi (greek uppercase)

$\mathfrak{C}$: script C (uppercase)

$\in$: math sign for " belongs to .."

$\cap$, $\cup$: set intersection, set union

Σ : sommation sign (sigma)

INTRODUCTION

The use of the watershed line has proved to be a powerful technique for segmenting images in many situations. We propose here to make an extensive review of the family of related techniques known as morphological segmentation.

S.Beucher and C.Lantuejoul were the first to apply the concept of watersheds and divide lines to segmentation problems. They used it for segmenting images of bubbles and SEM metallographic pictures [1]. We will illustrate the basic notions by dealing with a similar problem at the beginning of this paper. With the attempt to segment more complex images, the major drawback of the method appears very soon indeed, namely the occurrence of severe over-segmentation. The solution to overcome this difficulty is due to F.Meyer [2] and has first been applied for the segmentation of electrophoresis gels [3]. Later, F.Meyer showed how to avoid under-segmentation and how to close broken contours [4]. A generalization of the method to graphs, yielding a pyramidal approach to segmentation is due to S.Beucher [5].

In this paper, we have adopted a pragmatic approach. We proceed step by step, from the simplest to more complex segmentation problems. By using a topographic terminology, we try to render the methodology as visual as possible. Each new tool is presented in the context where it is needed.

We introduce all basic concepts in the first part of the paper. In the second part we illustrate the generality of the method by applying it to a variety of situations : color, motion, 3D images, holographic images etc. We have deliberately postponed to technical appendices the various algorithmic constructions of the watershed line.

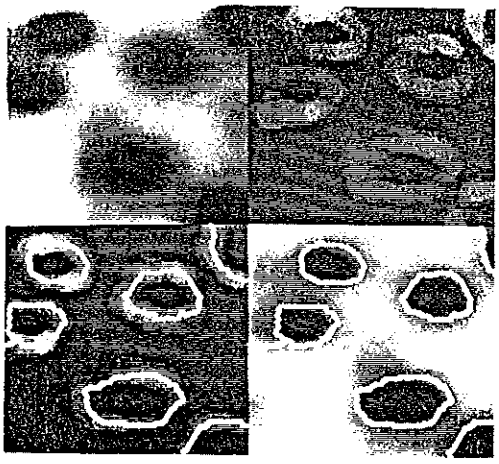
I) A TUTORIAL EXAMPLE : SEGMENTATION OF ELECTROPHORESIS GELS

I-1) Presentation of the problem

Pict.1 presents a simple case of segmentation : single dots from an electrophoresis gel, without overlapping. We will consider, hereafter, each grey-tone image as a topographic surface. The lighter the grey-tone of a pixel, the higher the altitude on the topographic surface. The dots in Pict.1 appear as domes with a round summit. Each dome has a unique summit. Our

problem is to find the best contour.

(a) (b)



(c) (d)

Picture 1

(a) Simple view of an electrophoretic gel, without overlappings.
(b) Morphological gradient of the electrophoretic gel of Pict.1.
(c) Divide line of the gradient superimposed with the gradient image
(d) Divide line of the gradient superimposed with the original image

Fig.1a presents a section of the image along a line. In the x direction we have the spatial distribution of points along a line. The y direction presents the grey-tone for each pixel.

This profile shows well why a simple threshold is not sufficient. With a low threshold, the lowest domes are correctly detected, but the highest domes are much too large. A higher threshold, while detecting correctly the higher domes, misses the lower.

Since absolute values cannot be used, we may try instead the variation of the function, for instance the modulus of the gradient. The simplest way to approximate this modulus has been described by Beucher [6]. It is obtained by assigning to each point x the difference between the highest and the lowest pixels within a given neighborhood of x. In other words, for a function f, it is the difference between the dilated function $f \oplus B$ and the eroded function $f \ominus B$. Fig.2 illustrates the construction.

The morphological gradient is itself a grey-tone function. The highest values now correspond to the most contrasted contours in the original image. Each dot of the original image becomes in the new image a regional minimum

surrounded by a closed chain of mountains, like a basin (A regional minimum is a plateau of uniform altitude without lower neighbors). The varying altitude of the chain of mountains expresses the variation of contrast along the contour of the original dot. An attempt to obtain optimal contours by a simple threshold would lead to similar difficulties as above : a low threshold would give too large contours, whereas high thresholds would miss some poorly contrasted parts of the contours. A low threshold followed by a skeletonization would lead to a thin but misplaced contour, as soon as the gradient profile is dissymmetric (see Fig.1d).

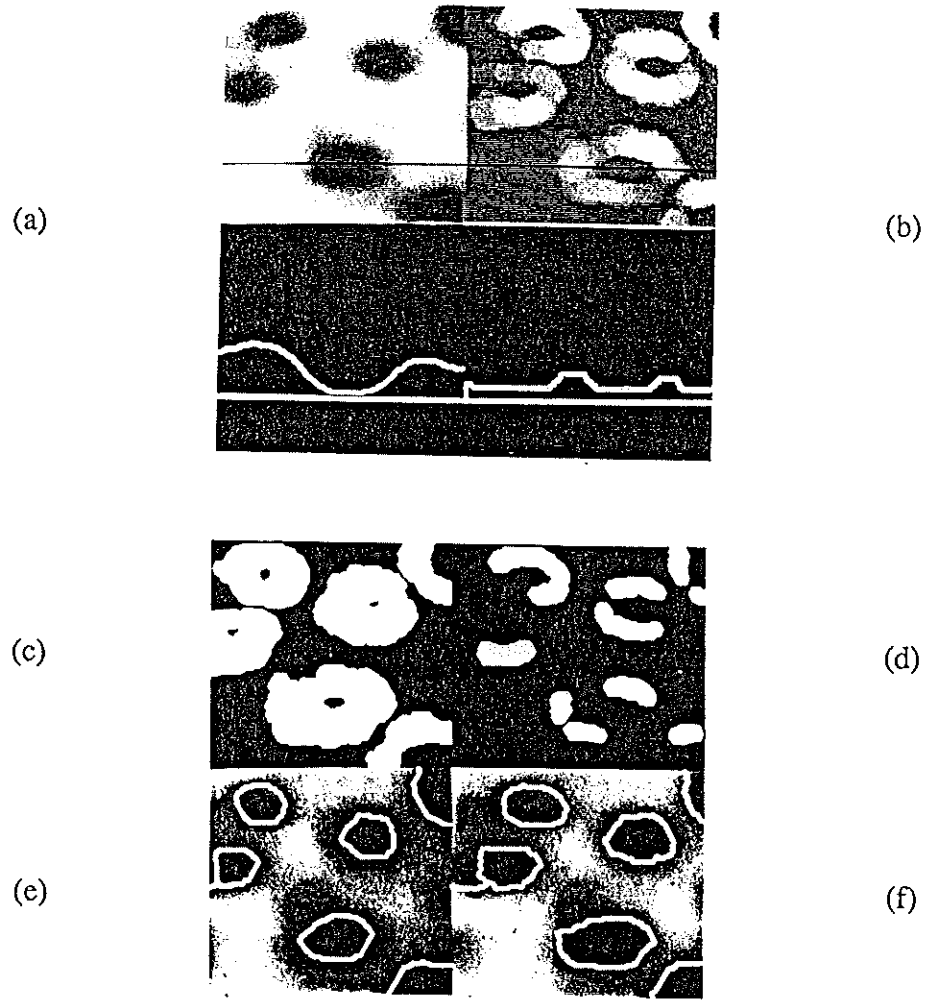


Figure 1

(a) Profile of the topographic surface along a scan line
(b) Profile of the gradient image
(c) Low threshold of the gradient function
(d) High threshold of the gradient function
(e) Skeleton of Fig.1c : misplacement of the contour.
(f) divide lines of the gradient

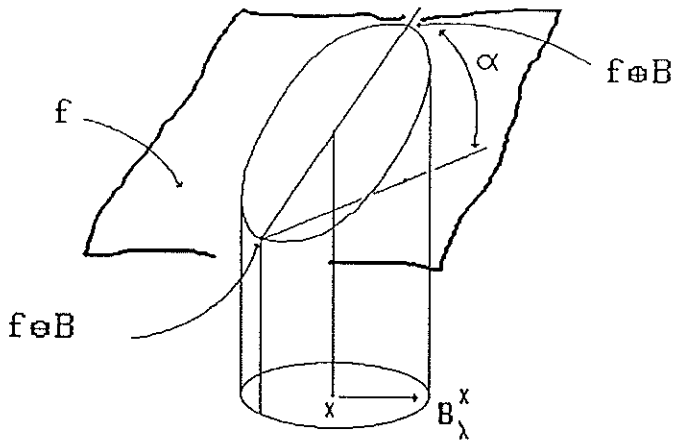


Figure 2

Construction of the morphological gradient of an image

If it is not completely obvious to find a procedure for placing the optimal contour on the gradient image, it is very easy to identify regions without a contour line. The regional minima of the gradient function constitute such regions. One minimum corresponds to the summit in the original function, another to the background. The chain of mountains which separates these minima possesses a crest line. It seems natural to consider this line as the optimal contour. Let's characterize it with some topographic considerations. We will define the catchment basin $\mathcal{C}(M)$ associated with a minimum M. If a drop of water falls at a given location x on a topographic surface, it will follow a line of swiftest descent and reach a regional minimum. Inversely, we associate with each regional minimum its catchment basin, or attraction zone : it is the set of all locations from which a drop of water will reach this regional minimum. Two neighboring catchment basins have a divide line in common : a drop of water falling on a divide line may equally be captured by each catchment basin. The optimal contour is the divide line. All these concepts are illustrated in Fig.3.

From the presentation we have made, the reader easily understands the origin of the name divide line (or watershed line) and catchment basins. However, from this presentation it is not easy at all to infer an algorithmic construction of the divide line. For this reason we switch to another way to

introduce the divide line : the topographic surface will be flooded not by falling rain, but by water coming from the ground, out of various sources. This presentation lends itself to a very simple algorithmic construction (Lantuejoul, Beucher,[1]). Moreover, it is easy, as we will see, to generalize this approach to more complex situations.

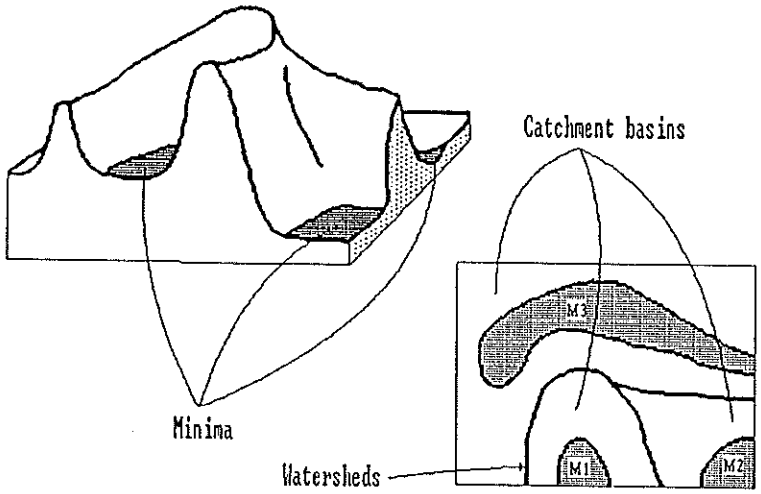


Figure 3

Regional minima, catchment basins and divide line

I-2) A presentation based on a flooding scheme

We will now adopt a flooding scheme. This time the flood does not come from above but from below : we bore a hole in each minimum of the relief and immerse the surface in a lake with a uniform vertical speed. The water, entering through the holes fills up the various catchment basins. We suppose that the immersion speed is slow enough to ensure a constant level in all the basins. In order to avoid the confluence of the floods coming from different minima, we build a dam along the lines where the floods would merge. After complete immersion only the dams emerge and separate the various catchment basins.

I-2-1) Notations

Let f and g be the functions we work on. The regional minima of a function are binary sets and are written $MR_i(f)$. Each i stands for the i-th connected component of the regional minima of f. The catchment basin

associated with a given regional minimum is then written CB (f,MR_i). The divide line is written DL(f).

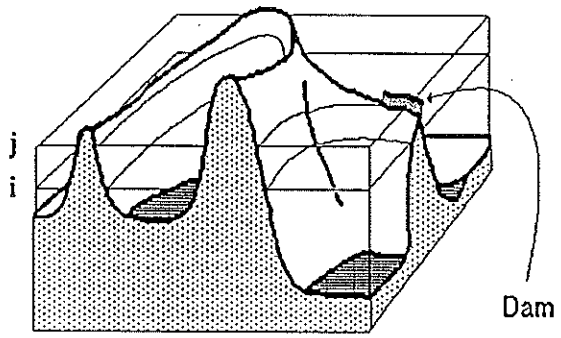
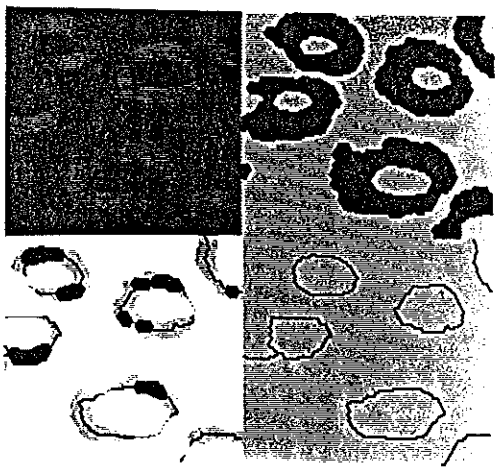


Figure 4

Flooding of the relief and construction of dams.

(a) (b)

(c) (d)



Picture 2

Gradient image (a) and its regional minima (b)
Various steps of the flooding (c) and final result (d)

I-2-2) Algorithm

The construction of the contour of the preceding example may be written :

- f = original image.
- g = f $\oplus$ B - f $\ominus$ B ;/* dilation minus erosion */

- Contour = DL(g) ; /* divide line of the gradient */
- Each dot is of the form : CB (f,MR_i) ;/* catchment basin of a regional minimum of the gradient function. */

II) SEGMENTATION OF ELECTROPHORESIS GELS : THE REAL SITUATION

II-1) Over-segmentation and its remedy

Applied to a real image, the basic procedure we have just described yields a severe over-segmentation (Pict.3b). The contours of the dots are found, but other contours are found as well. This is due to the fact that the gradient image is sensitive to noise : in addition to the centers of the dots and the exterior region, the gradient image has many more regional minima (Pict.3c). And each minimum is surrounded by a divide line. To avoid over-segmentation, we need some additional information : suppose we know before flooding which minima correspond to the centers of the dots, and which to the background. If we come back to our flooding scheme, we will bore a hole only in these dots before immersing the relief. It is the only difference with the preceding algorithm : the flooding and the building of dams take place as previously. The catchment basins of the minima which are not pierced, are filled up by a flood coming from the neighboring catchment basin : as soon as the water reaches the saddle point between both basins, the water rushes through the pass and fills the so far empty basin (see Fig.5). No dam is constructed between these two basins. A dam is only constructed for separating floods originating from different pierced minima. In the end, both spots and background will be covered by the flood, except the divide line which separates them.

In fact it is not even necessary to choose the minima to be pierced. Any region will do. Nor is it necessary that the various markers be connected particles. It is sufficient that they are labeled. Two particles with the same label will be considered to belong to the same region and no dam will be erected between them, if their flooded areas happen to merge. This method to suppress oversegmentation is due to F.Meyer [2].

II-2) Notations

Let (X_λ) be the set of labeled markers. The catchment basin associated

with a labeled binary set (X_γ) is then written CB (f, $X_\gamma \in (X_\lambda)$). Each catchment basin does not only depend on its own marker but on the neighboring markers as well. The divide line is written DL(f,(X_λ)).

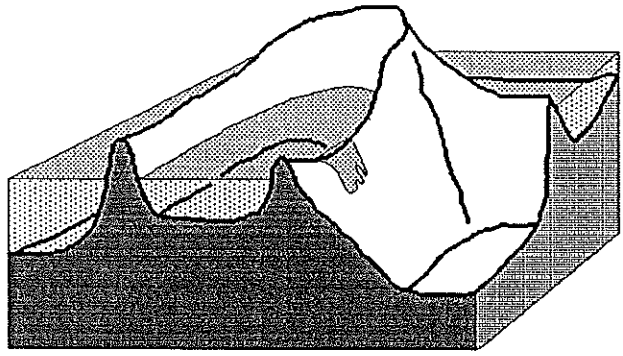


Figure 5

The flood originating from a minimum fills up a neighboring catchment basin. A dam is erected in order to separate the floods originating from different pierced minima.

II-3) Choice of the minima

We apparently solved our problem of over-segmentation. The possibility to choose the initiators of each region makes this algorithm very powerful. It considerably stabilizes the result, since the gradient is extremely sensitive to noise. But have we not shifted the problem ? The knowledge of the gradient function is not sufficient anymore, we need a marker for each region of interest. How to solve this new problem ? In fact this new problem is very often much simpler. Let's give some hints.

- It is not uncommon that the interpretation of the image is made in an interactive mode. Imagine a surgeon planning his operation looking at a tumor in CT imaging. It only takes a short time to mark some zones of interest with a light-pen or a mouse. This task needs not be very precise either. Then the machine can delineate exactly the zones of interest. In contrast an accurate contour tracking demands much more time and much more precise instruments. Not to speak of the three-dimensional case, where it is easy to mark some points

inside and outside the regions of interest ; whereas a three dimensional contour drawing is quite impossible. As is the case in the example given below, namely the contouring of the left ventricle of the heart.

- It is often possible to design a top-hat transformation (M eyer, [7]) in order to mark the inside parts of the objects.

- If the image is made of dark particles against a light background we could use for instance the following procedure : detecting regional maxima in the initial image for the inside markers and regional minima for the background. Often the image has to undergo a severe smoothing in order to suppress small irrelevant details. One may use local averaging, alternate sequential filters, convolutions with a Gaussian, etc.

- In some cases a multi-resolution approach may be adopted for identifying the regions. Smoothing and sub-sampling produce coarser resolutions preserving the presence of the most prominent objects, but not their shape. Once the regions are detected, one may go down to finer resolutions and replace each coarse marker by its catchment basin.

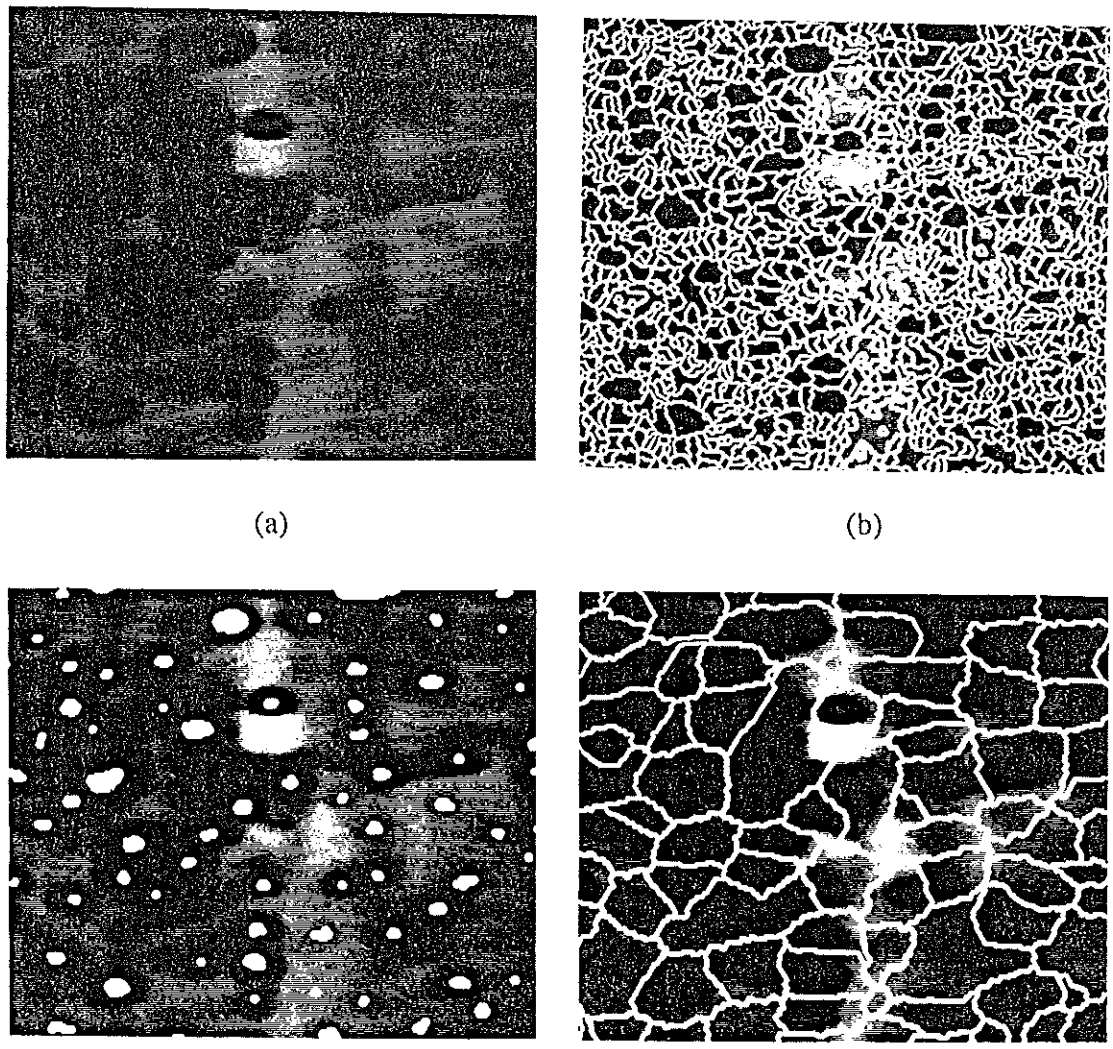
- Sometimes it is easier to identify the inside of the particles than the outside. It is precisely the case in electrophoresis gels. In some regions of severe overlapping, the exterior of the dots may be darker than some isolated small dots in other regions. The detection of the dots themselves however is easy : the detection of the regional minima is sufficient. Seen as a topographic surface, the dots are small basins. The divide line of this surface is an ideal marker for the outside region of the dots. The complete list of the transformations is given in the next section and illustrated from Pict.3b to Pict.3g.

II-3) Segmentation of the electrophoresis gels, second solution

This second solution is given by the following algorithm :

- f = original image. (Pict.3a)
- g = f $\oplus$ B - f $\ominus$ B ; /* dilation minus erosion */ (Pict.3f)
- Inside_markers = MR (f) ;(Pict.3c)
- Outside_markers = DL (f) ;(Pict.3d)

- Markers = Inside_markers ∪ Outside_markers ;(Pict.3e)
- Contour = DL (f, Markers) ; /* divide line of the gradient associated to the markers */ (Pict.3g)
- Each dot is of the form : CB (g, Inside_marker ∈ Markers) ;/* catchment basin of a regional minimum of the original function within the gradient function. */

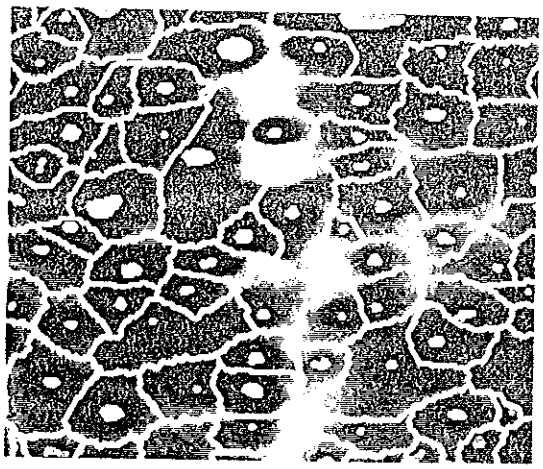


(a) (b)

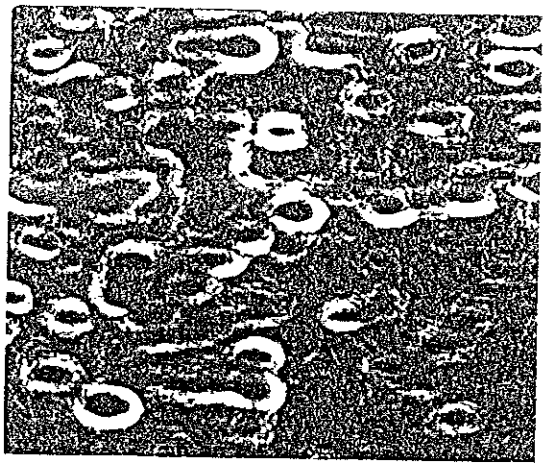
(c) (d)

Picture 3

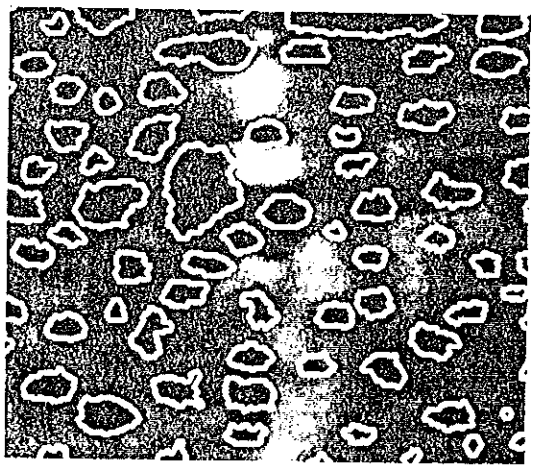
(a) Initial image of the electrophoresis gels, (b) over-segmentation of the watershed, (c) Regional minima (inner markers), (d) Divide line of the initial function (outer marker).



(e)



(f)



(g)

Picture 3 (continued)

(e) Union of outer and inner markers, (f) morphological gradient, (g) Final result of the segmentation.

III) BROKEN CONTOURS AND OVERLAPPING PARTICLES

III-1) Description of the problem

In the previous section we showed how to avoid an over-segmentation of the images. Sometimes however the opposite happens : an under-segmentation occurs. Fig.6a and Fig.7a are two typical examples. Fig.6a sketches a broken contour. As may be seen on the perspective view, this gradient function does not form a closed chain of mountains, separating an inner and an outer regional minimum. The human eye however reconstitutes easily the missing part. In the non missing parts of the gradient function, the crest line figuring the contour is not well defined everywhere. Several disconnected peaks exist

within relatively large and flat zones. In this case, the algorithm 1 would not produce any contour, since inside and outside parts do not belong to distinct catchment basins.

Of course, if one finds the means to introduce an inner and an outer marker, algorithm 2 produces a perfectly closed contour. Its location however would depend heavily upon the location of the markers : in the zones without any relief, the contour would be equidistant of the inner and the outer marker. In the zones with relief, it would follow the crest lines. The final result may be a very winding contour, if the initial markers are not correctly placed. In order to produce a smooth contour, inner and outer markers have to be placed in the right position. The following algorithm provides an optimal and automatic placement of them.

Another basic situation is illustrated in Fig. 7a. It is a schematic representation of an electrophoresis gel with a very poor grey-tone resolution. Three dots overlap. Two of them appear as emerging peaks. The third one appears as a plateau of uniform altitude. Its round, half individualized shape indicates the presence of a dot ; its grey-tone distribution is perfectly homogenous. This case is a mixture of two known situations : on one hand overlapping convex binary objects, easily recognized from their shape. On the other hand neighboring grey-tone domes, identified by their regional maxima. The basic algorithm of the first example solves both pure cases. In the binary case we start with the construction of the distance transformation of the binary sets ; as a result we get two domes. In each case we hawe now domes : the watershed line of the inverse image separates them. Algorithm 1 however clearly fails in segmenting a mixture of both situations. We have to resort to a new algorithm, explained in the next section.

The two previous situations have in common the fact that the human eye immediately recognizes a broken contour or overlapping particles. The segmentation algorithms presented so far fail, because they solely rely on grey-tone information such as regional minima, pass points and divide lines. The human eye takes another important factor into account , namely the shape of the objects. The present section shows how to take into account both shape and grey-tone information for the segmentation.

III-2) Lower-complete, upper-complete and complete functions

A solution will quite naturally appear if we analyze more closely the

flooding scheme used so far. The minima are pierced and the relief immersed at a given time 0. The immersion speed being of constant vertical speed 1, the flood reaches a height t at time t. Since the flood is viscous, a plateau is not immersed at once but progressively. The neighbors of lower pixels are the first to be reached, and each other point is reached at a time proportional to its distance to the lower border. Hence the time when each point is reached by the flood introduces an ordering relation between the points :

x > y iff | (x has a higher altitude than y) or
| (x and y have the same altitude but y is the first to be reached by the flood, i.e. is closer to a downside border of its plateau than x).

This dual ordering relation is the same as the order induced by the composite altitude defined by :

$$\text{Rank}(x) = L * f(x) + \text{lower\_distance}(x)$$

where L stands for the size of the image, and lower_distance(x) is defined by :

lower_distance(x) = 0 if x belongs to a regional minimum
= the length of the shortest path of constant altitude between x and a downside border point if not.

It is easy to check that with this ordering relation, each point has a lower neighbor, except the points of the regional minima. We call such a function a lower complete function. The notation for this order relation is f^-.

By duality we construct upper complete functions : 1) we reverse the function, 2) we define the same ordering relation as above , 3) we reverse the result again. The new ordering relation is written f^+ and defined by :

x > y iff | (x has a higher altitude than y) or
| (x and y have the same altitude but x is the first to be reached by the flooding of the inverse function, i.e. x is closer to an upside border of its plateau than y).

It is possible to construct a grey-tone function inducing the same ordering

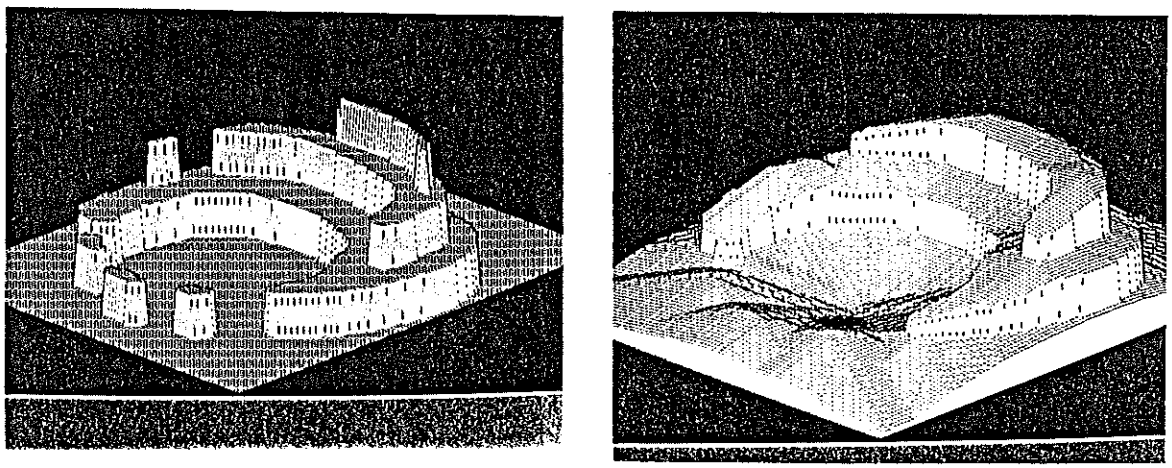
relation :

$$\text{Upper\_complete}(f,x) = L * f(x) + L - \text{upper\_distance}(x)$$

where L stands for the size of the largest plateaus (or of the image, since any plateau is smaller than the image), and the upper_distance(x) is defined by :

upper_distance(x) = 0 if x belongs to a regional maximum
= the length of the shortest path of constant altitude between x and an upside border point if not.

We have applied this transformation to a broken contour (perspective image in Fig.6a). The result is presented in Fig.6b. One sees clearly a minimum emerging in the center of the particle outlined by its contour, while inside and background were connected in the initial image. The original hierarchy is however preserved : the crest in the contour remains apparent with its relative altitude. The construction of the divide line on this new relief will produce a perfectly closed contour.



(a) (b)

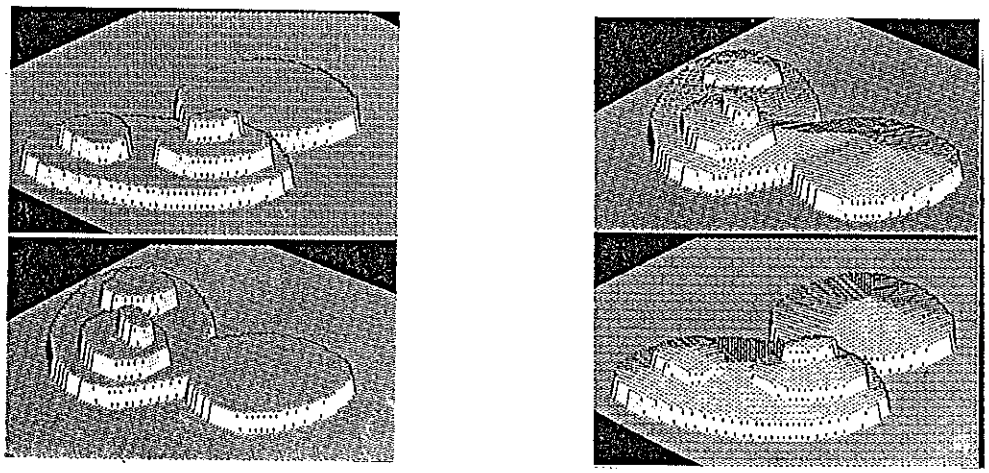
Figure 6

(a) Gradient image representing a broken contour. Inside and outside regions are connected. (b) Construction of the upper complete function associated with the preceding relief. A minimum appears in the center. The upstream of the saddle points yields a prefectly closed and well placed contour line.

III-3) Separation of particles and closing of contours

It is now easy to separate the particles of Fig.7a and close the contour of Fig.6a. We do not construct the divide line of the initial function f. But we use the upper completed function f^+ instead. The regional minima of this new function will be pierced and initiate the flooding. The successive flooding levels will correspond to the successive thresholds of the function f^+. The closing of the contour in Fig.6a and the separation of the particles in Fig.7a is achieved by the same algorithm :

- f = original image.
- Markers = MR(f^+) ;
- Contour = DL(f^+, Markers) ; /* divide line of the gradient associated with the markers */



(a) (b)

Figure 7

(a) Overlapping dots in an electrophoresis gel. Only two dots possess a regional minimum in their center. The other dots are plateaus of uniform grey-tone. Only their shape helps recognizing them as dots. (b) Construction of an upper complete function. A new regional minimum is created in the center of the right dot.

Remark : The algorithm 3 uses both shape and grey-tone information for marking the centers of the particles and for finding their divide lines. As such it is also able to treat pure binary cases : separation of particles on behalf of their shape.

IV) LAST GENERALIZATION

We are now ready for the last generalization which will lead us to the final and complete solution for the segmentation of electrophoresis gels.

Pict.4a shows a complete view to be processed. After a small alternate sequential filter, a first top-hat transformation is applied in order to produce a mask of all dots (this image called Mask is represented in Pict.4b). It permits to eliminate the slightest irrelevant dots :

- f = original image ;
- Mask = Top-hat(f) ;

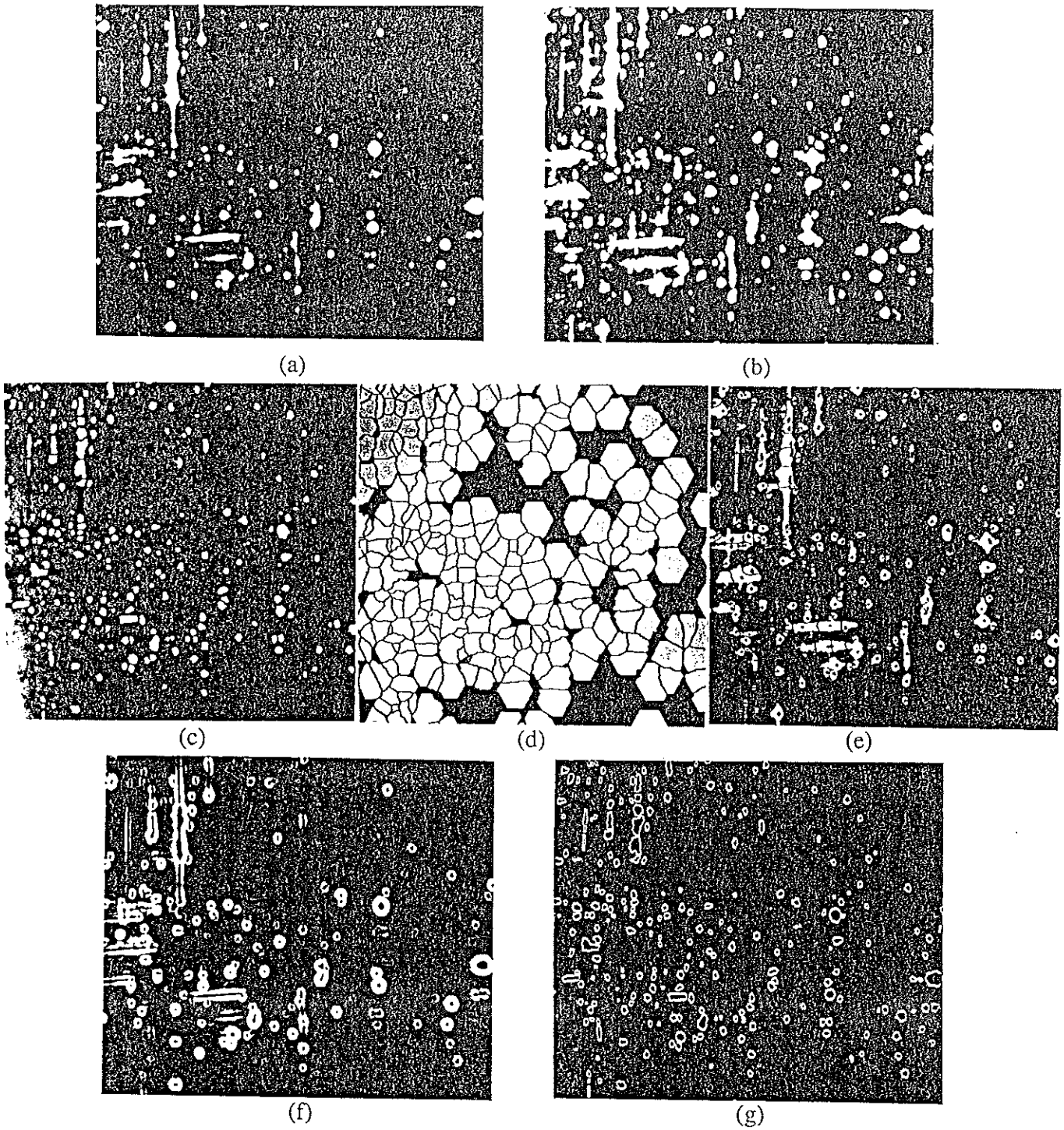
We then associate with the initial function an upper complete function. All regional minima of this function are centers of dots (see Pict.4c). With each center is associated a catchment basin. It is sufficient to construct them within the mask of the spots constructed above (Pict.4d). The boundaries of the catchment basins are good markers for the background :

- $g = f \oplus B - f \ominus B$; /* dilation minus erosion */
- Inside_markers = $MR(f^+)$;(Pict.4c)
- Outside_markers = $DL(f^+)$; (Pict.4d)
- Markers = Inside_markers $\cup$ Outside_markers ; (Pict.4e)

The morphological gradient is illustrated in Pict.4f. We may now assign to the markers obtained from the function f^+ their catchment basins within the gradient image as explained in situation 2 :

- Contour = DL(g , Markers) ; /* divide line of the gradient associated with the markers */
- Each dot is of the form : CB (g, Inside_marker $\in$ Markers) ;/* catchment basin of a regional minimum of the original function within the gradient function. */

The result is satisfactory except in some elongated trails of juxtaposed spots. The recognition of the existence of several spots is solely due to the existence of some prominences in the shape of the trail. Otherwise the grey-tone content is constant within the trail. Thus no gradient information is generated in the trail.



Picture 4

(a) Initial image of an electrophoretic gel, with many overlappings. (b) Rough mask of the dots obtained by a top-hat transformation. (c) Center of the dots (all minima of the upper-completed function). (d) Catchment basins associated to the centers of dots of the original function. The construction of these basins is made only within Pict.4b. It gives the maximal extension zone of each dot.(e) The set difference of the extension zones and the centers yields a rough contour with the correct homotopy. The complementary set is the union of inner and outer markers.(f) Morphological gradient. (g) Final result : construction of the divide line of the relief ($g^+ f^+$).

In absence of any information between the centers of the dots and the marker of the background, the divide line of the gradient will be placed at an equal distance from either. The spots are truncated.

Beside imposing minima to the flooding, we will modify the relief to be flooded as well. We adopt a composite relief, based on both grey-tone images $g^+$and f^+. We denote this relief by Order(g^+, f^+).

$$x > y \text{ iff } \begin{cases} g^+(x) > g^+(y) \ ; \\ \text{or} \\ g^+(x) = g^+(y) \text{ and } f^+(x) > f^+(y) \ ; \end{cases}$$

With this order relation, G is predominant wherever it exists. If g is broken or incomplete, the upper-distance function of g takes care of it. In the remaining plateaus, f^+ determines the flooding. In the empty region within the gradient of the trail, first the centers of the spots start growing until they touch the separating line of the background. In this way the spots are not truncated (Pict.4g).

As a summary we have the following algorithm :

- f = original image ;
- Mask = Top-hat (f) ;
- G = f $\oplus$ B - f $\ominus$ B ;/* dilation minus erosion */
- Inside_markers = MR(f^+) ;
- Outside_markers = DL(f^+) ;
- Markers = Inside_markers $\cup$ Outside_markers
- Contour = DL(Order(g^+, f^+),Markers) ; /* divide line of the gradient associated with the markers */
- Each dot is of the form : CB(Order(g^+, f^+), Inside_marker $\in$ Markers) ; /* catchment basin of a regional minimum of the original function within the gradient function. */

V) MOVEMENT AND 3D IMAGES

The segmentation of moving objects is absolutely similar to the treatment of 3D objects. We will treat them together in this paper.

V-1) Moving objects

We have chosen for illustrating the method the problem of segmenting the moving heart in nuclear medicine. The problem has been treated by F.Friedlander and was the subject of his PhD thesis at our lab [8]. The study was made in collaboration with Prof.Barrat from Hopital Haut Lévêque, Bordeaux and the company Sopha-Medical. The aim was the segmentation of the left ventricule of the heart in nuclear medicine. A sequence of 16 images of the beating heart are taken per second (Pict.5a). As usual in nuclear medicine, the images are very noisy. The usual methods of segmentation [9] solved this segmentation problem for each image separately. This method has the advantage of the simplicity but two major drawbacks :

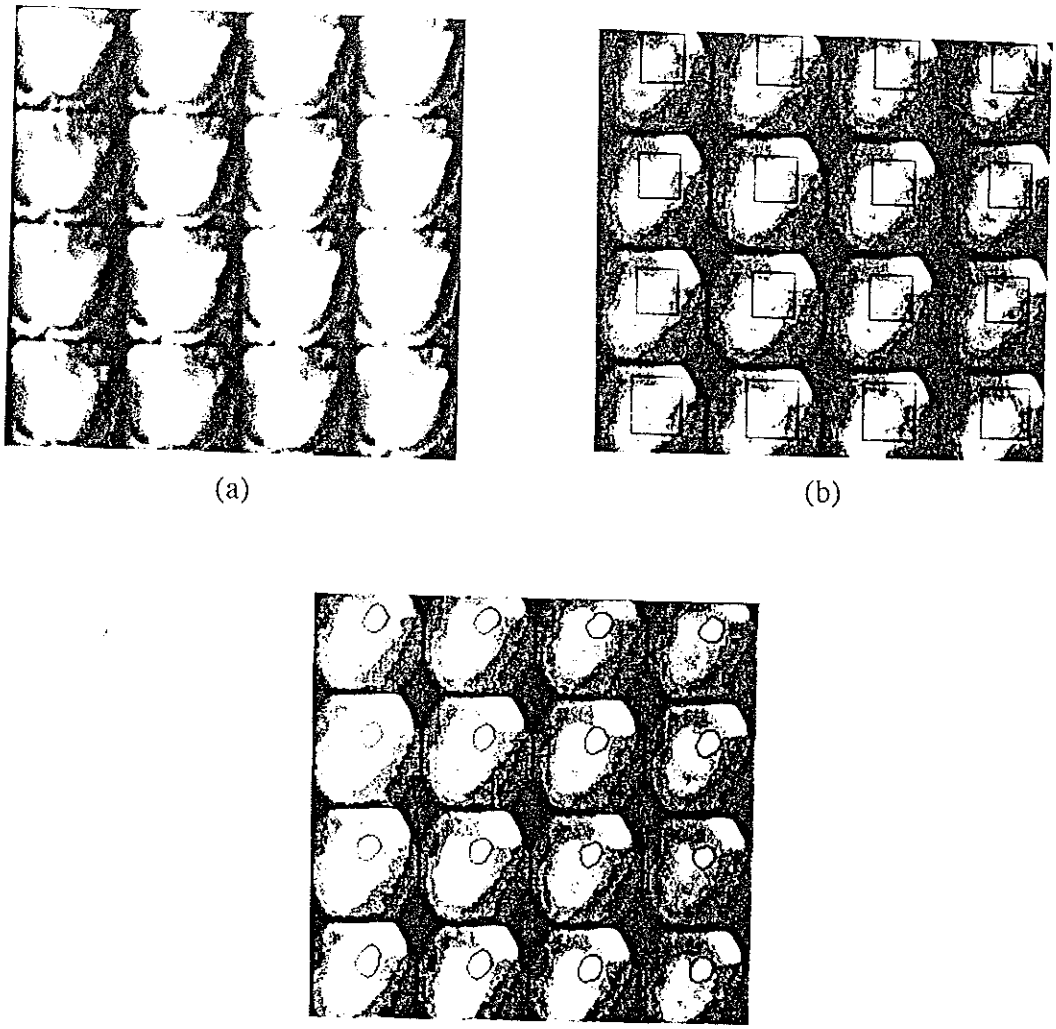
- the segmentation of a given view cannot take advantage of the information present in the view taken just before or just after. The risk of errors is increased for this reason.
- since the successive views represent the movement of the heart, any error in only one of the views will appear as an extremely pathological behaviour of the heart.

In order to make the best of the information present in the sequence of images, we decided to stack them together, as if they were slices of a three-dimensional object. This is of course a very peculiar object without a physical meaning : a three-dimensional grey-tone image with two spatial dimensions and one temporal dimension. For this reason the three dimensionsl body is not isotropic : the spatial dimensions offer a much higher resolution than the temporal dimension. A careful design of non isotropic structuring elements compensates for this fact.

For entering the situation 2, we need inner and outer markers of the ventricule. In our case, their automatic detection was not requested, and the physician marks with a mouse a point within the center of the ventricule and a square box around it (see Pict.5b) . This marking is made only once and is valid for the complete sequence of images.

The image undergoes a median filtering for elimination of noise and smoothing of the contours. We finally need a three-dimensional contour function. The morphological gradient presented above did not give very good results in our case. A three-dimensional modified laplacian filter gave better results. This contour function is then flooded starting from the markers producing a closed three-dimensional contour of the ventricule. Being part of

a three-dimensional body, all sections are segmented in a coherent way, as may be seen in Pict.5c.



(a) (b)

(c)

Picture 5

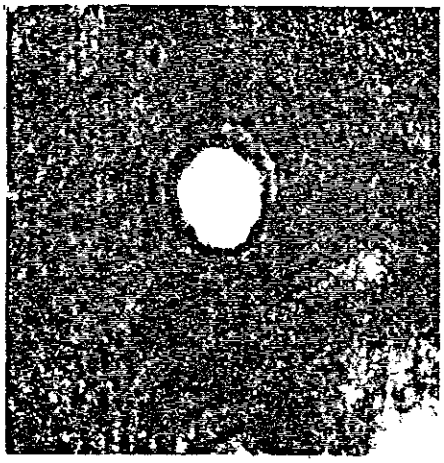
(a) The 16 images taken in sequence of the beating heart. (b) The inner and outer markers, drawn by the physician. (c) The final segmentation obtained for each section of the 3D body.

The validation of the algorithm has been made by Prof.Barat in Bordeaux. The method has given better result than the standard segmentation algorithms based on 2D segmentation. The method is very reproducible : in particular the result is not sensitive to the manual placement of the markers.

V-2) Holographic images. In search of focused objects

This example is another application of segmentation in the three

dimensional space [5]. Pict.6 represents a photograph of drops of water in suspension in the air (magnified view of a mist). The particular aspect of these drops of water comes from the technique which has been used to produce the picture : the mist being unstable and uneasy to handle, a holographic picture is made, then starting from this hologram, a 3D restitution is performed. Finally, this 3D picture is observed by means of a video camera with a small depth of focus optical device. As a result, we get sections of the 3D sample. These sections in fact do not correspond to real 2D cuts in the 3D sample because many drops of water may appear in the same image even if they do not cross the section. This is due to the optical sectioning. As a result, we cannot avoid to see drops which are not close to the optical plane of the section These drops appear with various contrasts and lightnesses : the closer the drop is to the section, the sharper its focus. The segmentation of the 3D picture is necessary to measure two parameters on every drop : its 3D coordinates and especially its depth and its volume. The depth corresponds to the section where the drop of water is best focused. The volume may be obtained simply by integrating over the apparent area of the drop in every section.



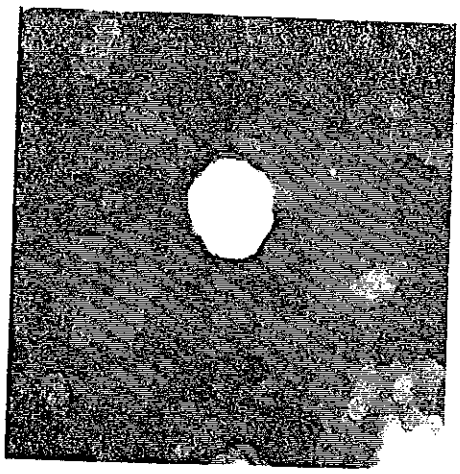
Picture 6

Section in the 3D holographic image of a mist

V-2-1) Markers of the drops

First, a filtering transform is performed on every picture f_i corresponding to the i-th section. In fact, many artifacts coming from the holographic technique must be suppressed from the images : noise, Moiré and diffraction patterns (Newton rings). This can be done by applying a

morphological alternate sequential filter Φ [10]. Pict.7 shows the result.



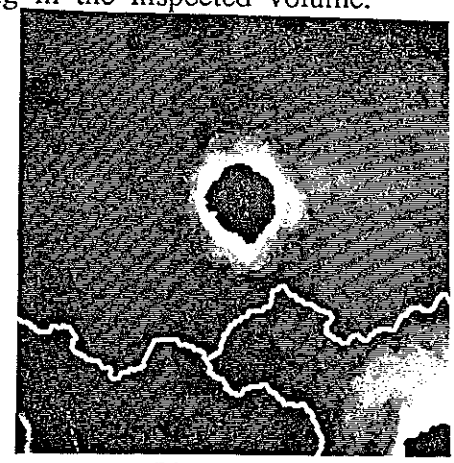
Picture 7

Image after an alternate sequential filtering

Each drop may be marked by simply detecting maxima of lightness in each picture $\Phi(f_i)$. But, owing to the fact that we do not know exactly in which section this maximum may occur, one solution consists in calculating a new two-dimensional image f defined as :

$$f = \underset{i}{\text{Max}} \ (\Phi(f_i))$$

and in directly detecting the maxima of this new image. Doing so, we are sure to detect every drop appearing in the inspected volume.



Picture 8

Inner and outer markers of the drops of water

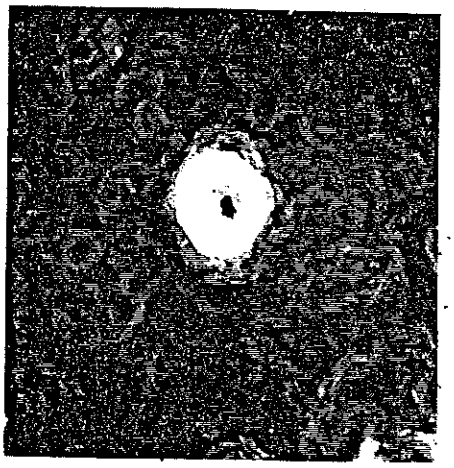
In order to avoid that two or more drops having approximately the same x and y coordinates but very different depth be merged by this transform, the

following operation can be performed : starting from any point (x,y) in the sections, we plot the lightness of this point in any section i as a function of i. If this function exhibits simply one maximum, it means that only one drop appears along the z axis. Conversely, if there are two or more drops, we detect as many maxima of lightness when we go through the sample. It is then possible to label every maximum by the number of drops corresponding in fact to this unique marker. The outer marker is given, as for the gels, by the watershed line of the function f (Pict.8).

V-2-2) Function used for the watersheds

Remember that we want to detect the contour of the drops corresponding to the best-focused image. A simple solution to reach that goal is to compute the gradient in every section and then to calculate the max of these gradient images. Because the best focus corresponds to the highest gradient, we are sure that the watershed of the max will coincide with the best contour. In fact, in order to increase the accuracy of the detection, we compute a 3D morphological gradient and take the sections of this gradient. The process is the following :

- calculate $grad(f) = f \oplus B - f \ominus B$ where B is a 3D ball.
- take the i-th section $grad_i(f)$ of grad(f).
- calculate $g = \max_i [grad_i(f)]$

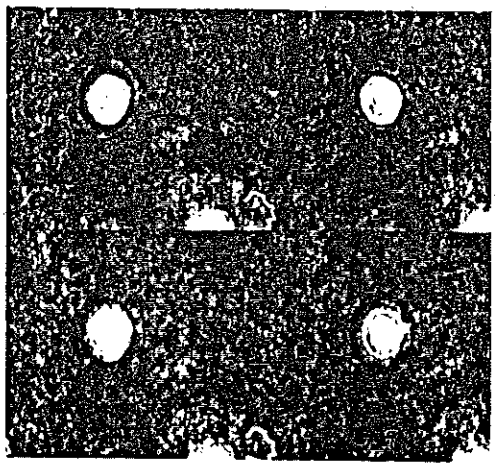


Picture 9

Projection of the 3D gradient by computing the max of all the sections

This two-dimensional function g (Pict.9) is then flooded with a flood

starting from previously selected markers and the watershed constructed. Pict.10 shows the result. We can easily see that the contour produced by this procedure corresponds to the best-focused section. Another advantage of this approach is that it is not necessary for the gradients $grad_i(f)$ to be perfectly centered on the markers of the drops. The modification process will automatically fill up the valleys which may occur when we compute the max. The only constraint is that the markers of the drops do not overlap the crest lines of the gradient. To avoid this, we simply thin the markers of the drops before the image modification.



Picture 10

The best contour superimposed on various sections of the 3D image

The same procedure can be applied to every section $grad_i(f)$ of the gradient image, using the same set of markers. We then obtain the contours of the drops in each section. It is then easy to compare the contour in every section with the best-focused contour and therefore to obtain the section i where this best focus occurs. The section number corresponds to the z-coordinate of the drop. Computing the volume is straightforward.

VI) USING MOTION FOR COMPUTING MARKERS

This example is a part of a rather complex software package used in a road traffic sensor using video cameras [11]. This sensor is able to compute various traffic parameters (traffic amount, speed of vehicles, occupancy...) for each traffic lane. To do so, an initialization process is performed at the beginning of the measurement session in order to segment the various traffic lanes in the scene. To achieve this segmentation, two images are generated,

starting from a sequence of images $\{f_i\}$, f_i being the view of the scene at time i. Typically, thirty seconds up to one minute sequences are used (750 to 1500 images). The first image is produced by simply calculating the average picture of the time sequence :

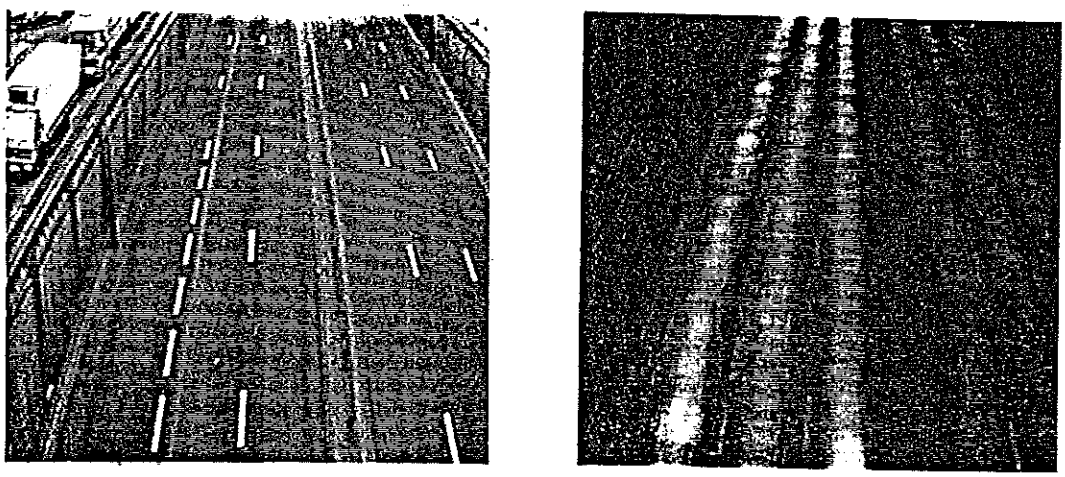
$$f_1 = \Sigma f_i / n$$

The result is a clean image of the road without any vehicle (Pict.11a).

The second picture is given by :

$$f_2 = \Sigma \mid f_{i+1} - f_i \mid /(n-1)$$

This picture which corresponds to the average of the first derivative along the time axis (Pict.11b) has the following interpretation : any region of the scene swept by moving objects (that is the traffic lanes) is characterized by high grey values of f_2.



(a) (b)

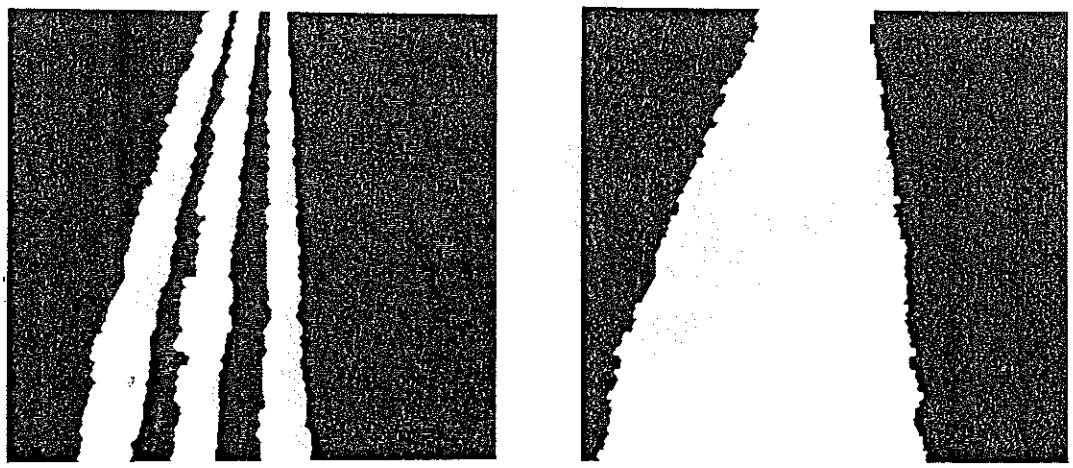
Picture 11

Generated images from motion. (a) Image of still objects. (b) Image of the regions in the scene where some movements occur.

VI-1) Markers

A fair marker of these lanes is given by the maxima of f_2 (Pict.12a). The outer marker is obtained at the end of a more complex algorithm. Starting from the previous image (binary markers of the lanes) we remark that we can connect these markers with a sufficiently large dilation. The size l of this dilation

corresponds to half the average distance between two vehicles running on adjacent lanes. So, performing a dilation of size $2l$ of the inner markers and taking the complementary set furnishes an appropriate outer marker (Pict.12b). Some refinements are used in the real process. The most important consists in using different sizes of dilation according to the position of the dilated point in the image. This technique provides a better marking because it takes into account the perspective deformation of the observed scene.



(a) (b)

Picture 12

Inner and outer markers of the traffic lanes. (a) Markers of the lanes. (b) Dilation of the inner markers generating after complementation the outer marker.

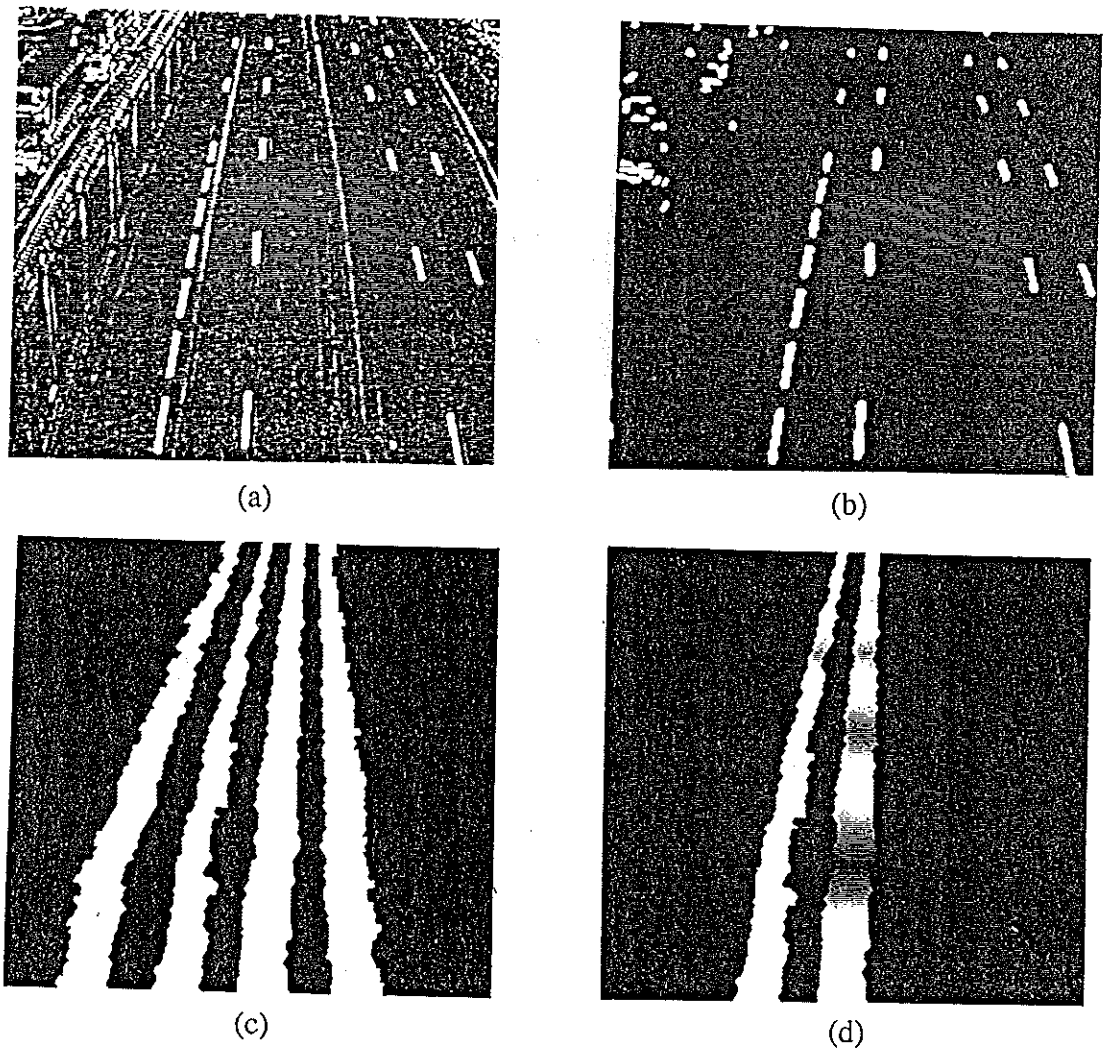
VI-2) Watersheds

One could simply perform a SKIZ (see appendix) of the markers to segment the lanes. This way of doing is equivalent to compute the watershed transform of the markers distance-function. But, for many reasons and especially because we want to check very accurately the position of every vehicle in its traffic lane, we impose that the lane boundaries go through the ground layout (continuous or dotted white lines). Because of the little shift which may exist in the markers image due to the fact that the video camera is not perfectly centered in the middle of the road, there is no reason that the white lines correspond to the crest lines of the distance function. So, we must build a function satisfying this constraint. This is done by the following procedure :

- Starting from the image f_1, extract the ground layout using the Top Hat

transform (see above) (Pict.13a).

- Threshold the previous image, and extract the ground layout. Remember that the signal/noise ratio of the f_1 picture is good (it's an average image), so the range of satisfying threshold values is large enough. Selecting the connected components of the threshold corresponding to the ground layout can be achieved by detecting elongated objects in the binary image. The result is given Pict.13b and is denoted L.



Picture 13

Steps for generating the function f to be segmented. (a) Ground layout extraction using Top Hat transform. (b) Filtered image. (c) Geodesic space and markers. (d) Geodesic distance function.

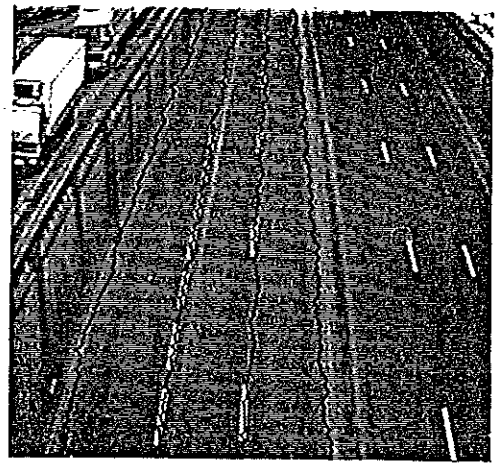
- Denoting by M the set of all the markers (traffic lanes and outer marker), we then compute the geodesic distance function of the set L inside the complementary set M^c (Pict.13c) : for any point x belonging to M^c, we calculate its geodesic distance to L, that is the length of the shortest path

included in M^c connecting x to any point of L (Pict.13d). This function d is obviously equal to the infinite in any connected component of M^c which do not contain ground white lines. For the sake of simplicity, in this case we simply fix its value to the greatest possible grey value $grey_{max}$ usable by the system. Finally, the function $grey_{max}$ - d is used because the ground layout corresponds to the crest lines of this late function. In order to take into account the situation where no ground layout is available, we define the function f to be transformed by watersheds as follows :

$$f = \sup (grey_{max} - d, k)$$

where k is the indicator function of M^c.

It is not necessary in this example to modify the function f because we are sure that its minima are composed of the connected components of M : f has been built in order to fulfill this condition. The result of the segmentation is given in Pict.14. As desired, the lanes boundaries pass through the ground lines.



Picture 14

Segmentation of traffic lanes : final result. The boundaries cross the ground layout.

VII) A HIERARCHICAL APPROACH TO SEGMENTATION

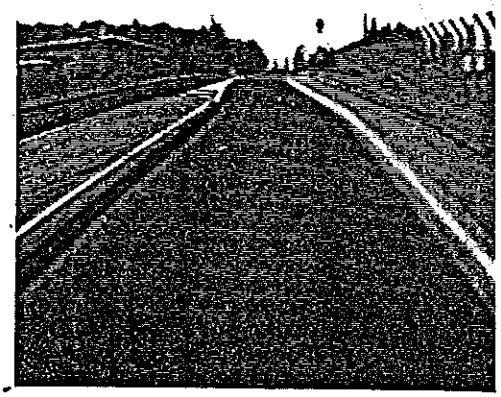
VII-1) Principle of the method

Sometimes, unfortunately, in some applications, the markers selection and

extraction are not so easy. Some pictures may be very noisy and image processing becomes more and more complex. Or, and it is the main problem, the objects to be detected may be so complex, because of a great variety of shapes, of grey levels, of sizes, that it is very hard to find reliable algorithms enabling their extraction. For that reason, we need to go a step further in the segmentation. For instance, the initial segmentation of Pict.15 yields a very unsatisfactory result : many homogeneous regions of the image are fragmented in small parts (Pict.16). Fortunately the watershed transform itself, applied on another level, will help us to merge the fragmented regions. Indeed, if we look at the boundaries produced by the segmentation, they have not the same importance. Those which are inside the homogeneous regions are weaker ; as we will see below, they will get a lower valuation during the further treatment . Now, in order to compare these boundaries, we need to introduce neighborhood relations between them. Two boundaries are considered as neighbors if they surround the same catchment basin. This yields a new non planar graph illustrated in Fig.9b. As a matter of fact, it is possible to extend mathematical morphology to graphs [12]. In our case, the weakest boundaries of the mosaic image correspond to regional minima of the new graph. We may flood the relief of the graph starting from these minima. All the boundaries inside the catchment basins are suppressed. Only the boundaries corresponding to the divide lines of the graph remain.

Let's detail this procedure through the following example.

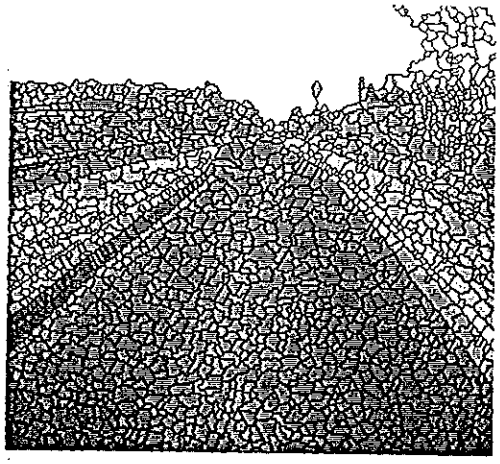
VII-2) Example of the first level of segmentation



Picture 15

Road scene viewed through the windscreen of a car (driver's view)

Consider again on Pict.15 the motor way scene, viewed through the windscreen of a car. Here, the problem consists in detecting the road and possible obstacles which may occur [13]. Only the road segmentation problem will be described in the sequel. As stated earlier,it is not easy to find good markers of the road. If we use a unrefined watershed of the gradient image to detect the contours of the road, the result (given Pict.16) is very disappointing. So, our first objective will be to build a procedure enable to reduce the over-segmentation. To do this, we will start from a simplified image of the initial scene : the partition image, also called mosaic image.



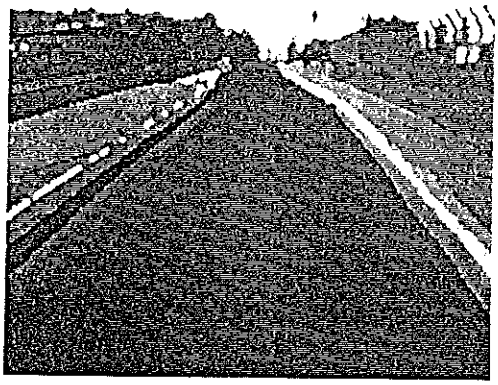
Picture 16

Result of the watershed transform applied to the original gradient image (over-segmentation).

Building this image is made through the following steps :

- Starting from the initial image, we calculate its morphological gradient, then the watersheds of the gradient image.
- Then, we label every catchment basin of the watershed with a single grey value. This value is the grey value in the initial image corresponding to the minima of the gradient (Fig.8).

The result is a simplified image (Pict.17), made of a patchwork of pieces of uniform grey level. We remark that, though the loss of information has been important, the simplified picture always contains the valuable information for segmentation : the contours of the initial image for instance have been preserved.



Picture 17

Simplified mosaic image of the scene

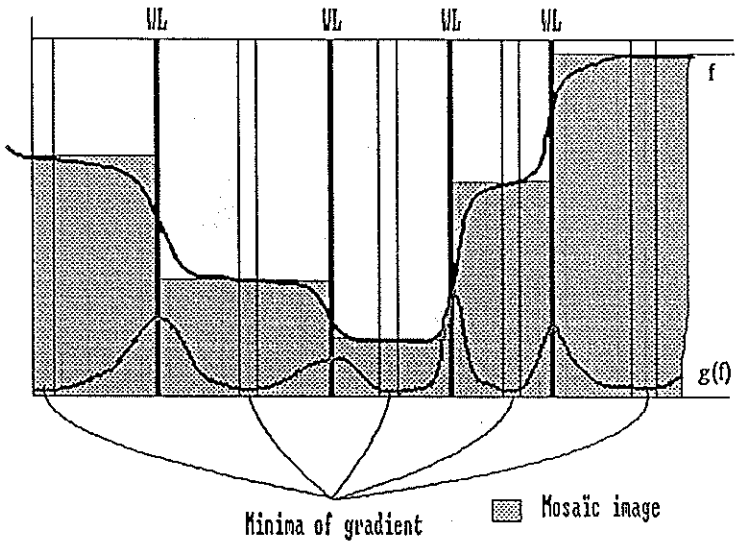
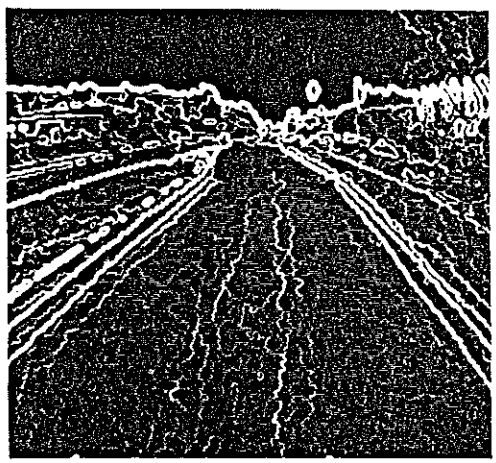


Figure 8

Principle of the computation of the mosaic image



Picture 18

Gradient of the mosaic image. This gradient is a thin gradient (zero thickness edges).

The boundaries between two tiles of the mosaic image are valued with the difference of the grey-tone of these tiles. The result is given Pict.18.

We compute then the neighborhood graph of these boundaries as explained above. The resulting divide lines on this graph identify the boundaries to be preserved (Pict.19b). The extraction of the road marker can be performed easily from that picture (Pict.19c).

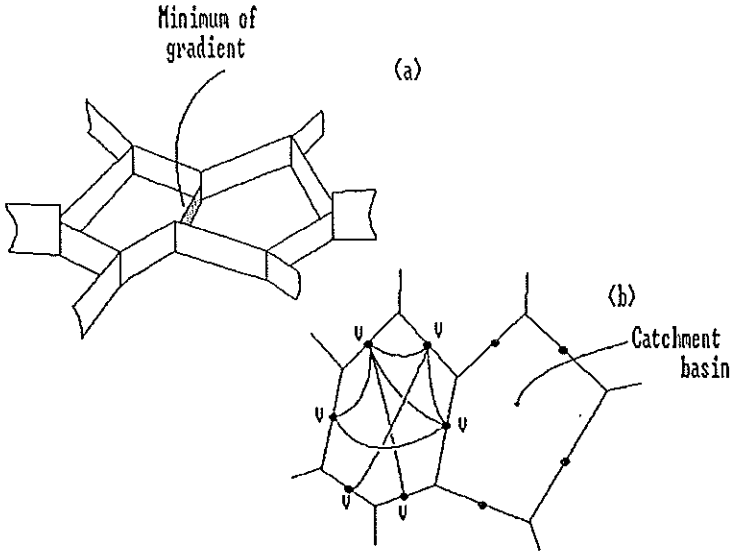
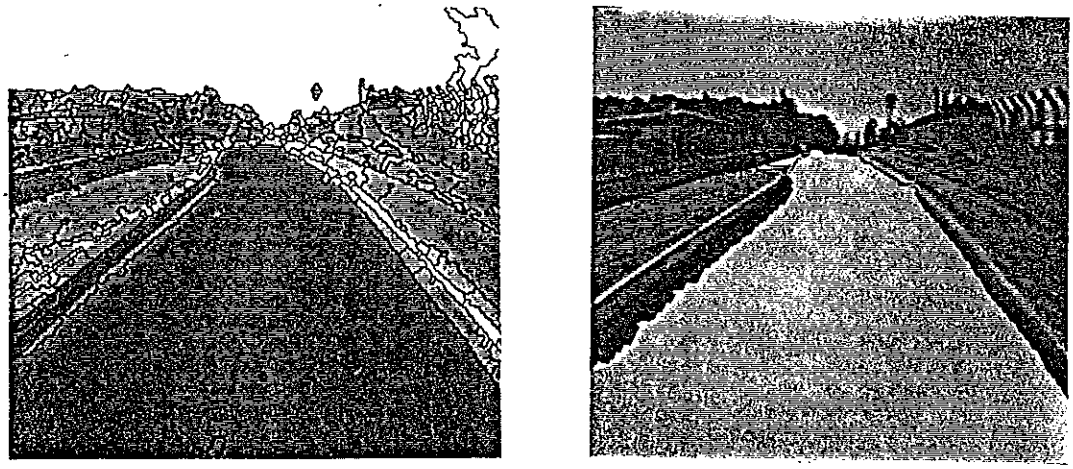


Figure 9

(a) Gradient of the mosaic image and (b) the corresponding graph used for the hierarchical approach.



(a) (b)

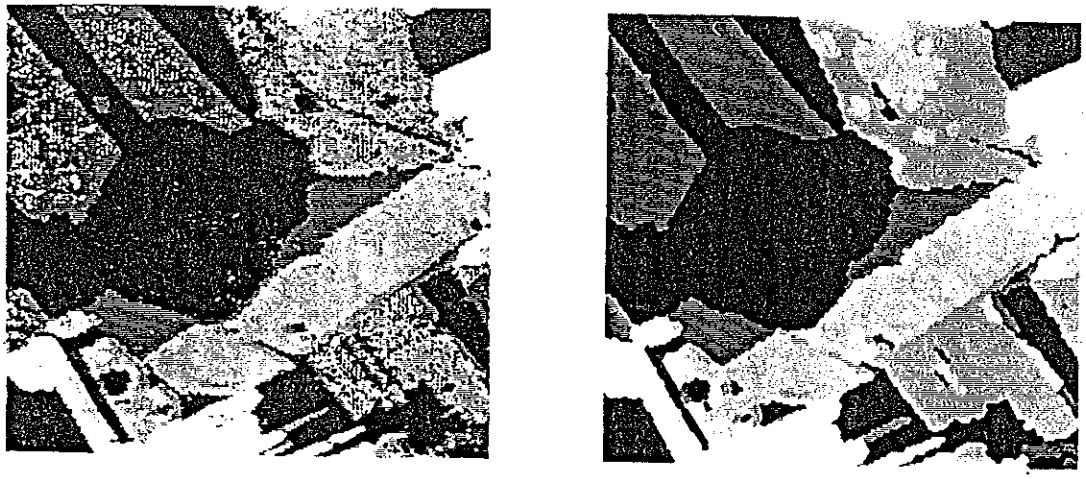
Picture 19

Detection of the road. (a) First level of hierarchy. (b) Marker of the road

VII-3) Further levels of the hierarchy

We started from a heavily fragmented image and we obtained after simplification a new mosaic. It is obviously possible to iterate this simplification process. By these means we get a hierarchy of simplification levels, the last one being always a uniform image. Now arises immediately the question : where to stop ? The answer is evidently case dependant (see [5]). In the case of the example treated above, the first level gives always good results.

VII-4) Another example : color image segmentation and compression



(a) (b)

Picture 20

Color image segmentation. (a) original image, (b) first level of hierarchy

We already emphasized the fact that image simplification deeply reduces the amount of information in the image. So, it can be used very efficiently for image compression, even if the original image is a color image. For instance, Pict.20a represents a color image (lightness component) of a microscopic section of a metamorphic rock. We can apply to the red, blue and green images the previous hierarchical approach in order to extract the various cristallographic phases. In most cases, performing morphological transforms on these three color images has no underlying physical meaning. Fortunately, watersheds transforms of the red, blue and green gradients are allowed because a change in the color will produce a higher gradient in at least one of the three color images. Pict.20b shows the result of the hierarchical approach performed in the three RGB images.

CONCLUSION

This approach of segmentation problems by means of the watershed transform and the use of markers has many advantages. Let us briefly recall some of them.

- The watershed transform provides closed contours, by construction. This fact is of first importance because we do not have to worry about the contours closing of objects, which could happen when using contours detection methods.
- When computing the watersheds, there is a tight fit between the contours which undoubtly appear in the image (that is, obvious contours of objects) and the divide lines of the gradient watershed, even when a severe over-segmentation occur. This explains, in particular, the good positioning of contours and the appropriate construction of the mosaic image used in the hierarchical approach.
- It is a general method which can be applied to a lot of situations. We gave some examples of its use in various applications. But, in fact, the examples are a little selection of the domains in image analysis where this technique has been used efficiently.
- The great advantage of this methodology is that it splits in two separate steps the segmentation process. Firstly, we have to detect what we want to extract : it's the markers selection. Then, we have to define the criteria which are used to segment the image. These criteria may be photometric (contrast variations) or based on the shape of the objects or a combination of both. This combination of different criteria is made easier through the use of powerful morphological tools (geodesic transforms, function completeness and so on) as it has been shown in the examples of electrophoresis gel and traffic lanes segmentations.

Our last argument could be considered too as a major drawback. The problem, which has already been raised, is that we seem to have shifted the difficulty. We replaced the detection of contours by a detection of some markers of the objects to be extracted, and despite the fact that there are many ways for designing these markers, they all seem to be *ad hoc* and there is no general method available to achieve this markers detection. This criticism is perfectly true. But, why such a general method should exist ? The

every day practice of image analysis shows, on the contrary, that in many problems you are not able to see the features you want to extract if you don't know *a priori* what you are looking for. For instance, you may completely miss some features appearing in an X-ray photograph which are, for a physician, a beyond doubt sign of a severe disease. That means that, in most cases, you have to build the objects you want to detect, building which is precisely performed in the image analyser, in a coarser way than the human brain, when detecting or generating the objects markers.

ALGORITHMIC APPENDIX

There are two basic families of algorithms. Family 1 simulates the flooding from regional minima or from other markers. Family 2 detects the dividing lines between two catchment basins : a drop of water falling on such a line has equal chance to be captured by one or the other of two neighboring catchment basins. In fact, as we will see, algorithms of family 2 are only an approximation of what we want. In some configurations they produce wrong results. In both cases we will start studying the binary situation, where the divide line becomes the skeleton by influence zone (SKIZ).

I) ALGORITHMS OF FAMILY 1 : BY FLOODING.

The case of binary pictures will be treated first. The binary algorithms will serve us as building blocks for the construction of the algorithms for grey-tone images.

I-1) Binary case

We have a binary set X, which may be seen as ground. Within X, a series of attracting zones are centered around a family of connected particles (Z_i). The catchment basin associated with a given set Z_i is defined as the subset of all the points of X, which are X-geodesic closer to Z_i than to any other particle Z_j (see Fig.1O). We have now to introduce the X-geodesic distance, we need some definitions :

- **A X-path P of length l** between two pixels s and t is a family of l+1 pixels $(x_0 = s, x_1, \cdots, x_l = t)$ such that :

$$\left| \begin{array}{l} \forall\ i \in [1,l] \quad x_i \in X \\ \forall\ i \in [1,l] \quad (x_{i-1},\ x_i) \text{ are neighbors.} \end{array} \right.$$

- The X-geodesic distance between two points s and t is the length of the shortest X-path with extremities s and t, if such a path exists and $+\infty$ if not.

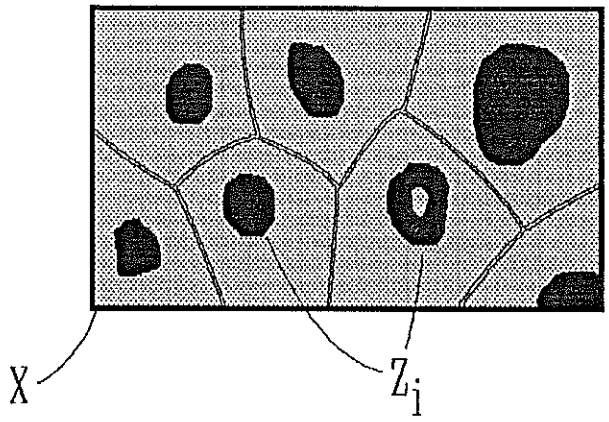


Figure 10

SKIZ of a family of particles Z_i within a field X

Notations :

The X-zone of influence of a set Z_i will be written :

ZI(Field = X ; Center = Z_i) .

The X-divide line : DL (Field = X ; Centers = (Z_i)).

Basic algorithm for the construction of the geodesic SKIZ.

- label all connected particles Z_i of the background with one of the classical transformations of the literature (see for instance [14]). At the end of the transformation each particle Z_i has the label i. We give to all other points of X the value Max_val, a value higher than any valid label.

- Repeat until convergence the parallel transformation :

  For any point x of the plane having the value Max_val :

  -a) if x has no labeled point in its neighborhood, do nothing.
  -b) if x has one and only one label j, in its neighborhood, give this label to x.
  -c) if x has two labeled points in its neighborhood, do nothing.

After convergence, each point with two labeled points in its neighborhood is a point of the divide line.

Remark :

- Despite its simplicity, this parallel transformation is seldom present on a hardwired machine. It is possible however to mimic the behavior of the preceding algorithm by using only grey-tone dilations commonly present in hardware [15].
- Another classical way for constructing the SKIZ consists in skeletonizing and pruning the set X.

I-2) Grey-tone images

I-2-1) Catchment basins of the regional minima

We suppose now that the relief is immersed step by step, each of height 1. It will give us the key of the basic algorithms for constructing the watershed line.

Notations :

- f the grey-scale under study, defined on the domain D_f.
- h_{min} and h_{max} the smallest and largest values taken by f on its domain.
- $T_h(f) = \{ x \in D_f ; f(x) \leq h \}$ the threshold at level h.
- M_i the regional minima and $\mathcal{C}(M_i)$ the associated catchment basins.

The catchment basins will be constructed step by step, corresponding to the step wise flooding. Let $\mathcal{C}_h(M_i)$ the part of the catchment basin associated with the minimum M_i, which is flooded at time h. Hence :

$$\mathcal{C}_h(M_i) = \mathcal{C}(M_i) \cap T_h(f)$$

If the altitude of the minimum M_i is j, then $\mathcal{C}_j(M_i) = M_i$ and $\mathcal{C}_k(M_i) = \varnothing$ for $k<j$.

The union of all catchment basins is written $\mathcal{C}(M)$, and its flooded part at time h is written : $\mathcal{C}_h(M)$. We have $\mathcal{C}_{hmax}(M) = \mathcal{C}(M)$.

Initialization :

$\mathcal{C}_{hmin}(M) = T_{hmin}(f)$; /* the deepest minima are the first catchment basins to be flooded */

Upward flooding.

We suppose that the flood has reached the level h-1, and that the corresponding catchment basins $\mathcal{C}_{h-1}(M)$ have been constructed. The threshold $T_h(f)$ is made of several connected particles. Each connected particle of

$\mathfrak{C}_{h-1}(M)$ is contained in one and only one connected particle of $T_h(f)$. The reverse is not true. There may be connected particles of $T_h(f)$ with zero, one or several connected particles of $\mathfrak{C}_{h-1}(M)$. Each type will contribute in a different way to $\mathfrak{C}_h(M)$:

- A particle X of $T_h(f)$ containing several connected particles $Y_j = \mathfrak{C}_{h-1}(M_j)$ of $\mathfrak{C}_{h-1}(M)$. In this case a dam has to be constructed within X for separating the converging floods. For finding the place where the floods will meet, one has to define the progression of the flood on the plateau X. Let's imagine the fluid with a given viscosity : starting from the sets Y_i the fluid will spread out with uniform speed within X. At all places where the fluids coming from two different sets Y_i and Y_j meet, a dam has to be erected. In more mathematical terms, the dam has to be erected along the geodesic skeleton by influence zones (SKIZ) of $\mathfrak{C}_{h-1}(M)$ within $T_h(f)$. And the catchment basin $\mathfrak{C}_{h-1}(M_j)$ is expanded to its geodesic zone of influence within $T_h(f)$:

$$\mathfrak{C}_{h-1}(M_j) = \text{Zone\_of\_influence } \{\text{Field} = T_h(f) \ ; \ \text{Particle} = \mathfrak{C}_{h-1}(M_j)\}$$

- A particle X of $T_h(f)$ with one connected particle $\mathfrak{C}_{h-1}(M_i)$ inside. X belongs completely to the catchment basin associated with the minimum M_i. In fact the previous procedure still holds : X is the geodesic influence zone of $\mathfrak{C}_{h-1}(M_i)$ within $T_h(f)$.

- A particle X of $T_h(f)$ without any connected particle of $\mathfrak{C}_{h-1}(M)$ inside. This is the typical case where a new minimum is just flooded. X has simply to be added to $\mathfrak{C}_{h-1}(M)$. This situation is easily recognized. It is impossible to reconstruct X using $\mathfrak{C}_{h-1}(M)$ as a marker. Hence the regional minima appearing at the height h are given by :

$$\text{Min}_h(f) = T_h(f) \ / \ \text{Reconstruction } \{\text{Particles} = T_h(f) \ ; \ \text{Markers} = \mathfrak{C}_{h-1}(M) \ \}$$

Remark :

The reconstruction of all particles with a marker inside is nothing but the geodesic dilation of infinite size of the markers within the particles.

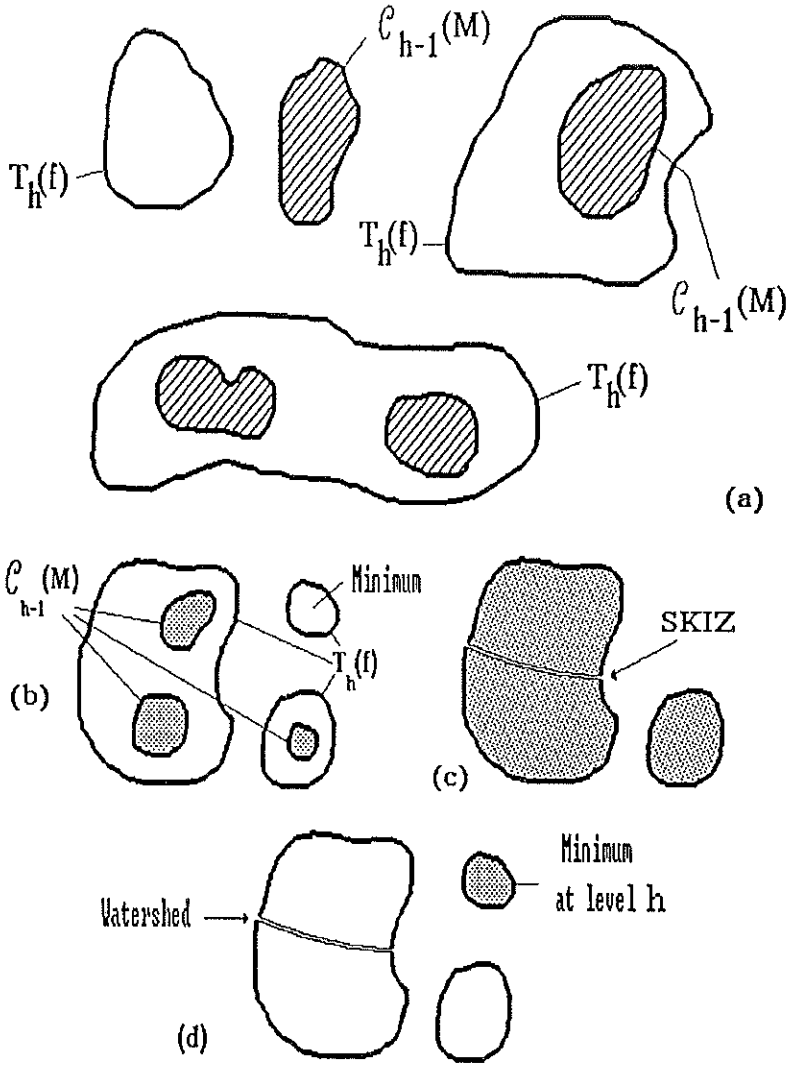


Figure 11

(a) The three possible inclusion relations of $\mathfrak{C}_{h-1}(M)$ *in* $T_h(f)$. *(b) Reconstruction of* $T_h(f)$ *from the markers* $\mathfrak{C}_{h-1}(M)$. *The particles which are not reconstructed correspond to the new minima* Min_h. *(c) Geodesic influence zones* IZ_h *of the particles* $\mathfrak{C}_{h-1}(M)$ *within* $T_h(f)$. *(d) Union of the new minima* Min_h *with the influence zones* IZ_h. *We get the new set* $\mathfrak{C}_h(M)$.

As a summary we get the following algorithmic construction of the catchment basins (algorithm 1) :

The set of the catchment basins of the grey-tone function f is equal to the set $\mathfrak{C}_{hmax}(M)$ constructed recursively :

1) $C_{hmin}(M) = T_{hmin}(f)$;

11) For $h \in [hmin+1, hmax]$: $C_h(M) = IZ_h \cup Min_h$;

with IZ_h = Zone_of_influence {Field = $T_h(f)$; Particle = $\mathcal{C}_{h-1}(M)$}

and $Min_h = T_h(f)$ / Reconstruction {Particles = $T_h(f)$; Markers = $\mathcal{C}_{h-1}(M)$ }

The divide line of f corresponds to the complement of the set $C_{hmax}(M)$, where dams have been erected. Note that $C_{hmax}(M)$ represents the strict catchment basins. By strict catchment basin we mean the attraction zones of one and only one minimum. If we add to this strict basin the divide line bordering it, we get a large catchment basin.

Remarks :

The grey-tone discretization well shows the basic construction of the watershed line. To be used on a grid, it still needs a spatial discretization.

The basic transformations to be performed level by level are the skeleton by zone of influence and particle reconstruction. The former has been presented in the preceding section. The latter can be made by the classical geodesic dilation of the markers within the particles :

Repeat until stability : Markers = (Markers $\circledast$ H) $\cap$ Particles ;

Or alternatively by a sequential algorithm presented in [16].

I-2-2) Catchment basins of imposed minima

If we come back to our flooding scheme, we will bore a hole at all places defined by a labeled set of markers (X_λ). A given set X_μ (made of one or several connected particles) will generate a region made of one or several catchment basins, identified with the label μ. The catchment basins of the minima which are not pierced, are filled up by a flood coming from the neighboring catchment basin : as soon as the water reaches the saddle point between both basins, the water rushes through the pass and fills the so far empty basin (see Fig.5). No dam is constructed between these two basins. A dam is only constructed for separating floods originating from different pierced minima. In the end, both spots and background will be covered by the flood, except the divide line which separates them.

Notations :

Let (X_λ) be the set of labeled markers. The catchment basin associated with a labeled binary set (X_γ) is then written :

$$CB(f, X_\gamma \in (X_\lambda)).$$

Each catchment basin does not only depend on its own marker but on the neighboring markers as well. The divide line is written :

$$DL(f, (X_\lambda)).$$

Algorithmic formulation (Algorithm 2) :

The function f is given, and a binary set X representing all the centers of the regions to recognize.

1) $C_{hmin-1}(M) = X$;

2) For $h \in [hmin, hmax]$:

$\mathfrak{C}_h(M)$ = Zone_of_influence { Field = $T_h(f) \cup X$; Particle = $\mathfrak{C}_{h-1}(M)$ }

The divide line of f corresponds to the complement of the set $C_{hmax}(M)$, where dams have been erected.

Remark :

The algorithm is simpler than the general one. The detection of new regional minima at each level by grain reconstruction is not necessary any more. At each level the catchment basins are extended within a field constituted by the union of the threshold $T_h(f)$ and of X. From the initialization through the end the number of particles of $C_h(M)$ remains constant and equal to the number of particles of X.

II) DETECTION OF THE DIVIDE LINES

In this section we will see how to directly detect divide lines. To do this, we have to resort to the theory of grey-tone skeletons elaborated by F.Meyer [17].

II-1) The theory of the grey-tone skeleton

II-1-1) Lower-complete, upper-complete and complete functions

As we explained earlier in the text how to close contours, we already presented lower and upper complete functions. Let's recall the definitions :

- a function is lower-complete, if each point has a neighbor of lower altitude, except the points of the regional minima.
- a function is upper-complete, if each point has a neighbor of higher altitude, except the points of the regional maxima.
- a function is complete, if it is both lower and higher complete.

We immediately remark that the flooding of a relief which starts from the regional minima induces a lower complete order : the points of the regional minima excepted, each point is flooded from a lower neighbor. The closing of contours necessitates the introduction of an upper complete relief.

A sequential algorithm for constructing a lower or an upper complete function is given in [17].

II-1-2) Divide points are particular crest points

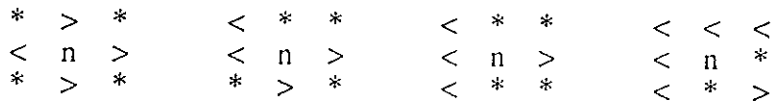
On a lower-complete or complete function one then defines a family of remarkable points belonging to the skeleton. These points are the crest points, and characterize all the zones where the relief has not a smooth regular slope. It is faster to make the inventory of the non crest points, where the relief has a gentle slope up.

On a hexagonal grid, the non crest points are the following (up to a rotation of $k\pi/3$) :

```
  >   >          >   >          >   *
* n   >  ;   *  n  *  ;   *  n   <
  <   *          <   <          <   <
```

where * means a pixel of indifferent value.

On a square grid, with 8-connectivity for the sets and 4-connectivity for the background, they are (up to a symmetry or a rotation of $k\pi/2$) :

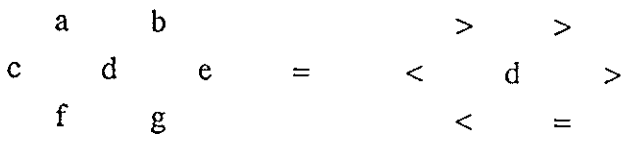


Crest points in a 8/4-connectivity.

All the other points are crest points.

II-1-3) Upstream construction

If we replace all crest points by an infinite altitude, we get a new function which is not lower-complete anymore. Local modifications permit to restore this lower completeness. On this new function, new crest points appear in the upstream of the preceding ones. The repetition of this mechanism permits to construct the complete upstream of the original crest points. More precisely, the construction of a dike places this point at an infinite altitude, which modifies the neighboring local topography. Let us consider the following configuration :



If the altitude of f becomes infinite, d has then two connected particles of higher altitude, separated by the points g and c. Hence, c and g lead to two disconnected downstream regions. On the other hand, if the altitude of c and not of f becomes infinite, the relief still belongs to a non crest type distribution. In the following figures, ordinary points are indicated in small font, crest points in bold font. Fig. 12a indicates the crest points of the initial relief. The altitude of these points then becomes infinite, which causes new crest points to emerge (Fig.12b), which become themselves infinite (Fig.12c), and the process is repeated (Fig.12d).

Thus, the skeleton gradually widens into a cone. This is due to the fact that we have not noticed that replacing certain points introduced new plateaus on the relief. Let us consider for instance the relief obtained after the first step (Fig 12b). The points of altitude 2 and noted in bold type belong to a plateau, but only the two extreme points have neighbors lower than themselves. The relief is then no longer lower complete. To render it lower complete again, we have to construct a complementary function g indicating for

each point of a plateau its distance to the border. For illustrative purpose, we multiplied the function of Fig.12b by 10 and added to the result the function g. This enables us to show the relief associated with (f,g) on one function (Fig.12d). On this new relief only the point of altitude 21 is a crest point. The resorption of the plateaus at all altitudes prevents the skeleton from widening to a cone. The next steps are shown in fig.12f, fig.12g and fig.12h.

Conventions for figures : Small characters = ordinary points of the set
Bold characters = crest points
Normal characters = other skeleton points

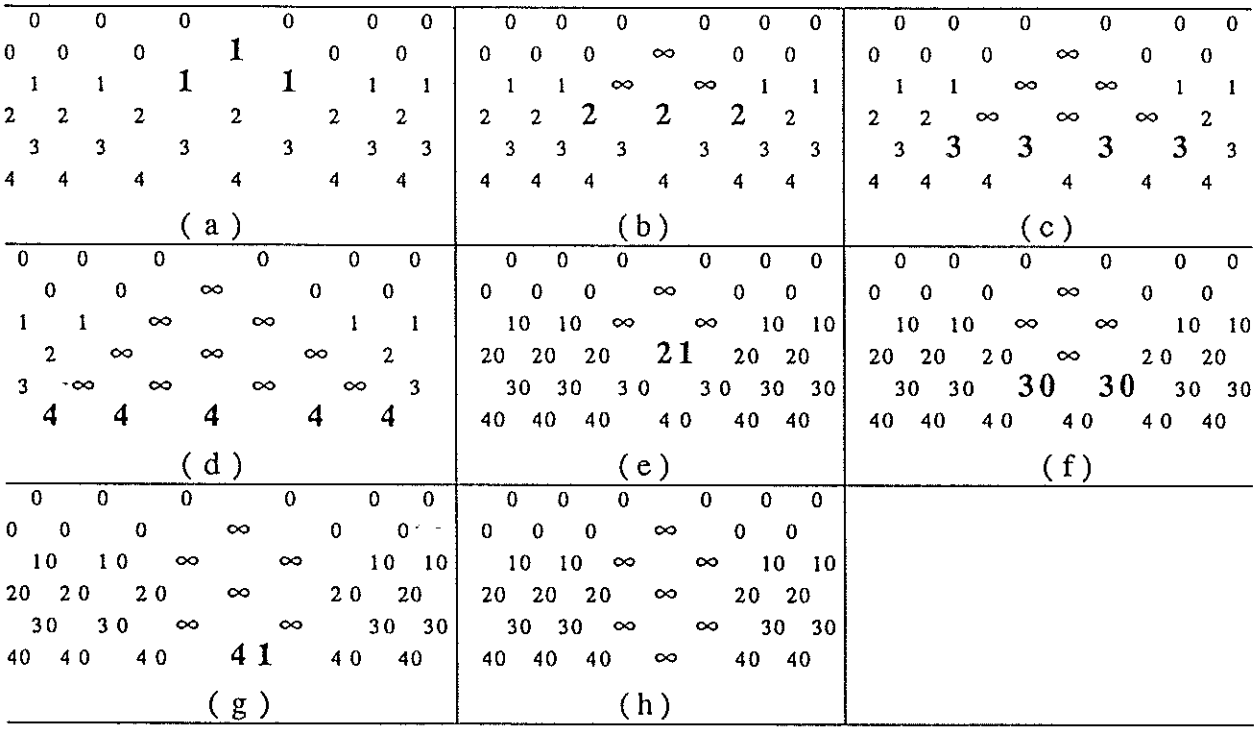


Figure 12

a) Crest-points of the initial relief. b,c,d) Previously detected crest points are given the value ∞, allowing new crest points to appear. The upstream of the initial crest points widens into a cone. e) The initial crest points are given the value ∞ and the new relief is made lower complete. A single new crest point appears on this new relief. f,g,h) The process is repeated for the upstream construction : previously detected crest points are given the value ∞, and the resulting relief is made lower complete. New crest points appear on this new relief.

Remark :

Instead of constructing a complementary relief g for the resorption of the plateaus, we may as well apply the following rules for distinguishing between border and non border points :

- Let x and y be two neighboring points of the same initial altitude. If all

neighbors of x, which were initially of lower altitude, belong at present to the skeleton, we call x a non border point. In contrast, if x still has lower neighbors which do not belong to the skeleton, then x is called a non border point.

- If x and y are both border points or non border points, they are still at the same altitude.
- If x is a border point and y a non border point, x is higher than y.

The square grid, with its double 4/8-connectivity relation introduces as always additional troubles. After the modification of the relief on skeleton points, three classes of new plateau points have to be considered :

- category a : points with a non skeleton 4-connected downstream neighbor.
- category b : points with a non skeleton 8-connected downstream neighbor but without a 4-connected non skeleton downstream neighbor.
- category c : points for which all downstream neighbors belong to the skeleton.

Three points of equal altitude on the initial relief have to be classified in the following order after relief modification on skeleton points :

category a < category b < category c

Analysis of the final result :

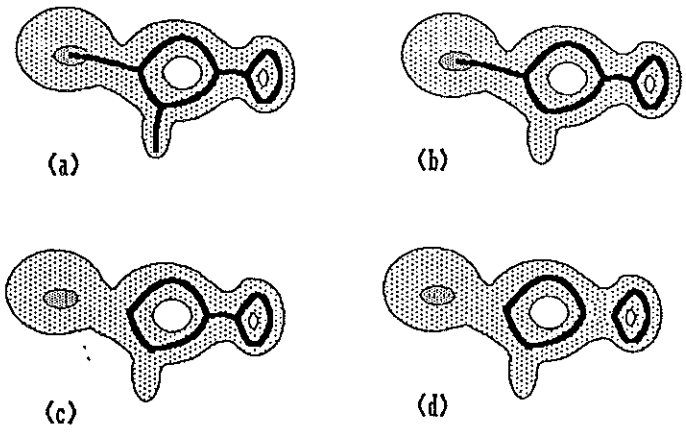


Figure 13

(a) Complete skeleton of a grey tone function. (b) Minimal skeleton. (c) Result of pruning the minimal skeleton : we have still an overset of the divide line . (d) Divide line. It can only be obtained by a regional analysis of the connectivity of the minima.

Fig.13 shows the following :

- the regional maxima of the function are linked by the skeleton (Fig.13a-b)
- the regional minima are separated by the skeleton (Fig.13a-d)
- pruning of the skeleton is not sufficient : in the present case we have still an overset of the divide line (Fig.13c).
- the skeleton is an over-set of the divide line (Fig.13d)

The question which now arises is how to suppress the spurious branches of the skeleton ? Or better, how to avoid that they appear ?

II-1-4) The minimal skeleton, an approximation of the watershed

It is possible to show that the crest points are the union of two categories of points :

- category 1 : points without upstream
- category 2 : separation points or separation couples of points : A local separation point is a point whose immediate downstream is made of two disconnected particles. In a local separation pair, each point has an upstream and a connected downstream, but within the neighborhood of the pair, these down streams are disconnected.

We have previously defined complete functions. They are characterized by the fact that each point has a lower and a higher neighbor, except the regional minima or maxima. The regional maxima are obviously without higher neighbors, hence they are crest points. Among all other points, having all a higher neighbor, the only crest points are of category 2 : divide points or divide couple of points. These points being divide points and having higher neighbors all belong to saddle zones of the relief. let's now turn this fact to the best account for constructing the divide line :

- construct a complete relief
- detect the crest points (regional maxima and saddle zones).
- construct the upstream of these points.

The result of the construction is called minimal skeleton.

II-1-5) Difference between minimal skeleton and divide line

The minimal skeleton is formed by taking the upstream of the saddle points. Nothing guarantees that both downstream directions starting from the

pass reach two different regional minima. It is mostly the case on real images. There are however some pathological situations where it is not true.

II-2) Complete functions for contour closing

We have seen that the construction of the minimal skeleton is made on a complete function. There are several ways to obtain complete functions :

- one first constructs a lower complete function, for which new possible regional maxima appear, but no regional minima. The resulting relief is then made upper complete. The construction of the minimal skeleton on this function produces an over-set of the divide line corresponding to situation 1.
- If one reverses the order of operations. One first constructs an upper complete function. New regional minima may then appear. The result is made lower complete. The detection of the minimal skeleton on this new relief produces an over-set of the divide line of situation 3. It allows to close broken contours. This method has been used in the series of Figures 6a-6d.

II-3) Other approximations of divide lines : grey-tone skeletons and pruning

The most common approximation on hardwired machines of the divide line consists in constructing a grey-tone skeleton using thinning transformations (see Serra [18] for grey-tone thinnings). The resulting skeleton is an over-set of the divide line with many spurious branches. All these branches have in common the possession of a loose end. It is possible to detect and recursively suppress these loose ends. This operation is called pruning. After convergence the result is an acceptable approximation of the divide line.

III) CONSTRUCTION OF THE WATERSHED LINE WITH IMPOSED MINIMA : HOMOTOPY MODIFICATION.

III-1) How to impose new minima upon a function ?

In the previous sections we presented the basic construction algorithms of the divide line. Given a function f, Algorithm 1 associates a catchment basin with each of its regional minima, whereas algorithm 2 associates a catchment basin only with the connected components of a binary set X. However, in some hardware environments, only algorithms of type 1 are available. Then,

the following question immediately arises : is there a function f', for which algorithm 1 gives the same catchment basins as algorithm 2 applied to the couple (f, X). The answer is yes. The set of the regional minima of f' will be identical to the set X. f' may be obtained by the following iterative algorithm :

- Initialization : f' = Max on X^c
  0 on X
- Iterative construction :

f' = Max(f, f'⊖ H) ; /* minimum function between f and the result of an erosion of f' by a disk of size 1 */

The following figure illustrates this construction on the profile of a function :

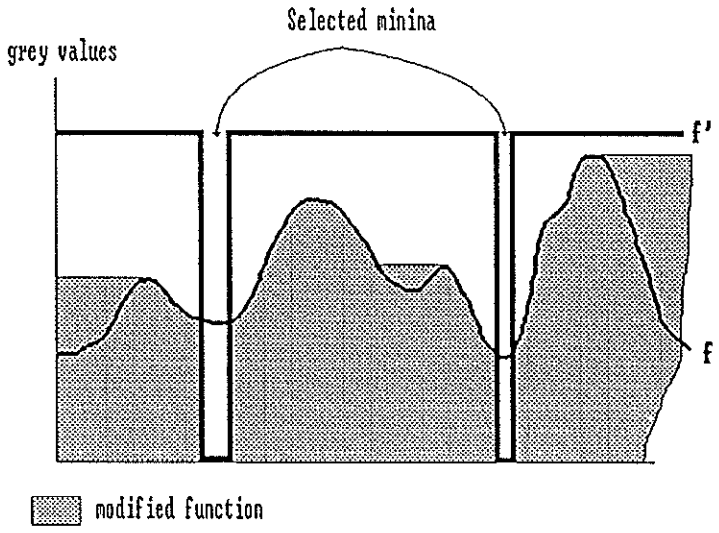


Figure 14

Image f, where all regional minima outside a set X are to be suppressed. Image f' equal to 0 within X and to Max outside X. Result of the preceding algorithm (dotted area) : all unwanted regional minima of f' are suppressed.

III-2) Interpretation as a grey-tone geodesic erosion

The operation f' = Max (f, f'⊖ H) can be interpreted as a grey-tone geodesic erosion [19] of size 1 of the function f' constrained to remain above the function f.

We first define an f-upper-admissible path of length l and of height H between two points x and y : it is a (l + 1)-uplet of pixels (x_0 = x, x_1, ...,x_{l+1} = y) such that :

- x_i and x_{i+1} are neighbors

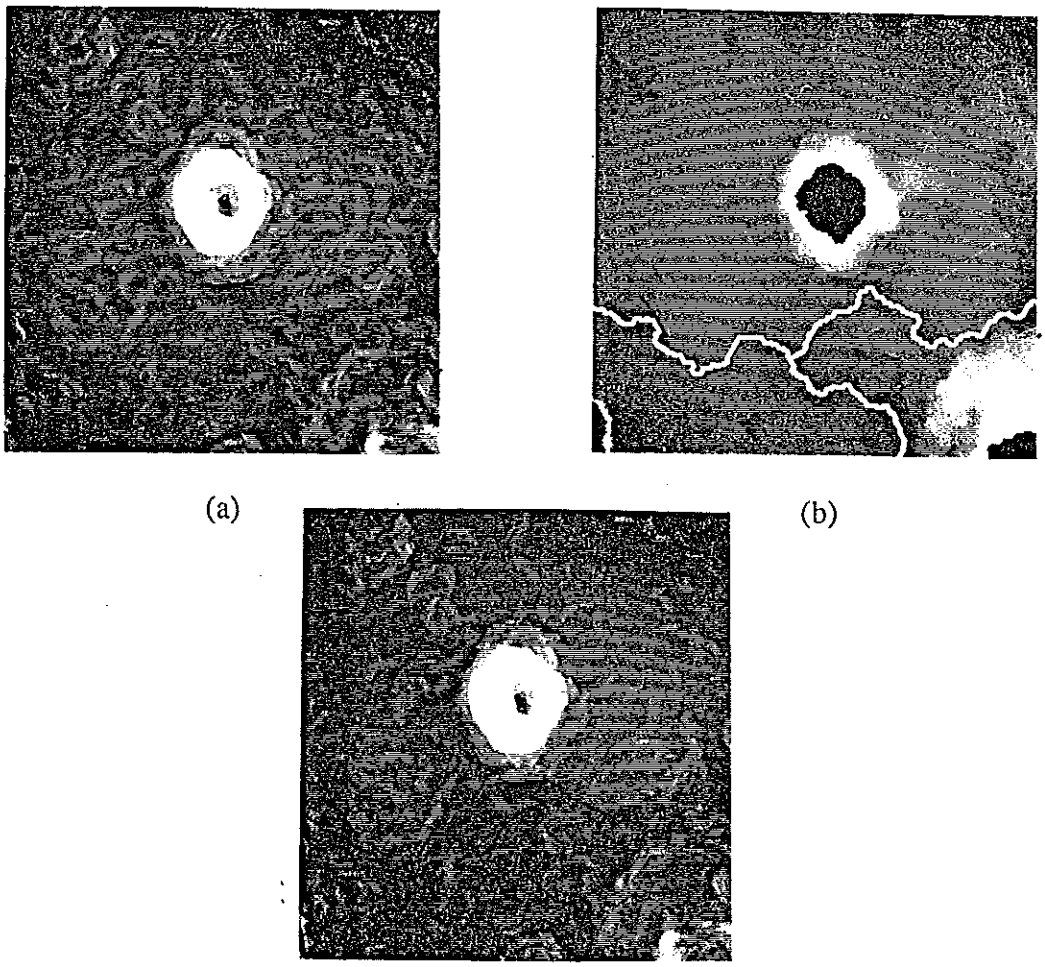
- $f(x_i) \geq H$ for any x_i.

The union of all f-upper-admissible paths of size l and of origin x constitute the f-geodesic ball of size l and center x. The geodesic erosion of size l of the function g above the function f at a point x is the smallest value of the intersection of the upper-graph of g with the f-geodesic ball of size l centered at x.

The function f' is then the erosion of infinite size of the function f'_0 (defined supra) above the function f.

A sequential algorithm for the construction of f' is presented in [16].

III-3) Filtering properties of this geodesic erosion



(a) (b)

(c)

Picture 21

(a) Original grey-tone image. (b) Binary sets of the wanted minima superimposed on the grey-tone image. (c) Grey-tone image with the imposed minima.

The connected components of X are now the regional minima of f'. All other regional minima of f have been filled. It results in a strong filtering effect.

If we consider $\tilde{f}$= Min(f'_0 , f). Then f' is a closing of $\tilde{f}$ (extensive, increasing and idempotent operation). We know that closings are strong morphological filters.

Remark :

Only the regional minima of f have changed. All regional maxima outside the set X are preserved. For this reason, we say that this operation changes the lower homotopy of the function f. The notions of lower and upper homotopy are extensively discussed in [17].

IV) CONSTRUCTION OF THE CATCHMENT BASINS BY UPSTREAM CLIMBING

IV-1) Some problems related with the discretization

The algorithm based on the flooding of the relief, level by level may be very efficient if one has a random access to all the points in the image, and will be relatively inefficient otherwise (the problem of the speed is different : an inefficient algorithm may be very fast on dedicated hardware). Indeed a parallel neighborhood transformation analyzing every point would be particularly inefficient since a point may only be assigned to a catchment basin when the flood reaches its level. The random access to all the points of the image greatly improves the performances. L.Vincent et al. have published a particularly efficient implementation [20], with the following features :

- in a first scan, the histogram of the image is computed and the addresses of all the points of the image are stored in increasing order (by insertion sorting).
- the flooding proceeds level by level. Since the points have been ordered, at any moment the algorithm addresses only the points of the current level to be flooded. The use of a FIFO in order to achieve a breadth first search within the plateaus further increases the efficiency.

However, random access is not always possible or may be very much time consuming. Imagine very large grey-tone images stored on hard disk. Since flooding is made level by level, one will have to shuffle the image back and forth between RAM memory and disk as many times as there are levels in the

image. Therefore, one may wonder whether it is possible or not to adopt another strategy ; propagate the information that a point belongs to a given catchment basin to its upstream neighbors, and so on recursively. This labeling of each basin would proceed by a hill climbing and not level by level. A first attempt has been published (Friedlander et al [21], Arcelli [22]) with some drawbacks. The major problem is that the upstream of a point soon becomes very large. Fig.15 shows several upstream paths on a regular relief all having the same slope.

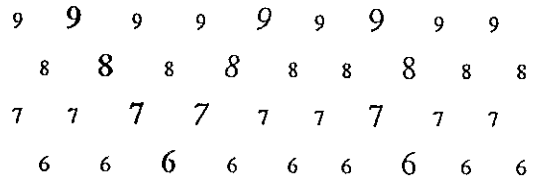


Figure 15

Non unicity of the lines of steepest descent

Three trajectories are indicated in normal, italic and bold fonts. All three follow the lines of swiftest descent, with the same slope (at least by using the hexagonal distance associated with the grid). With as many possibilities, the upstream of a point will be much too large. Hence the catchment basins will be too large as well and will highly overlap. This is illustrated by the following figure where we have two regional minima of value 0. The catchment basin which is exclusively associated with the left (resp. the right) one is indicated in standard (resp. italic) characters. The area in bold characters is the overlapping zone.

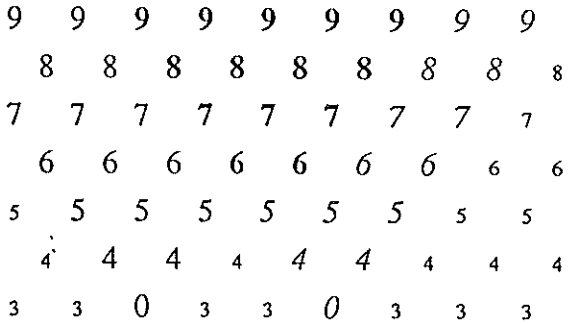


Figure 16

a straightforward definition of the catchment basin leads to excessively thick divide zones

For this reason, Friedlander firts defines a broad catchment basin : it is the set of the pixels that can be reached by following a never descending path starting from a regional minimum. The zone where two or more broad catchment basins overlap is referred to as a "watershed zone". Its complement constitutes the "restricted catchment basins". Finally, the catchment basins themselves are obtained via the SKIZ of the restricted catchment basins. This last step is an approximation which may be very crude. The algorithm presented in the next section avoids this drawback and is due to F.Meyer.

IV-2) Steepest slope upstream climbing

We have to find an upstream construction, which follows a line of steepest slope on a digital grid, although these lines are poorly defined on grids. Furthermore, we would like that the upstream of a point does not widen into a cone (except in some cases where we have non unicity of the upstream like in a greek amphitheater). The algorithm for the construction of the grey-tone skeleton detailed previously yields such an upstream construction. Indeed, once the crest points of the initial relief are found, the upstream construction completes the skeleton by following lines of steepest slope and without widening into a cone. If we use the same upstream construction for propagating the labels of the regional minima upwards, we obtain a perfectly efficient construction of the catchment basins. It may be done by a sequential algorithm. Alternatively, all potential upstream points of each labeled point may be stored in a FIFO. In this way, the upstream of a point is completely constructed before another point is considered.

REFERENCES

1. S.BEUCHER and C. LANTUEJOUL : Use of watersheds in contour detection, in *Proceedings, Intern. workshop on image processing, CCETT/IRISA, Rennes, France,* Sept. 1979.
2. F.MEYER : sequential algorithms for cell segmentation : maximum efficiency?, in *Proceedings, Intern. symposium on clinical cytometry an d histometry, Schloss Elmau,* April 1986.
3. S.BEUCHER : *analyse automatique de gels d'électrophorèse bi-dimensionnelle et morphologie mathématique, Internal note, Centre de*

Morphologie Mathématique, Fontainebleau, France, Feb. 1982.

4. F.MEYER : The perceptual graph : a new algorithm, in *Proceedings ICASSP 82, IEEE Internat. Conf. on Acoustics, Speech and Signal Processing, Paris,* May 1982, 1932-1935.
5. S.BEUCHER : *Segmentation d'images et Morphologie Mathématique, Doctorate Thesis, School of Mines, Paris (in press).*
6. S.BEUCHER and F.MEYER : *Méthodes d'analyse des contrastes à l'analyseur de textures, Internal note, Centre de Morphologie Mathématique, Fontainebleau, France,* Sept. 1977.
7. F.MEYER : Contrast features extraction, in *Proceedings, 2nd European Symposium on Quantitative analysis of microstructures in Material Sciences, Biology and Medecine, Caen,* Oct. 1977, 374-380.
8. F.FRIEDLANDER : *Le traitement morphologique d'images de cardiologie nucléaire, Doctorate Thesis, School of Mines, Paris,* Dec. 1989.
9. J.L.BARAT, J.P.COLLE, D.COMMENGES, A.J.BRENDEL, J.OHAYON, V.MAGIMEL PELONNIER, F.LECCIA, P.BESSE, D. DUCASSOU : Reproductibilité et valeur diagnostique comparatives dans l'analyse quantitative de la fonction ventriculaire gauche par tomographie d'émission gamma et angiographie de contraste, in *J. Biophysique et Med. Nucl.,* **8** (2-3), 1984.
10. J.SERRA : *Image Analysis and Mathematical Morphology, Vol. 2 : Theoretical Advances,* Academic Press, London, 1988.
11. S.BEUCHER, J.M.BLOSSEVILLE, F.LENOIR : Traffic spatial measurements using video image processing, in *Proceedings, SPIE, Advances in Intelligent Robotics Systems, Cambridge Symposium on optical and Optoelectronic Engineering, Cambridge, Mass.,* Nov. 1987.
12. L.VINCENT : Graphs and mathematical morphology, in *Signal Processing, special issue on Math. Morph.,* **16** (4), April 1989, 365-388.
13. S.BEUCHER, M.BILODEAU, X.YU : Road segmentation by watersheds algorithms, in *Proceeding, PROMETHEUS workshop, Antibes, France,* April 1989.
14. A.ROSENFELD and A.KAK : *Digital Picture Processing, 2nd edition, Vol. 2,* Academic Press, 1982.
15. F.MEYER : *Séparation de particules volumiques imbriquées, Internal note, Centre de Morphologie Mathématique, Fontainebleau, France,* Nov. 1989.
16. F.MEYER : Algorithmes séquentiels, in *Proceedings, Onzième Colloque GRETSI, Nice, France,* 1987, 543-546.
17. F.MEYER : Skeletons and perceptual graphs, *Signal Processing,* **16**(4), April 1989, 335-363.
18. J.SERRA : *Image Analysis and Mathematical Morphology* , Academic Press,

London, 1982.

19. S.BEUCHER : *Spring School in Mathematical Morphology, Centre de Morphologie Mathématique, Fontainebleau, France,* 1980.
20. P.SOILLE and L.VINCENT : Watersheds in digital spaces : *an efficient algorithm based on immersion simulations, Internal note, Centre de Morphologie Mathématique, Fontainebleau, France,* 1990.
21. F.FRIEDLANDER and F.MEYER : A sequential algorithm for detecting watersheds on a grey level image, in *Proceedings, 7th Internat. Conf. on Stereology,ICS77, Caen, France,* 1987, 663-668.
22. C.ARCELLI and G.SANNITI DI BAJA : Computing Voronoi diagrams in digital pictures, *Pattern Recognition Letters,* 4(5), 1986.