# Morphological multiscale and interactive segmentation

Fernand Meyer
Ecole des Mines de Paris
Centre de Morphologie Mathématique
35, Rue Saint Honoré
F-77305 Fontainebleau FRANCE

## Abstract

In the context of object oriented coding, new demands on segmentation have appeared, like multiscale segmentation and video-objects editing. The task is particularly difficult as the type of image content is not known a priori.

We present old and new morphological algorithms for segmentation in a unified framework : flooding a topographic surface. This permits to introduce the classical segmentation paradigm like the watershed, but also new multiscale segmentation methods with good psychovisual properties.

## 1 Introduction

Image segmentation in mathematical morphology is essentially based on one method: the watershed of a gradient image from a set of markers. By marker, we mean a binary set included in the object of interest ; its exact location or shape has no importance. The watershed construction will grow the markers until the exact contours of the objects have been found. The strategies for finding good markers are extremely diverse and problem dependant In all cases listed in [1] [8], the objects to be segmented belonged to a well defined class, with constant features and were to be found in a known environment. This made the detection of markers feasible, if not always easy.

In the last years there has been a growing demand for segmentation in the domain of telecommunications and more generally in the multimedia industry. The development of games, teleworking, teleshopping, television on demand, videoconferences etc has multiplied situations where images and sequences have not only to be transmitted but also manipulated, selected, assembled in new ways. This trend is well illustrated by the new ISO norm MPEG-4, which allows to compress but also to manipulate the content of video and audio sequences. The norm allows to represent the shape of the video-objects present in the sequence but does offer no means for segmenting or tracking them in a sequence. Furthermore, in some cases, like in games, the scenes may be segmented beforehand, but in other situations, the receiver of the information may want to select and track an arbitrary object in the sequence, requiring real time segmentation. This evolution is most challenging for segmentation techniques : one has to segment complex sequences of colour images in real time, be automatic but also able to deal with user interaction

Object oriented coding represents an even greater challenge for segmentation techniques. Such encoders segment the scene into homogeneous zones for which contours, motion and texture have to be transmitted. Depending upon the targeted bitstream and the complexity of the scene, a variable number of regions may be transmitted [4]. Hence an automatic segmentation with a variable number of regions is required for sequences for which the content or even content type is not known a priori. Hence, there is no possibility to device a strategy for finding markers, and as a consequence the traditional morphological segmentation based on watershed and markers fails.

This situation has triggered the development of new techniques of multiscale segmentation, where no markers are required. When going from a finer to a coarser scale fusions of adjacent regions take place, resulting in a coarser partition. In this paper we analyse both traditional and multiscale segmentation in the common framework of flooding a topographic surface.

# 2 Uniform flooding of a topographic surface : the watershed transform

## 2.1 Grids, connectivity, flat zones, regional minima and maxima

We consider here images in a digital framework. Images are represented on regular graphs where the nodes represent the pixels. The edges of the graph define the neighboring relations : an edge connects each pair of neighboring pixels. With the help of these neighboring relations, we may define flat zones or plateaus of functions

**Definition 1** *Two pixels $x$,$y$ belong to the same flat-zone of a function $f$ if and only if there exists a n-tuple of pixels $(p_1, p_2, ..., p_n)$ such that $p_1 = x$ and $p_n = y$, and for all $i$, $(p_i, p_{i+1})$ are neighbors and have the same grey-tone: $f_{p_i} = f_{p_i+1}$.*

Each neighboring pixel of a flat-zone has an altitude which is either higher or lower than the altitude of the flat-zone. Ordinary plateaus have lower and higher neighbors.

**Definition 2** *A regional minimum (resp. maximum) is a flat-zone without lower (resp. higher) neighbors*

## 2.2 The topographic distance function

Let us now consider a drop of water falling on a topographic surface $f$ for which the regional minima are the only plateaus. If it falls outside a plateau, it will glide along a path of steepest descent. If the altitude of a pixel $x$ is $f_x$, the altitude of its lowest neighbor defines the erosion $\varepsilon f_x$ of size 1 at pixel $x$. Hence the altitude of the steepest descending slope at pixel $x$ is $\text{slope}(x) = f_x - \varepsilon f_x$. If $x$ and $y$ are two neighboring pixels, we will define the topographic variation $\text{topvar}(x, y)$ between $x$ and $y$ as $\text{slope}(p)$ if $f_x \neq f_y$ and as $\frac{\text{slope}(x)+\text{slope}(y)}{2}$if $f_x = f_y$.

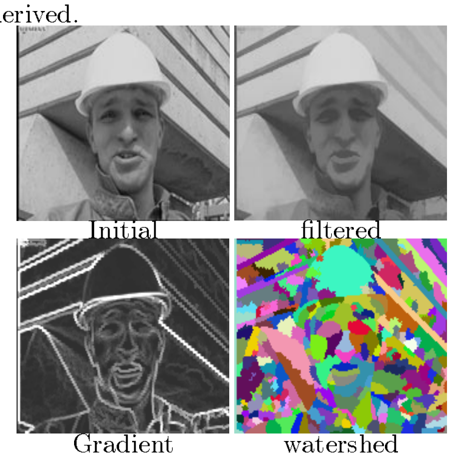
If $\pi$ is a path $(x = p_1, p_2, ..., y = p_n)$ between two pixels $x$ and $y$, we define the topographical variation along the path $\pi$ as the sum $\sum_{i=1,n-1} \text{topvar}(p_i, p_{i+1})$ of the elementary topographical variations along the path $\pi$. The topographical distance between two pixels $x$ and $y$ is defined as the minimal topographical variation along all paths between $x$ and $y$. By construction the trajectory of a drop of water falling on the surface is a geodesic line of the topographic distance. If two pixels $p$ and $q$ belong to such a geodesic line, the topographic distance between these two pixels is equal to the difference of altitudes between both pixels : $\mid f_p - f_q \mid$. A pixel $p$ belongs to the upstream of a pixel $q$ if and only if the topographic distance between both pixels is equal to $\mid f_p - f_q \mid$. The topographic distance between all other couples of pixels is higher than the difference of altitude between these pixels.

Let us now transform the topographic surface by putting all regional minima at altitude 0.

**Definition 3** *We call catchment basin* $\text{CB}(m_i)$ *of a regional minimum $m_i$ the set of pixels which are closer to $m_i$ than* *to any other regional minimum for the topographical distance*

A more general description of the topographic distance, also valid for images with plateaus may be found in [6]. Within each catchment basin, the set of pixels closer to the minimum than a given topogaphic distance $h$ are all pixels of this basin with an altitude below $h$. There exist extremely efficient shortest path algorithms on graphs which may be applied on the grid, in order to construct the catchment basins of all the minima. The most used uses the shortest path algorithm of Moore [9]. The aim is to compute the shortest path form the minima to all other pixels. The pixels for which the length of the shortest path is known are ordered according to this length. The node with the lowest value is expanded and the shortest paths of its neighbors computed. Its implementation is particularly efficient using hierarchical queues [5], with the additional advantage to flood correctly the plateaus: the flood progresses from descending borders of the plateaus inwards with uniform speed. Initially the minima are labeled and the hierarchical queues permit to perfectly mimic the flooding of the topographic surface with a flood of uniform altitude ; during the flooding the label of each minimum is propagated to the corresponding catchment basin.

In practice, we use a biased algorithm in order to create a partition, the pixels which are at an equal distance of two minima are assigned arbitrarily to one of the adjacent catchment basins. As a result of this choice, the union of the catchment basins forms a partition. This partition is the finest partition from which all other segmentations will be derived. In the future, the smallest entities we are dealing with are the catchment basins and not the pixels. This process is illustrated by the following figures. On the top row we see the original image ; applying a morphological filter called leveling [7] produces a simplified version, where the flat zones have been enlarged. The modulus of the gradient image is then computed for the filtered image and serves as topographic surface. The catchment basin of the gradient image serve as finest segmentation from each coarser segmentations will be derived.




## 2.3 The neighborhood graph

The neighborhood relations between catchment basins are best summarized by the neighborhood graph:

- the nodes of the graph represent the catchment basins of the topographic surface
- an edge links two nodes of the graph if the corresponding catchment basins are neighbors

Fig.1a represents a topographic surface which is currently flooded. Along the lines where two flooding fronts would merge, a dam is erected, materialising the watershed line. In fig.1b we see on one hand the catchment basins with their labels and limits, and on the other hand a graph structure summarizing their relationships : the grey round dots represent the minima ; they are linked by thin lines, representing the edges of the neighborhood graph. The weight of the edge measures the difficulty of crossing the border. As simplest weight, one may take the altitude of the lowest pass point between two adjacent catchment basins. This weight distribution permits to represent the propagation of the flooding of the topographic surface from any set of sources.

Other weight distributions are possible. As an example one may chose the mean altitude along the watershed line on the common boundary between adjacent catchment basins. In case of colour images, the modulus of a colour gradient can be chosen.

More generally, the weights express a dissimilarity between adjacent tiles of a mosaic. This dissimilarity is expressed by a symmetric binary relation. This formulation permits generalisations for instance to colour or moving images.

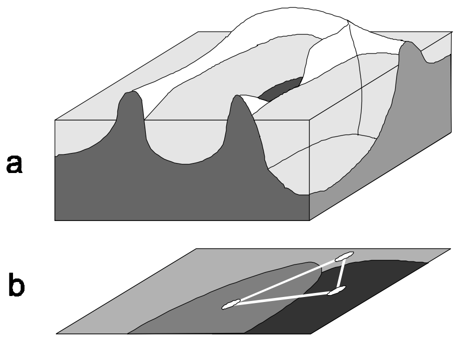



Figure 1: a) Construction of the watershed line of a topographic surface by flooding.
b) Creation of the associated neighborhood graph.

## 2.4 The flooding dendrogram

The watershed construction has created a fine partition on the topographic surface, where each tile is a catchment basin. Flooding of the topographic surface also produces the most natural hierarchical segmentation. Let us grow a flood with uniform altitude over the complete topographic surface (it may be seen as boring a hole in each regional minimum and plunging the topographic surface in water ; this ensures a flood with uniform altitude). As the level of the flood increases, new lakes appear in the minima and existing lakes merge, until the whole surface is flooded. The evolution of the lakes, paritions and fusions may be summarized in a tree structure. The minima, where each lake appears for the first time are the terminal nodes of the tree. Every time two previously disconnected lakes merge, a new node is created, representing the new lake and linked by two edges to the two merging lakes. This is illustrated by fig.2

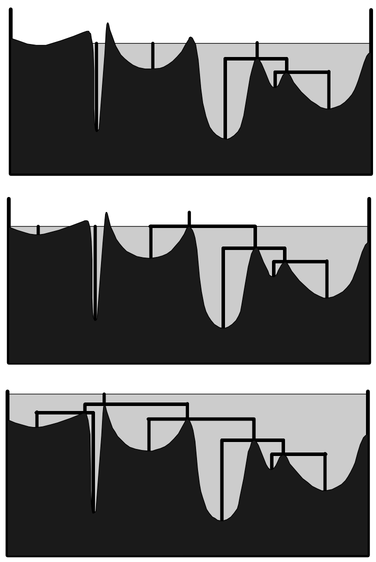


Figure 2: Construction of the flooding dendrogram. Top: three lakes exist. Center: two lakes merge, creating a new node. Bottom : complete dendrogram.

The associated multiscale segmentation is easy to obtain : for each level of flood, all catchment basins touched by a common lake are merged. As the level of the flood increases, new lakes merge. This implies that from a finer segmentation to a coarser, only fusions of regions occur. Since these fusions are illustrated by a dendrogram, the associated series of partitions is called segmention tree. The next section shows that all information about the segmentation tree is contained in the minimum spanning tree of the neighborhood graph.

## 2.5 The flooding ultrametric distance and the minimum spanning tree

The minimum level of flood for which two catchment basins are touched for the first time by a common lake constitutes in fact an ultrametric distance. An ultrametric distance is defined by the following axioms :

* reflexivity : $d(x,x) = 0$
* symmetry: $d(x,y) = d(y,x)$
* ultrametric inequality : for all $x, y, z$: $d(x,y) \leq \max\{d(x,z), d(z,y)\}$

The first two axioms are obviously verified. The last one may be interpreted as follows ; the minimum altitude of the lake containing both catchment basins $x$ and $y$ is lower than or equal to the altitude of the lowest lake containing all three catchment basins $x$ and $y$ and $z$. The level of this last lake is precisely $\max\{d(x,z), d(z,y)\}$.

As the weights of the edges in the neighborhood graph represent the altitude of the pathpoint between adjacent catchment basins, the interconnection of lakes as the level of the flood increases is perfectly represented on the neighborhood graph. Between two nodes $x$ and $y$ there are many paths on the graph ; each of them has a highest edge, called sup-section of the path. Consider now the level $h$ of the lowest lake containing $x$ and $y$ ; it contains entirely all paths between $x$ and $y$ for which the highest edge has a valuation $h$. These paths are the paths of lowest sup-section. If all edges of the neighborhood graph have different weights, then these paths reduce to only one. The union of all paths of lowest sup-section between any couple of nodes constitutes the minimum spanning tree of the neighborhood graph.

The distance between two nodes is computed as follows : there exist between the two nodes a unique path on the MST. The highest weight encountered on this path is by definition the flooding distance between both nodes.

The closed balls of the ultrametric distance precisely correspond to the segmentation tree induced by the minimum spanning tree. The balls of radius 0 are the individual nodes, corresponding to the catchment basins. Each ball of radius $n$ is the union of all nodes belonging to one of the subtrees of the MST obtained by cutting all edges with a valuation higher than $n$. A closed ball of radius $R$ and centre $C$ is the set of nodes which belong to the same subtree of the MST, obtained by cutting the edges at altitude higher than or equal to $R$ and containing $C$. Obviously replacing the centre $C$ by any other node of the subtree yields the same subtree.

The radius of a ball is equal to its diameter, i.e. to the largest distance between two elements in the ball, which is equal to the altitude of the highest edge in the corresponding subtree. Flooding at a lower altitude would cut the subtree into smaller subtrees. Given a partition $P_k$ in $k$ classes, the diameter of the class $C_i$ is defined as the maximal distance between two elements of $C_i$ . In our case it represents the height above which $C_i$ would be disconnected in two parts. One may then define the diameter of the partition $P_k$ as the maximal diameter on its classes. The partition in $k$ classes obtained by cutting the $(k-1)$ highest edges of the minimum spanning tree is the partition in $k$ classes for which the diameter of the partition is minimum

Cutting the $(k-1)$ highest edges of the minimum spanning tree creates a forest of $k$ trees. This is the forest of $k$ trees of minimal weight contained in the neighborhood graph. In many situations, the segmentations obtained like that are not extremely useful. A first solution for obtaining useful segmentations is to use segmentation seeds called markers (see section 4). A second solution is to use other ultrametric distances than the flooding distance (see section 3)

# 3 Synchronous flooding of a topographic surface

The MST represents the set of paths along which any type of flooding progresses: if a source is placed at any minimum, the flood will invade progressively the complete topographic surface, always following the edges of the MST. The edges of the minimum spanning tree represent the path points on which the floods coming from different sources will meet for the first time. These points also will play an important role, if one floods the topographic surface according to some other progression rules. In particular, we will place sources at each minimum and flood the surface in such a way tha all lakes share some characteristics: same depth (contrast criteria) , same area (size criteria) or same volume (size and contrast criteria). This is called synchronous flooding. As the flooding proceeds, the level of some lakes cannot grow any further, as the level of the lowest path point has been reached. In the figure 3, the flooding is synchronized by the depth of the lakes. We have represented 4 successive levels of the flood. In the first image, the rightmost catchment basin contains two lakes, the smallest of which being absorbed by the catchment basin of the biggest. The successive images contain respectively 4, 3, 2 and 1 catchment basins.

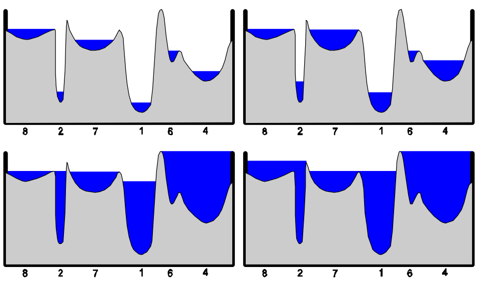


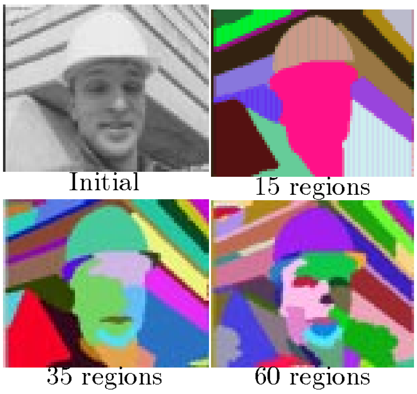

Figure 3: Four levels of height synchronous flooding

It is possible to organise all kinds of synchronous floodings : for instance, the lakes share the same volume or the same area. To each type of synchronous flooding corresponding different fusions of neighboring catchment basins. We remark that any stage of synchronous flooding represents a closing of the topographic surface. For the depth, it will be h-closing [3], for the area, area closings [12] and for the volume, volumic closings [11]. M.Grimaud [2] was the first to study the absorption of catchment basins for depth criteria ; C.Vachier [10] extended his work for area and volume criteria.

## 3.1 Absorption of catchment basins during synchronous flooding

When the water reaches the lowest path point of a catchment basin, it stops ; the resulting lake then entirely belongs to the neighboring catchment basin. At this very moment, the size (depth, area, or volume, depending upon the type of flooding) of the smallest lake (say A) may be ascribed to the minimum of the disappearing catchment basins and also to the path point, i.e. the edge of the MST, which is the link to the absorbing catchment basin (say B). Furthermore, a relation may be defined : lake A is absorbed by lake B. After completion of the flooding, each minimum and each edge of the MST will have a weight. These weights permit to define a new ultrametric distance, associated to the particular type of flooding which has been used. Hence we will have a depth, area and volume ultrametric distance. As previously, the distance between two nodes of the MST is defined as the weight of the highest edge alon the unique path linking both nodes.

The volumic ultrametric distance has particularly good psychovisual properties : the resulting segmentation trees offer a good balance between size and contrast. as illustrated in the following figures. The topographical surface to be flooded is a colour gradient of the initial image (maximum of the morphological gradients computed in each of the R, G and B colour channels). Synchronous volumic flooding has been used, and 3 levels of fusions have been represented, corresponding respectively to 15, 35 and 60 regions.




### 3.2 Depth, area and volumic distance measurements through uniform flooding

Synchronous flooding is not as easy to program as uniform flooding. For this reason, we also developed a method for measuring depth, area or volume distances during uniform flooding. During uniform flooding the events of interest are the merging of neighboring lakes, forming so called critical lakes. Just before merging, a measurement (following the choice made beforehand ; depth, area or volume) is performed on each of the lakes and the smallest measurement is assigned to the critical lake which is formed. But as there is a one to one correspondance between the critical lakes and the edges of the MST, the measurement performed on the smallest lake may be assigned to the corresponding edge of the MST (the edge through which the junction between both lakes occur). If labels are given to the minima, these labels may be propagated in the dendrogram with the following rules : when two lakes merge, the largest gives its label to the new critical lake. The measure made on the smallest is assigned to all smaller lakes with the same label. The resulting label distribution on the dendrogram is illustrated for depth measurements in fig.4.

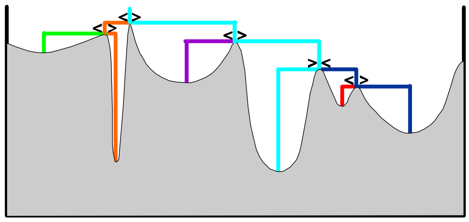


Figure 4: Labellisation of the dendrogram based on the depth of the lakes.

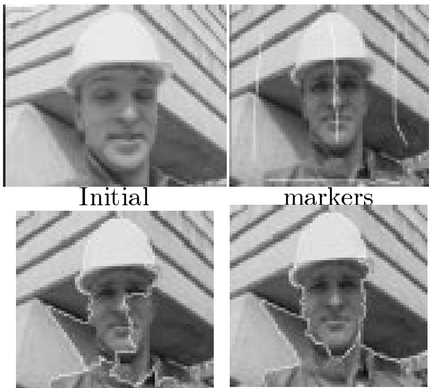
## 4 The watershed from markers

### 4.1 Minimum spanning forests rooted in markers

Starting from a partition, we have represented it as a neighborhood graph. Its minimum spanning tree constitutes a first way to generate a segmentation tree, by cutting all edges above some thresholds and merging all regions of the fine partition covered by a subtree, we obtain the successive partitions of the segmentation tree. As the segmentations associated to the flooding ultrametric distance favour too much the contrast of the regions and offer a bad balance between contrast and size, we have developed new means to assign weight distributions to the edges of the MST which produce segmentation trees with a better psycho-visual quality.

In many situations one has a seed for the objects to segment. It may be the segmentation produced in the preceding frame when one has to track an object in a sequence. It may also be some markers produced by hand, in interactive segmentation scenarios. As a result, some nodes of the minimum spanning tree may be identified as markers. The resulting segmentation associated to these markers will then still be a minimum spanning forest, but constrained in that each tree is rooted in a marker. The algorithm for constructing the minimum spanning forest is closely related to the classical algorithms for constructing the MST itself (see ref??. for more details). Each marker gets a different label and constitutes the initial part of a tree. The edges are ranked and processed in increasing order. The smallest unprocessed edge linking one of the tree $T$ to an outside node is considered ; if this node does not already belong to another tree, it is assigned to the tree $T$. If it belongs to another tree, the edge is discarded and the next edge is processed.

Segmenting with markers constitutes the classical morphological method for segmentation. However, in the literature, only the flooding ultrametric distance has been used [8]. With the multiscale approach described in this paper, we have at our disposition new ultrametric distances which better represent the relative importance of the various structures in an image. This is illustrated by the following figures, where the same set of markers has been used alternatively with the flooding distance and with the volumic distance. The superiority of the volumic distance clearly appears here : it correctly detects the face, whereas the flooding distance follows the boundary of a shadow and cuts the face in two.



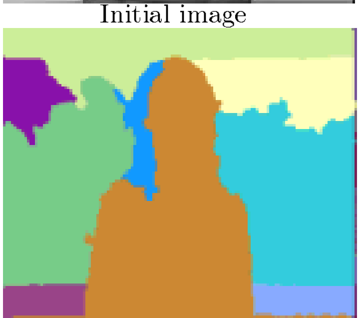
Initial markers

flooding dist. volumic dist.

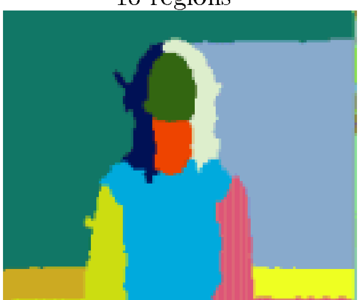
## 4.2 Interactive segmentation with the segmentation tree

Besides the traditional segmentation technique based on markers, new interactive segmentation techniques may also be developed. A toolbox for interactive editing is currently constructed in our lab (F. Zanoguera) , based on the hierarchy of segmentations. A mouse position is defined by its x-y coordinates and its depth in the segmentation tree. If the mouse is active, the whole tile containing the cursor is activated. The following pictures illustrate the possibilities of such a device. Starting from the segmentation with 10 regions, the person in foreground has been resegmented in 6 regions whereas the background has been merged in 4 regions. Fusions are obtained by using a coarser level of the pyramid, resegmentations by using a finer level

Initial image



10 regions



10 regions

The same type of techniques may be used in object oriented coding. The objects of interest may be segmented and represented with a finer degree of detail than the background for instance.

## 5 Conclusion

Considering an image to segment as a topographic surface has not only a high pedagogical interest, it also is the source of new multiscale segmentation techniques. The image of flooding has traditionally served to present the watershed transform Synchronous flooding offers new ways to merge neighboring regions ; according the criteria which are chosen, one can focus the segmentation on the contrast, on the area. Considering the volume of the lakes during flooding seems to offer a particularly good balance between size and contrast of the objects chosen at each level of the segmentation. The watershed transform may be used once at the beginning of the process for the construction of the finest segmentation. The resulting tiles are represented in a weighted graph structure. Moreover, the minimum spanning tree conveys all useful information for the subsequent steps. The weights assigned to the edges completely express the ultrametric distance which governs all subsequent steps. The classical morphological segmentation with markers itself gets a new flavour, as it may be used with weight distributions on the MST, which lead to more meaningful segmentations.

## References


[1] S. Beucher. *Segmentation d'Images et Morphologie Mathématique.* PhD thesis, E.N.S. des Mines de Paris, 1990.

[2] M. Grimaud. *La Géodésie Numérique en Morphologie Mathématique. Application a la Détection Automatique de Microcalcifications en Mammographie Numérique.* PhD thesis, E.N.S. des Mines de Paris, 1991.

[3] M. Grimaud. New measure of contrast : dynamics. *Image Algebra and Morphological Processing III, San Diego CA, Proc. SPIE*, 1992.

[4] B. Marcotegui, F. Marqués, R. Morros, M. Pardas, and P. Salembier. Segmentation of video sequences and rate control. *Annales des Télécommunications*, pages 380–389, 1996.

[5] F. Meyer. Un algorithme optimal de ligne de partage des eaux. In *Proceedings* 8$^{\underline{\text{ème}}}$ *Congrès AFCET, Lyon-Villeurbanne*, pages 847–857, 1991.

[6] F. Meyer. Topographic distance and watershed lines. *Signal Processing*, pages 113–125, 1994.

[7] F. Meyer. The levelings. In H. Heijmans and J. Roerdink, editors, *Mathematical Morphology and Its Applications to Image Processing*, pages 199–207. Kluwer, 1998.

[8] F. Meyer and S. Beucher. Morphological segmentation. 1(1):21–46, Sept. 1990.

[9] E. Moore. The shortest path through a maze. In *Proc. Int. Symposium on Theory of Switching*, volume 30, pages 285–292, 1957.

[10] C. Vachier. *Extraction de Caractéristiques, Segmentation d'Image et Morphologie Mathématique.* PhD thesis, E.N.S. des Mines de Paris, 1995.

[11] C. Vachier and F. Meyer. Extinction values: A new measurement of persistence. In I. Pitas, editor, *1995 IEEE Workshop on Nonlinera Signal and Image Processing*, pages 254–257, 1995.

[12] L. Vincent. Grayscale area openings and closings, their efficient implementation. In J. Serra and P. Salembier, editors, *First Workshop on Mathematical Morphology and its Applications to Signal Processing*, pages 22–27, Barcelona, Spain, May 1993.