

Stochastic Watershed Hierarchies

Fernand Meyer

▶ **To cite this version:**

Fernand Meyer. Stochastic Watershed Hierarchies. ICAPR 2015 ; The Eighth International Conference on Advances in Pattern Recognition, Indian Statistical Institute, Jan 2015, Kolkata, India. hal-01111749

HAL Id: hal-01111749
https://minesparis-psl.hal.science/hal-01111749

Submitted on 30 Jan 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Stochastic Watershed Hierarchies

Fernand Meyer
MINES ParisTech
PSL Research University
(CMM) Center for Mathematical Morphology
Fontainebleau, France
Email: fernand.meyer@mines-paristech.fr

***Abstract*—We present a segmentation strategy which first constructs a hierarchy, i.e. a series of nested partitions. A coarser partition is obtained by merging adjacent regions in a finer partition. The strength of a contour is then measured by the level of the hierarchy for which its two adjacent regions merge. Various strategies are presented for constructing hierarchies which highlight specific features of the image. The last part shows how the hierarchies lead to a final segmentation.**

Keywords: Mathematical Morphology, Segmentation, Stochastic Watershed, Waterfall, Hierarhies.



Fig. 1. A gray tone image, its gradient and the watershed segmentation associated to the minima of the gradient.

I. Introduction

If one considers an image as a topographic surface, the watershed transform partitions the domain of the function into catchment basins [4], [11]. Each catchment basin represents the attraction basin of a regional minimum of the function. The watershed partition counts numerous tiles, as many tiles as there are regional minima in the gradient images : the image is oversegmented and has to be regularized.

For segmenting a scene, the watershed is applied to its gradient image. Fig.1 presents a grey tone image, its gradient and the watershed partition, which is oversegmented and has to be regularized. A popular regularization method chooses a reduced number of minima and suppresses all others [8]. This method is called marker driven watershed segmentation, as the minima to be kept are marked. This method is robust as the resulting segmentation is insensitive to a large extent to the shape or position of the markers.

The quality of the result depends on a correct estimation of the contour strength as given by the gradient image. The modulus of the gradient is computed within a small window ; it is sensitive to noise and other perturbations of the images such as blur. For this reason a purely local estimation of the contour strength is often not satisfactory. In this paper we propose more robust but also more selective methods for estimating the contour strength. Furthermore, these methods may be tailored such as to take into account specific features of the image. The fine segmentation is modeled as an edge weighted graph on which all subsequent processings are performed.

II. From the image to a fine partition

A. The Region Adjacency Graph

To be efficient, we work at two resolutions. The lowest level is the pixel level: the initial image is segmented and a fine partition produced. The highest level is the level of regions, of partitions and families of partitions. We suppose that the fine partition produced by an initial segmentation (for instance the watershed transform presented above) contains all contours making sense in the image. We define a dissimilarity measure between adjacent tiles of the fine partition. As an example, in the case of the watershed transform, it may be the lowest altitude or the mean altitude of the watershed line separating two adjacent basins. In the case of color images the dissimilarity may be derived from various color distances.

Partition and dissimilarity between adjacent tiles are then modelled as an edge weighted graph, the region adjacency graph or RAG: each node represents a tile of the partition ; an edge links two nodes if the corresponding regions are neighbors ; the weight of the edge is equal to the dissimilarity between both regions. Working on the graph is much more efficient as working on the image, as there are fare less nodes in the graph as there are pixels in the image.

B. Reminders On Node And/or Edge Weighted Graphs

A *non oriented graph* $G = [N, E]$ contains a set N of vertices or nodes and a set E of edges ; an edge being a pair of vertices. The nodes are designated with small letters: p, q, r...The edge linking the nodes p and q is designated by e_{pq}.The partial graph associated to the edges $E' \subset E$ is $G' = [N, E']$.

Edges and/or nodes may be weighted. Denote by $\mathcal{F}_e$ and $\mathcal{F}_n$ the sets of non negative weight functions on the edges and on the nodes respectively. The function $\eta \in \mathcal{F}_e$ takes its value η_{pq} on the edge e_{pq}. The function $\nu \in \mathcal{F}_n$ takes the weight ν_p on the node p.

A path, π, is a sequence of vertices and edges, interweaved in the following way: π starts with a vertex, say p, followed by an edge e_{ps}, incident to p, followed by the other endpoint s of

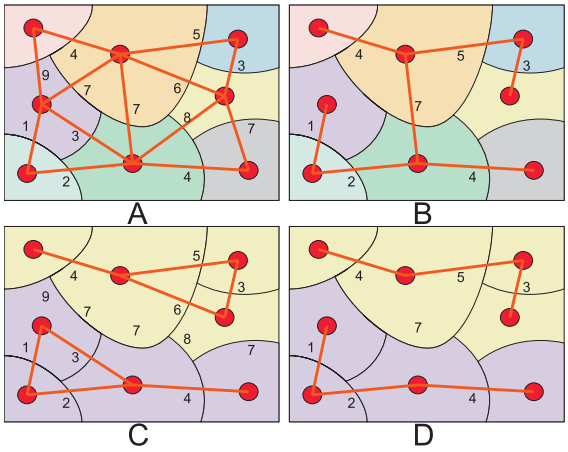


Fig. 2. A: a partition represented by an edge weighted graph
B : the minimum spanning tree of the graph
C : two connected subgraphs, obtained by cutting all edges with a weight above 6. They span a partition of 2 regions.
B : two connected subtrees, obtained by cutting all edges with a weight above 6. They span the same partition of 2 regions

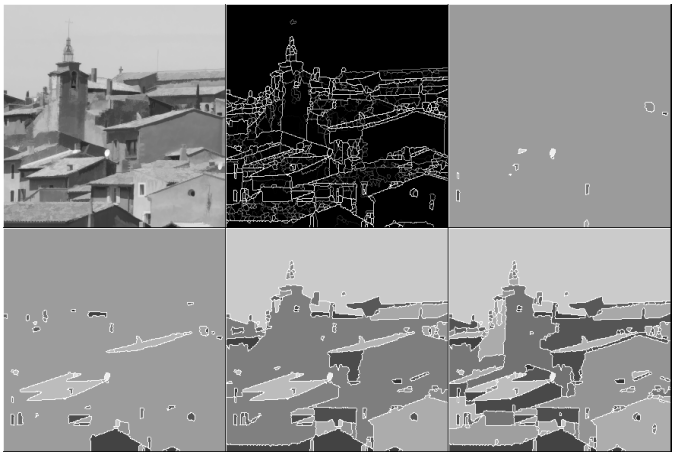


Fig. 3. An image, its watershed partition, followed by 4 levels of the ultrametric gradient hierarchy, going from coarse to fine.

e_{ps}, and so on. *A connected subgraph* is a subgraph where each pair of nodes is connected by a path. A *cycle* is a path whose extremities coincide. *A tree* is a connected graph without cycle. A *spanning tree* is a tree containing all nodes. *A forest* is a collection of trees.

C. The Ultrametric Hierarchy [3]

1) The Ultrametric Distance: We define the altitude of the path π as the weight of the highest edge along the path. Among all paths between two nodes p and q, the paths with the lowest altitude are called critical paths ; their altitude constitutes an ecart δ_{pq} (and not a distance as $\delta_{pq} = 0 \nRightarrow p = q$) between p and q :

- $\delta_{pp} = 0$
- for (p, q, s) : $\delta_{ps} \leq \delta_{pq} \vee \delta_{qs}$. (Proof: Concatenating a critical path between p and q of altitude δ_{pq} and a critical path between q and s of altitude δ_{qs} is a path between p and s whose altitude is higher or equal to the altitude of a critical path between p and s). This inequality is called ultrametric inequality and the ecart δ_{pq} ultrametric distance

2) The Minimum Spanning Tree: In a spanning tree, there exist one and only one path linking two nodes. If this path is always a critical path, the spanning tree is a minimum spanning tree, i.e. the sum of the weights of its edges is minimal. Fig.2B represents a minimum spanning tree T of the graph of fig.2A. Any two nodes are linked by a unique path in this tree, and this path is a critical path. Cutting the highest edge in this path disconnects the MST in two trees, each of them containing one node. Cutting the edges above a threshold λ in the RAG or in the MST disconnects the same nodes as illustrated by fig.2C and D.

3) The Ultrametric Hierarchy: The open ball of center p and radius ρ is $\mathrm{Ball}(p, \rho) = \{q \mid \delta_{pq} < \rho\}$ and the closed ball $\overline{\mathrm{Ball}}(p, \rho) = \{q \mid \delta_{pq} \leq \rho\}$. The following lemmas are easy to prove.

Lemma 1: Each element of a ball $\mathrm{Ball}(p, \rho)$ is centre of this ball

Lemma 2: Two balls $\mathrm{Ball}(p, \rho)$ and $\mathrm{Ball}(q, \rho)$ with the same radius are either disjoint or identical.

Since each node p belongs to one and only ball, namely $\mathrm{Ball}(p, \rho)$, the balls with the same radius form a partition of the nodes.

For increasing radii, the balls are increasing : $\lambda < \mu \Rightarrow \mathrm{Ball}(p, \lambda) \subset \mathrm{Ball}(p, \mu)$. Hence the balls $\mathrm{Ball}(p, \mu)$ form a coarser partition as the balls $\mathrm{Ball}(p, \lambda)$. And $\mathrm{Ball}(p, \mu)$ is the union of all balls $\mathrm{Ball}(q, \lambda)$ for $q \in \mathrm{Ball}(p, \mu)$. A series of partition where the tiles of a coarse partition are obtained by unions of tiles of finer partitions is called a hierarchy.

One would expect that the coarse levels of the hierarchy represent the most salient features of an image and the finer levels constitute minor details and refinements. If the fine partition represented by the RAG is the watershed partition associated to a gradient image, this is unfortunately not the case. Fig.3 presents an image followed by the watershed partition of its gradient image. The contour separating two tiles is weighted by the ultrametric distance between the tiles. The next 4 images show 4 hierarchies obtained for decreasing values of the radius of the balls $\mathrm{Ball}(p, \lambda)$. The coarsest levels of the hierarchy only contain small and contrasted objects of the initial image. The larger structures appear only later. This is due to the fact that the contour surrounding a large object is more likely to have a weaker portion with low values. For this reason this region is more likely to merge with neighboring regions ; it appears as an isolated region for relatively small values of λ in the balls $\mathrm{Ball}(p, \lambda)$.

4) A Minimum Spanning Forest Associated To Markers: Cutting all edges of the MST with a weight higher than λ creates a spanning forest. Among all forests with the same number of trees, this forest is a minimum spanning forest MSF_λ, the sum of its weights being minimal. Two nodes p and q belonging to the same tree of the forest have a distance $\delta_{pq} < \lambda$: they belong to the same ball $\mathrm{Ball}(p, \lambda)$. Hence the

trees of the forest and the balls $\mathrm{Ball}(p, \lambda)$ induce the same partition of the nodes. As shown above, this partition often does not well represent the salient features of an image.

More interesting partitions are obtained with the same number of trees, if we chose the roots of the trees. We select a subfamily (m_i) of nodes, (also called markers) within N and construct a minimum spanning forest where each tree is rooted in a marker. Each minimum spanning forest is obtained by cutting some edges of the MST. Consider two consecutive markers m_1 and m_2 on the MST, such that there exists no other marker along the path along the MST joining both markers. In order to get a forest, one has to cut an edge along this path ; in order to minimize the total weight of the edges, one cuts the highest edge. The same process applied to all pairs of edges produces the desired minimum spanning forest [9].

Consider two consecutive markers m_1 and m_2. Suppose that the highest edge e_{pq} on the path of the MST linking both markers has a weight λ. If p and m_1 (resp. q and m_2) are connected after cutting e_{pq}, then the altitude of the path linking p with $m1$ (resp. q with m_2) is lower then λ. This gives us a criterion for recognizing whether a given edge of the MST belongs or not to the MSF associated to a family of markers: the edge e_{pq} with weight λ does not belong to the MSF if and only if there exists 2 paths with an altitude lower than λ, one linking p with a marker and another linking q with another marker. Or equivalently, if the balls $\mathrm{Ball}(p, \lambda)$ and $\mathrm{Ball}(q, \lambda)$ contain each at least one marker. This criterion will be used all along of this paper for deriving various feature driven hierarchies.

5) A Hierarchy Based On Prioritized Markers: The previous section has explained how to associate a partition to a family (m_i) of the nodes taken as markers. Let F be the minimum spanning forest associated to these markers. Suppose that we add a new marker n. A marker m_k of the family is a neighboring marker of n if there exists a path between n and m_k along the MST on which there is no other marker. Such a path belongs to the tree T_k rooted in m_k. The highest edge along this path has to be cut. Like that the tree T_k is cut in two parts. Hence, by adding new markers, one obtains finer partitions [9].

Consider now a family of markers ranked according to some priority (m_i). We want to construct a hierarchy associated to this family. The coarsest level of the hierarchy is the partition associated to the markers with the highest priority. Every time we add a marker, we obtain a finer partition, as a tile of the coarser partition is cut in several parts. Our goal is to define new weights θ_{pq} for the edges e_{pq} such that cutting all edges with a weight above k produces a minimum spanning forest associated to the k markers with the highest priorities.

The edge e_{pq} with weight $\eta_{pq} = \lambda$ does not belong to the MSF if the balls $\mathrm{Ball}(p, \lambda)$ and $\mathrm{Ball}(q, \lambda)$ contain each at least one marker. If there is no marker at all in one of the balls $\theta_{pq} = 0$. If μ_p and μ_q are the highest priorities of the markers present respectively in $\mathrm{Ball}(p, \lambda)$ and $\mathrm{Ball}(q, \lambda)$, then, by choosing all markers with a priority higher or equal than $\mu_p \wedge \mu_q$, there will be a marker in each of the balls. If we assign to the edge e_{pq} the weight $\mu_p \wedge \mu_q$, we obtain the desired result.

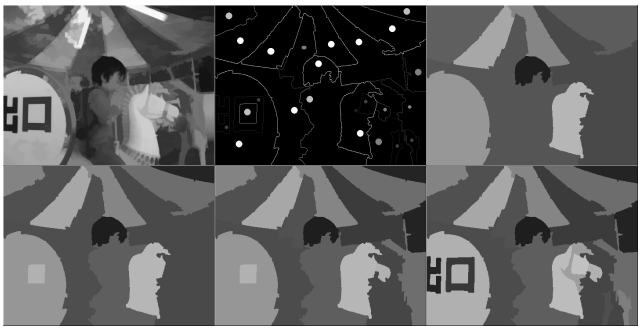


Fig. 4. A number of prioritized markers have been chosen: they appear as disks whose shade of grey is brighter for higher priorities. The associated hierarchy is illustrated through the saliency of the contours in the second image and through 4 levels of the associated hierarchy.

The algorithm visits all edges of the MST in the order of increasing weights. Repeat until all edges are processed:

Let e_{pq} the current edge to process with a weight λ.

If μ_p and μ_q are the highest priorities of the markers present respectively in $\mathrm{Ball}(p, \lambda)$ and $\mathrm{Ball}(q, \lambda)$, we assign to the edge e_{pq} the weight $\mu_p \wedge \mu_q$.

Illustration: In fig.4 a number of prioritized markers have been introduced ; they appear in the second image as disks whose brightness is proportional to the priority. The saliency of the hierarchy is indicated in the same image. The boundary between two regions has a shade of grey proportional to the hierarchy level for which it disappears. The last 4 images represent 4 partitions of the associated hierarchy with decreasing coarseness.

III. The Stochastic Watershed

The last section has shown how ranking the markers generates a hierarchy. We now replace deterministic markers by stochastic markers. The seminal idea, introduced by Angulo [1], [2], is to spread random germs all over the image and to use them as markers for the watershed segmentation. Large regions, separated by low contrast gradient from neighboring regions will be sampled more frequently than smaller regions and will be selected more often. On the other hand, strong contours will often be selected by the watershed construction, as there are many possible positions of markers which will select them. Evaluating the strength of the contours by simulation offers a great versatility : various laws for the implementation of point patterns, various shapes for the markers themselves may be used. The method suffers however from a serious handicap, if the contour strength is evaluated through simulations, as each of them requires the construction of a watershed segmentation. We show below, not only how simulations may be avoided, but also how to imagine scenarios which would be difficult or even impossible to simulate [10].

A. Principle Of The Method

We imagine that we draw random germs on the domain where the image is defined and compute the probability of each piece of contour to appear in the associated segmentation. We have to assign to each edge e_{pq} of the MST with an initial weight η_{pq} a new weight θ_{pq} equal to the probability to appear as a contour. In a first stage we only consider points a markers. Later we will also considers arbitrary, stochastic or deterministic, sets as markers.

As shown above, the edge e_{pq} with weight $\eta_{pq} = \lambda$ does not belong to the MSF if the balls $\mathrm{Ball}(p, \lambda)$ and $\mathrm{Ball}(q, \lambda)$ contain each at least one marker. Thus the probability θ_{pq} is equal to the probability that there is at least one random marker in each of the balls $\mathrm{Ball}(p, \lambda)$ and $\mathrm{Ball}(q, \lambda)$.

For the sake of simplicity, we chose a Poisson distribution of germs over the domain. We fix the number of germs to be equal to ω ; the distribution is then uniform. Consider a set X of area A within a domain D of area S. The probability that there falls no one germ within the domain X is then $\left(1 - \frac{A}{S}\right)^{\omega}$. And the probability that there is at least one germ in X is then $1 - \left(1 - \frac{A}{S}\right)^{\omega}$

1) Absorption Of The Smallest Region:

a) Area Oriented Absorption: Consider the edge e_{pq} with weight $\eta_{pq} = \lambda$ and the balls $\mathrm{Ball}(p, \lambda)$ and $\mathrm{Ball}(q, \lambda)$. Let a_p and a_q be the areas of these balls. We place a deterministic marker in the region with the largest area and a random marker in the smallest. The probability θ_{pq} is then equal to the probability that there exists at least one random marker in the smallest region of area $a_p \wedge a_q$, i.e. $1 - \left(1 - \frac{a_p \wedge a_q}{S}\right)^{\omega}$.

b) "Volume" Oriented Absorption: The previous criterion is based on the area of the balls $\mathrm{Ball}(p, \lambda)$ and $\mathrm{Ball}(q, \lambda)$. For high values of λ this area is likely to be larger than for small values. However, in order to reinforce the influence of the contrast, one may multiply the areas of the balls $\mathrm{Ball}(p, \lambda)$ and $\mathrm{Ball}(q, \lambda)$ by the value λ. This product λa_p may be considered as a kind of volume. Let $\lambda_{\max}$ be the highest weight of the edges of the MST. The probability that no marker falls within the volume λa_p within the total volume $\lambda_{mas} S$ is then $\left(1 - \lambda \frac{a_p \wedge a_q}{S * \lambda_{\max}}\right)^{\omega}$. Remark that whereas the absolute values of λ depend upon the global contrast of the image, the evaluation of the contour strength is nevertheless relatively robust against the change of contrast, as it is based on the ration $\lambda / \lambda_{\max}$.

c) Contrast Oriented Absorption: Consider again the watershed segmentation. If the image on which the watershed is constructed is a gradient image, the significant features are the levels of the pass points between adjacent regions ; the level of the minima, often near to 0, has not much signification. In other situations one has to construct the watershed on images of a different type, for which the levels of the minima is significant. For instance, the micro-aneurisms in a retina appear as dark spots for which the level of the minima is significant. Another example is the segmentation of text on a document (see fig.9) In such situations, the noise often appear also as dark spots with less contrast. With the stochastic watershed less contrasted regions get absorbed by more contrasted regions. We measure the contrast of the ball $\mathrm{Ball}(p, \lambda)$ as the difference between λ and the deepest value ζ_p taken by the image in $\mathrm{Ball}(p, \lambda)$. We put a hard marker in the most contrasted region and compute the probability that there is a marker in the less contrasted region for ω markers uniformly distributed in the range $[0, \zeta_p]$, yielding $\left(1 - \frac{\lambda - \zeta_p \wedge \zeta_q}{\lambda_{\max}}\right)^{\omega}$.

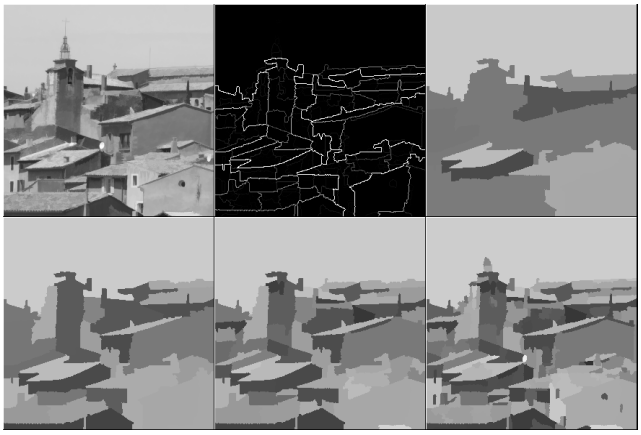


Fig. 5. The surfacic stochastic watershed hierarchy. The saliency of the contours followed by 4 levels of the hierarchy.

2) The Symmetrical Stochastic Watershed:

a) The Area Based Stochastic Watershed: We now consider the distributions of markers in both balls $\mathrm{Ball}(p, \lambda)$ and $\mathrm{Ball}(q, \lambda)$. In short we write $B_p = \mathrm{Ball}(p, \lambda)$ and $B_q = \mathrm{Ball}(q, \lambda)$. The weight θ_{pq} of the edge e_{pq} is then equal to the probability of the event:

$E = \{$there is at least one marker in $B_p\}$ and $\{$there is at least one marker in $B_q\}$

The opposite event is the union of two non exclusive events:

$\mathrm{no}\, E = \{$there is no marker in $B_p\}$ or $\{$there is no marker in $B_q\}$.

Its probability is:

$P(\mathrm{no}\, E)$ = P $\{$there is no marker in $B_p\}$ + P $\{$there is no marker in $B_q\}$ - P $\{$there is no marker in $B_p \cup B_q\}$

And $P(E) = 1 - \left(1 - \frac{a_p}{S}\right)^{\omega} - \left(1 - \frac{a_q}{S}\right)^{\omega} + \left(1 - \frac{a_p + a_q}{S}\right)^{\omega}$

Fig.5 presents the surfacic stochastic hierarchy: the initial image, the new saliency of the contours, followed by 4 levels of the hierarchy.

3) The Volume Based Stochastic Watershed: For stressing more the strength of the gradient separating both regions $\mathrm{Ball}(p, \lambda)$ and $\mathrm{Ball}(q, \lambda)$, we replace the measures of the areas a_p and a_q by the pseudo volumes λa_p and λa_q. The markers being distributed in a total volume $S \times \lambda_{\max}$.

The probability that there exists at least one markers in both "volumes" is then:

$P(E) = 1 - \left(1 - \lambda \frac{a_p}{S \times \lambda_{\max}}\right)^{\omega} - \left(1 - \lambda \frac{a_q}{S \times \lambda_{\max}}\right)^{\omega} + \left(1 - \lambda \frac{a_p + a_q}{S \times \lambda_{\max}}\right)^{\omega}$

4) The Symmetrical Stochastic Watershed Within Transformed Domains: Until now we considered the domains $\mathrm{Ball}(p, \lambda)$ and $\mathrm{Ball}(q, \lambda)$ only through their area or the deepest

value taken by the image within the balls. In order to now take into account also their shape, we apply an anti-extensive morphological operator ψ on the balls: $\psi(X) \subset X$. The area of $\psi(X)$ is thus smaller than the area of X. The most common operators are the erosion and the opening. This opens a large choice of possibilities : erosion or opening, type of structuring elements (often disks or segments in various directions), size of the structuring element, etc.

We define $\beta_p = \text{area}[\psi\,\text{Ball}(p,\lambda)]$. The probability θ_{pq} to be assigned to the edge e_{pq} is then $1 - \left(1 - \frac{\beta_p}{S}\right)^{\omega} - \left(1 - \frac{\beta_q}{S}\right)^{\omega} + \left(1 - \frac{\beta_p+\beta_q}{S}\right)^{\omega}$

It is noteworthy that this assignment of probabilities cannot be obtained by the simulation method used by Jesus Angulo, consisting in introducing real random germs in the image and constructing the watershed partition for each new simulation.

5) The Symmetrical Stochastic Watershed With Non Punctual Markers: The computation which follows corresponds to the experiment where one uses random markers, which are not reduced to points. We suppose that Z_x is a marker implanted at a random position x. For the sake of simplicity we suppose that Z is the same marker everywhere, and its implementation is random. It is possible to imagine and compute the probabilities using random markers (for instance disks with random radii, segments with random or regionalized length and orientation etc.). Recall that the structuring element Z_x hits a set X if its center x belongs to the dilation of X by $Z : x \in X \oplus Z$.

Taking the same notations as above : the edge e_{pq} will be cut for a random distribution of markers, if the 3 following events are verified:

- $A1 = \{\exists$ random marker Z hitting $B_p\} = \{\exists$ random point marker belonging to $B_p \oplus Z\}$
- $A2 = \{\exists$ random marker Z hitting $B_q\} = \{\exists$ random point marker belonging to $B_q \oplus Z\}$
- $A3 = \{\nexists$ random marker Z hitting B_p and $B_q\} = \{\nexists$ random point marker belonging to $(B_p \oplus Z) \cap (B_q \oplus Z)$

The balls B_p and B_q before and after dilation by an horizontal segment, and the intersection $(B_p \oplus Z) \cap (B_q \oplus Z)$ of both dilated sets are illustrated in fig.6

We have to compute $P(A1 \text{ and } A2 \text{ and } A3) = P(A1 \text{ and } A2 \mid A3) \times P(A3)$.

If S_{pq} is the area of $(B_p \oplus Z) \cap (B_q \oplus Z)$, then $P(A3) = \left(1 - \frac{S_{pq}}{S}\right)^{\omega}$.

And $P(A1 \text{ and } A2 \mid A3) = 1 - P(\text{not } A1 \text{ or not } A2 \mid A3) = 1 - P(\text{not } A1 \mid A3) - P(\text{not } A2 \mid A3) + P(\text{not } A1 \text{ and not } A2 \mid A3)$

The conditional probability $P(. \mid A3)$ means that all punctual germs have been distributed outside $(B_p \oplus Z) \cap (B_q \oplus Z)$, that is in an area $S - S_{pq}$. And the event $(\text{not } A1 \mid A3)$ means that there is no germ falling in $B_p \oplus Z$, knowing that there is also no germ falling in $(B_p \oplus Z) \cap (B_q \oplus Z)$, i.e. there is no germ falling in $(B_p \oplus Z) \setminus (B_q \oplus Z)$, domain with an area $S_{p \setminus q}$. Thus the probability is equal to $\left(1 - \frac{S_{p \setminus q}}{S - S_{pq}}\right)^{\omega}$. Exchanging the roles of p and q, we get $P(\text{not } A2 \mid A3) = \left(1 - \frac{S_{q \setminus p}}{S - S_{pq}}\right)^{\omega}$.

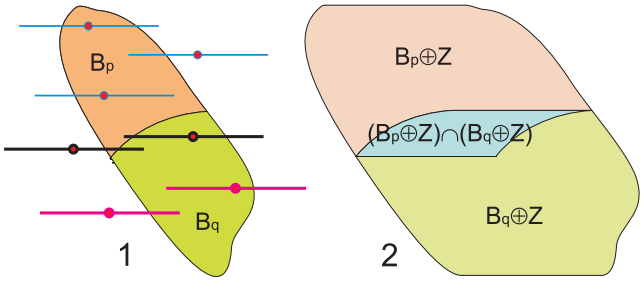


Fig. 6. 1: The two balls B_p and B_q and a number of structuring elements hitting the balls. The black ones hit both balls.
2: The dilated balls $B_q \oplus Z$ and $B_p \oplus Z$ and their intersection $(B_q \oplus Z) \cap (B_p \oplus Z)$ in cyan colour.

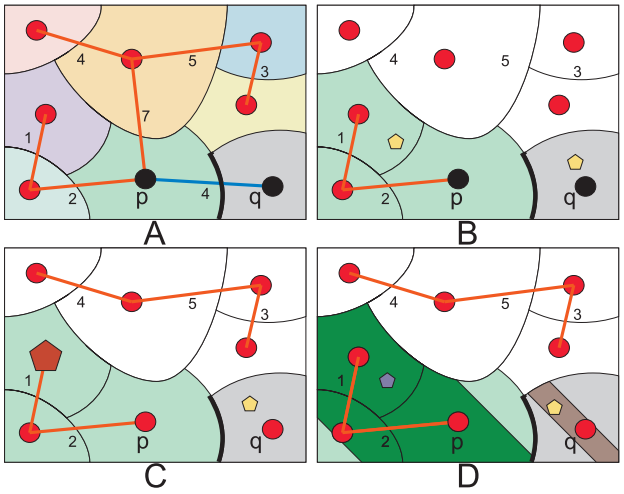


Fig. 7. A: Partition and its minimum spanning tree
B: Area stochastic watershed: a random marker in both colored regions B_p and B_q
C: Area oriented absorption stochastic watershed: a fixed marker in the largest region B_p and a random marker in the smaller region B_q
D: Area stochastic watershed with transformed domains: a random marker in the regions obtained by a linear opening of B_p and B_q.

The event $\{\text{not } A1 \text{ and not } A2 \mid A3\}$ means that there is no punctual germ in $(B_p \oplus Z) \setminus (B_q \oplus Z)$ nor in $(B_q \oplus Z) \setminus (B_p \oplus Z)$. If $S_{p \triangle q}$ is the area of $(B_p \oplus Z) \setminus (B_q \oplus Z) \cup (B_q \oplus Z) \setminus (B_p \oplus Z)$, we obtain the probability $P(\text{not } A1 \text{ and not } A2 \mid A3) = \left(1 - \frac{S_{p \triangle q}}{S - S_{pq}}\right)^{\omega}$.

Putting everything together, we get the new weight $\theta_{pq} = \left\{1 - \left(1 - \frac{S_{p \setminus q}}{S - S_{pq}}\right)^{\omega} - \left(1 - \frac{S_{q \setminus p}}{S - S_{pq}}\right)^{\omega} + \left(1 - \frac{S_{p \triangle q}}{S - S_{pq}}\right)^{\omega}\right\} \times \left(1 - \frac{S_{pq}}{S}\right)^{\omega}$

6) Illustration: Fig.7A presents the partition and a minimum spanning tree derived from the dissimilarities between adjacent regions. We want to evaluate the strength of the blue edge e_{pq}, having a weight equal to 4. This edge will get a new weight θ_{pq} according various scenarii:
Fig.7B: Area stochastic watershed : All edges with a weight above or equal to 4 are cut, leaving two trees representing the regions $B_p = \text{Ball}(p,4)$ and $B_q = \text{Ball}(q,4)$. Two yellow

polygons symbolize two random markers in these balls. θ_{pq} is the probability that at least one marker falls in each of the regions B_p and B_q.
Fig.7C: Area oriented absorption: A non random marker (large red polygon) is placed in the largest ball B_p. θ_{pq} is the probability that a random marker (yellow polygon) falls in the smallest region B_q
Fig.7D: Area stochastic watershed with transformed domains: Both balls B_p and B_q are submitted to an opening γ by a segment in the direction $2\pi/3$. θ_{pq} is the probability that at least one marker falls in each of the opened regions γB_p and γB_q.

B. More Hierarchies

1) The Waterfall Hierarchy: Starting with the MST of the RAG, we keep for each node one and only one of its lowest neighboring edges. We create like that a spanning forest. Assigning the same label to all nodes of each tree yields the level 2 of a hierarchy. The next level is obtained by retaining for each tree one and only one of the edges linking this tree with a neighboring tree. A number of tree has merged, creating a forest with less trees, inducing the partition of level 3 of the hierarchy. The same process may go on, creating at each stage a new level of the hierarchy. This hierarchy has first been described in the context of flooding a topographic surface, and called waterfall hierarchy [5]. If the nodes represent the catchment basins of a topographic surface, and a basin is flooded, then it overflows, creating a waterfall, into a neighboring basin and this overflow occurs along its lowest edge.

It is possible to produce the waterfall hierarchy in one pass through the edges of the MST, with initial weights τ_{pa}. We will assign to each edge e_{pq} of the MST a new weight θ_{pq} expressing the level of the waterfall hierarchy. The algorithm visits all edges of the MST in the order of increasing weights τ. Repeat until all edges are processed:

Let e_{pq} the current edge to process with a weight λ. The initial weights τ of the edges of $B_p = \mathrm{Ball}(p, \lambda)$ and $B_q = \mathrm{Ball}(q, \lambda)$ are lower than λ. Hence all the new weight θ of these edges has already been computed. The highest weight taken by the function θ in the ball B_p is called diameter of the ball and we write $\operatorname{diam} B_p$. The waterfall level θ_{pq} is then equal to: $\theta_{pq} = 1 + \min(\operatorname{diam} B_p, \operatorname{diam} B_q)$. The diameter of the ball $B_p \cup \{u\} \cup B_q$ is then $\max(\operatorname{diam} B_p, \theta_{pq}, \operatorname{diam} B_q)$.

Fig.8 presents the waterfall hierarchy: the initial image, the waterfall saliency of the contours, followed by 4 levels of the waterfall hierarchy.

2) Cascading And Combining Hierarchies: All hierarchies described so far are fully characterized by their ultrametric distance (UD) ; these distances have as support the same MST spanning the nodes/regions of the same fine hierarchy and differ by the weights of the edges. Each operator described so far takes as input a set of weights of the MST and produces a new set of weights on the same MST ; in some cases additional measurements taken in the image are needed. This new MST may then be submitted to the same process and a second hierarchy produced, taking into account different features of the image.

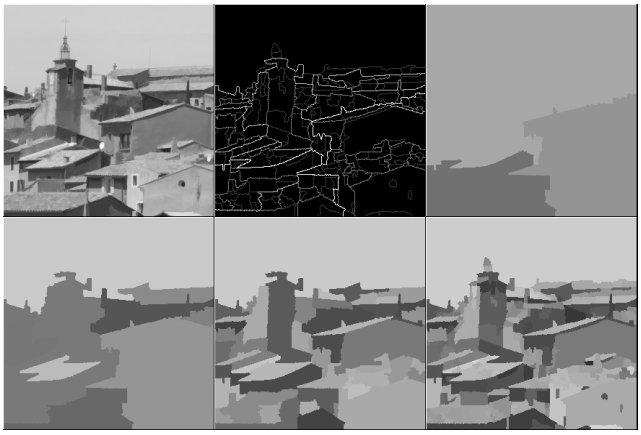


Fig. 8. The waterfall hierarchy. The saliency of the contours followed by 4 levels of the waterfall hierarchy.

3) The Lattice Of Hierarchies: Two hierarchies $\mathcal{A}$ and $\mathcal{B}$ may be compares through their UD $\chi_{\mathcal{A}}$ and $\chi_{\mathcal{B}}$: $\mathcal{B} \leq \mathcal{A} \Leftrightarrow \forall p, q \in E \quad \chi_{\mathcal{A}}(p, q) \leq \chi_{\mathcal{B}}(p, q)$

As $\forall p \in E : \mathrm{Ball}_{\mathcal{B}}(p, \rho) \subset \mathrm{Ball}_{\mathcal{A}}(p, \rho)$, the hierarchy $\mathcal{A}$ is coarser than the hierarchy $\mathcal{B}$.

Consider now a family of hierarchies $(\mathcal{A}_i)_{i \in I}$, with the associated UD χ_i. The infimum $\wedge \mathcal{A}_i$ is the largest hierarchy which is smaller than each $\mathcal{A}_i$ and its UD is $\chi_{\wedge \mathcal{A}_i} = \bigvee_i \chi_i$. The infimum of hierarchies is particularly useful for dealing with color images, as the hierarchies produced for each color component may be combined. The supremum $\vee \mathcal{A}_i$ is characterized by its UD, the largest UD below $\bigwedge_i \chi_i$.The supremum of hierarchies retains the contours which are present in various hierarchies, and thus emphasizes the strength of these contours.

IV. Conclusion: Tailor The Hierarchy Which Is Best For Your Problem

A hierarchy aims at proposing a reduced but sufficient set of contours in an image, ranked by their pertinence and importance. There is nothing like an optimal hierarchy, adapted to all types of images or objects to segment.

Each method presented above constructs a particular hierarchy, although they are all derived from the same fine partition. Hierarchy $\mathcal{A}$ will highlight some contours of this fine partition and neglect others, whereas hierarchy $\mathcal{B}$ makes another choice. For this reason will a particular hierarchy inform us about the image content. For instance, if a contour appears strong in a stochastic watershed hierarchy based on openings with large horizontal structuring elements but is weak for vertical structuring elements, it informs us about the local orientations in the image. Combining the weights of the same piece of contour obtained for various hierarchies constitutes a powerful signature which may serve for object recognition or image matching.

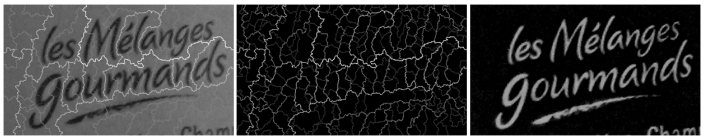


Fig. 9. Waterfall hierarchy used for the segmentation of text.

If we have to segment an image we have to design a hierarchy which discards the structures of no interest and highlights those which are interesting for us. In order to design a useful hierarchy which will facilitate the further processing we have to analyze carefully the image and determine which features best characterize the image and the objects to detect : size, contrast, orientation, color, texture etc. ? We then have to design one or several hierarchies which highlight these features.

Let us now give some cues on how to use hierarchies for extracting among all weighted contours the contours of the objects we want to detect.

A. Marker Based Segmentation

Marker based segmentation has been rephrased above as the construction of a hierarchy, in which the coarsest level of the hierarchy corresponds to the desired segmentation; each region containing one and only one marker. We have shown also how to use a family of prioritized markers. We have seen that the hierarchies may be cascaded. If the first hierarchy highlights correctly the contours of interest, it will be possible to extract the regions of interest with a reduced number of markers. The segmentation will be more robust and less sensitive to the shape or size of the markers.

B. Robust Ans Pararmeter Free Top-Hats

The top hat, the residue of an opening or a closing, is a useful operator for detecting text on a non uniform background. The waterfall hierarchy is a hierarchy which does not depend on any parameter. Consider the image 9: the text is dark on a brighter background. The watershed segmentation of this image (and not on its gradient) produces a fine segmentation. The waterfall hierarchy analyses how the structures of the image are nested. The ranking of the contours is an enumeration of these nested levels, is completely independent of the contrast of the image and does not depend upon any parameter. We take the contours of the last but one level in the waterfall hierarchy and derive from it a ceiling function which is equal to the initial image along these contours and is white everywhere else. The highest flooding of the image under this function fills completely the text. The residue produces a bright text on a uniformly dark background.

C. Interactive Segmentation With Hierarchies

Interactive segmentation strategies are particularly efficient when applied on hierarchies. As the hierarchy is constructed beforehand, the computing time after each interaction is greatly reduced. The library of routines for interactive segmentation, as well for segmenting multimedia images [14] as for segmenting medical images [13].

Local resegmentations or mergings : a first partition is chosen in the hierarchy and then adapted locally by resegmenting regions which are two large or inversely by merging adjacent regions. Both operations are simply obtained by going up or down in the hierarchy.

Magic wand To extract a region with uniform color, most drawing/painting software packages have a function called ”magic wand”. For each position of the mouse, the color is determined and the connected region composed of all pixels with more or less the same color, depending on some tolerance threshold, is selected. This procedure is often helpful, but fails in some situations, when there is a progressive change of color shade, as is the case with the yellow apple in fig.10. The darker part of the apple is not selected and an irregular contour produced. On the contrary, using a hierarchy has the advantage of providing well defined contours. The hierarchy based magic wand selects the largest region in the hierarchy such that its mean color remains within some predefined limits.



Fig. 10. On the left, initial image ; center: all pixels which are within a colour tolerance of an initial pixel. On the right, result of the magic wand.

Lasso Another classical interactive tool is the lasso : the user draws an approximate contour around the real contour as shown in fig.11a. The classical solution consists in applying the magic wand defined above to each pixel belonging to the approximate contour. For each such position one gets a piece of the background. The union of all such pieces constitutes the background. As shown on fig.11b, the result is not very satisfactory. Using a hierarchy, one may select the union of all regions of the hierarchy contained in the contour yielding a much better result as shown on fig.11c.



Fig. 11. Construction of all tiles of a hierarchy entirely included in an approximate outside contour

Intelligent Brush: An intelligent brush segments an image

by ”painting” it: it first selects a zone of interest by painting. Contrary to conventional brushes, the brush adapts its shape to the contours of the image. The shape of the brush is given by the region of the hierarchy containing the cursor. Moving from one place to another changes the shape of the brush, when one goes from one tile of a partition to its neighboring tile. Going up and down the hierarchy modifies the shape of the brush. In fig.12, on the left, one shows the trajectory of the brush ; in the centre, the result of a fixed size brush, and on the right the result of a self-adapting brush following the hierarchy. This method has been used with success in a package for interactive segmentation of organs in 3D medical images [13].



Fig. 12. Comparison of the drawing with a fixed size brush and a self adaptive brush.

D. Energy Minimization In A Hierarchy

An additional way to construct hierarchies is through energy minimization, which becomes a tractable problem if it is applied on a hierarchy. Given a hierarchy $\mathcal{A}$, one wants to extract a partition π, whose regions verify an optimality criterion. The regions all belong to the hierarchy but not necessarily to the same hierarchical level. Philippe Salembier et al proposed to construct optimal partitions in the context of image coding ; the aim is to produce a partition where each region is described by a simplified model under the constraint that the encoding cost is not too high [12].

Laurent Guigues [6] analyzed the types of energies which may be minimized within hierarchies. His work has been continued and extended by Ravi Kiran and Jean Serra [7]. The energies contain two terms, a data fidelity term and a regularization term ; the value of the first increases and the second decreases, by climbing in the hierarchy towards coarser levels. Both terms are linked by a scale parameter. As an example, consider the Mumford-Shah model where $D(R_i)$ represents the total variance of the image in the region R_i of the partition and the second term measures the length of the contours present in the partition ; we get like that a kind of energy: $E(\pi, \lambda) = \sum_{R_i \in \pi} D(R_i) + \lambda C(\pi)$ where λ is a scale parameter. For each scale parameter an optimal partition is easily extracted from the hierarchy through dynamic programing. For increasing values of λ one obtains a series of nested partitions, i.e. a new hierarchy. There are then various strategies for approximating or finding the global minimum. The following example in fig.13 is due to the courtesy of Jean Stawiaski, from Philips Medical Systems. It presents the various steps for segmenting a tumor: initial image, gradient, fine segmentation, saliency of the initial contours, saliency of the surfacic stochastic watershed, extraction of the contours minimizing the Mumford Shah functional.

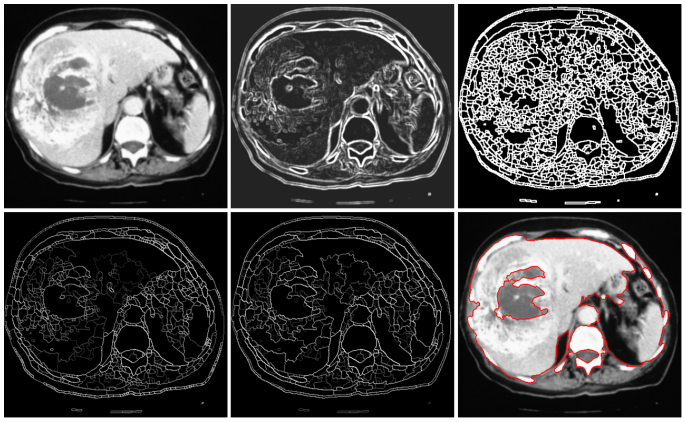


Fig. 13. Segmentation of a tumor in a liver: initial image, gradient, fine segmentation, saliency of the gradient, saliency of the stochastic watershed, final segmentation

REFERENCES

[1] J. Angulo and D. Jeulin. Stochastic watershed segmentation. *ISMM07 : Mathematical Morphology and its applications to Signal and Image Processing*, pages 265–276, 2007.

[2] Angulo, J., Velasco-Forero, S., Chanussot, J.: Multiscale stochastic watershed for unsupervised hyperspectral image segmentation. In: IEEE IGARSS'2009 (IEEE International Geoscience and Remote Sensing Symposium). (2009) 93–96

[3] J. P. Benzécri. *L'analyse des données 1. La taxinomie*, chapter 3, pages 119–153. Dunod, 1973.

[4] S. Beucher and C. Lantuéjoul. Use of watersheds in contour detection. In *Proc. Int. Workshop Image Processing, Real-Time Edge and Motion Detection/Estimation*, 1979.

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

[6] Laurent Guigues, Jean Pierre Cocquerez, and Hervé Le Men. Scale-sets image analysis. *International Journal of Computer Vision*, 68(3):289–317, 2006.

[7] B Ravi Kiran and Jean Serra. Global–local optimizations by hierarchical cuts and climbing energies. *Pattern Recognition*, 47(1):12–24, 2014.

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

[9] F. Meyer. Minimal spanning forests for morphological segmentation. *ISMM94 : Mathematical Morphology and its applications to Signal Processing*, pages 77–84, 1994.

[10] Fernand Meyer and Jean Stawiaski. A stochastic evaluation of the contour strength. In *DAGM-Symposium'10*, pages 513–522, 2010.

[11] Roerdink and Meijster. The watershed transform: Definitions, algorithms and parallelization strategies. *FUNDINF: Fundamenta Informatica*, 41:187–228, 2000.

[12] P. Salembier. Morphological multiscale segmentation for image coding. *EURASIP Signal Processing, 359-386*, 38(3):359–386, august 1994.

[13] Jean Stawiaski, Etienne Decenciere, and François Bidault. Interactive liver tumor segmentation using graph-cuts and watershed. In *Workshop on 3D Segmentation in the Clinic: A Grand Challenge II. Liver Tumor Segmentation Challenge. MICCAI, New York, USA*, 2008.

[14] M. F. Zanoguera, B. Marcotegui, and F. Meyer. An interactive colour image segmentation system. In *Wiamis'99 : Workshop on Image Analysis for Multimedia Interactive Services*, pages 137–141. Heinrich-Hertz Institut Berlin, 1999.