Morphological probabilistic hierarchies for texture segmentation

Dominique Jeulin

Research Article **Open Access**

Dominique Jeulin*

Morphological probabilistic hierarchies for texture segmentation

Abstract: A general methodology is introduced for texture segmentation in binary, scalar, or multispectral images. Textural information is obtained from morphological operations on images. Starting from a fine partition of the image in regions, hierarchical segmentations are designed in a probabilistic framework by means of probabilistic distances conveying the textural or morphological information, and of random markers accounting for the morphological content of the regions and of their spatial arrangement. The probabilistic hierarchies are built from binary or multiple fusion of regions.

1 Introduction

In many cases images contain regions with different textures, rather than objects on a background or regions with homogeneous grey level easily segmented by thresholding.

Automatic texture extraction is required in different areas such as for instance industrial control [7] or remote sensing [25, 26]. In these two last cases, in every pixel of images multivariate information is available, like results of morphological transformations applied to grey level or to binary images [6, 7], or like the wavelength response of a sensor in multispectral images [25, 26]. In this context a typical approach of segmentation makes use of pixel classification by means of multivariate image analysis [6, 7], sometimes combined with a watershed segmentation based on some multivariate gradient [25, 26].

In what follows, we introduce a hierarchical probabilistic segmentation of textures based on multivariate morphological information available on every pixel. After a short review of earlier works, a reminder on morphological texture descriptors and texture classification is given. Then a probabilistic approach of hierarchical segmentation of textures, based on a probabilistic distance and on the introduction of various random markers, is developed. This work is an extension of preliminary results given in [17].

2 Earlier works on probabilistic segmentation of textures

Previous publications on probabilistic segmentation of textures are rather sparse. We present here for comparison a short review of some papers related to this area.

In the early work [30], objects (rather than textures) are labelled by relaxation operations to remove some ambiguity in their classification. A heuristic process involves a probabilistic model, called stochastic labeling,

*Corresponding Author: Dominique Jeulin: MINES ParisTech, PSL Research University, CMM-Centre de Morphologie Mathématique, 35, rue Saint Honoré, F-77305 Fontainebleau, France, Tel.: +33-1-64-69-47-95, Fax: +33-1-64-69-47-07, E-mail: dominique.jeulin@mines-paristech.fr

using compatibility constraints between the labels of all objects. They are implemented as empirical coefficients of a linear or a non-linear operator acting on the probabilities of the labels attributed to pairs of object. The choice of the operators is justified by their behaviour during iterations, namely a convergence to a fixed point, starting from an initial guess, but without any probabilistic basis. In a later work [13], the probabilistic interpretation is rejected and the approach is reformulated in the context of local optimization, leading to a relaxation labeling by a projection operator.

In [28] a probabilistic classification (or attribution of labels) of nodes of a graph or of pixels is performed in a Bayesian framework. For this, the probability for a node to get a given label is considered as a random variable, which is estimated from information on its neighbours in an iterative calculation involving a conditional independence of the probabilities, knowing the labels. An updating operator of the labels by means of neighboring probabilities is introduced. Its convergence is ensured by the existence of fixed points.

In [20] segmentation algorithms rely on stochastic and deterministic relaxation principles. Using the Gauss Markov random field model the segmentation is given by the approximate solution of an optimization problem, namely the MAP (maximum *a posteriori*) or the expected classification error rate per pixel. The possibility of hierarchical segmentation from the change of the energy involved in the Gibbs formulation is briefly mentioned.

A probabilistic bottom-up aggregation approach [1] merges adjacent regions by a graph coarsening procedure. Various criteria (intensity and texture by means of outputs of edge filters) are accounted for, and combined by a mixture of probabilities involving likelihoods estimated from information in surrounding regions. Prior probabilities promote the fusion of adjacent regions with long common boundaries.

In [19] a supervised segmentation of textures is proposed. Starting from filter responses, a similarity measure between textures is given from the data of the Kullback-Leibler divergence between histograms obtained on local windows around a training set of pixels. This similarity measure is then used for every pixel and introduced in a Bayesian Markov random field (MRF) for the segmentation (in fact classification) of each pixel. In [18] the same method is applied to the fusion of existing regions (instead of local windows), and using an active contour based segmentation.

In contrast with some of these previous developments, our approach does not make use of any probabilistic or statistical model of random field for the image. The randomness is introduced by the process of sampling random points in the image, or by the use of random markers. Furthermore, we operate on regions of a partition resulting from an oversegmentation process and will build hierarchies based on a probabilistic content.

3 Morphological texture descriptors

We will consider images as domains $\mathcal{D}$ in the n dimensional space $\mathbb{R}^n$. Every pixel x is described by a set of morphological parameters or transformations building a vector with dimension p in the parameter space $\mathbb{R}^p$. For instance in the case of 2D multispectral images, we have $n = 2$ and p is the number of channels of each spectrum. Many types of transformations can be used. From experience, some standard families of morphological transformations Ψ [21, 31], performed on an initial image, are efficient as texture descriptors [6, 7]: dilations $\delta(\rho)$, erosions $\varepsilon(\rho)$, openings $\gamma(r)$ or closings $\varphi(\rho)$ by convex structuring elements with size ρ. These operations are as well defined for binary images as for scalar grey level images. Efficient texture descriptors, from the point of view of pixel classification, are increments of transformations with respect to the size ρ. Thus for a binary image A, vectors of description are obtained for each type of transformation, with components $I_\alpha(x)$ where I_α is the indicator function of the set $\Psi(A, \rho_\alpha) \triangle \Psi(A, \rho_{\alpha-1})$, $\triangle$ being the set difference, and α ranging from 1 to s, with $\rho_0 = 0$. For a grey level image $Y(x)$, the components are given by the increments $Z_a(x) = \left| \Psi(Y(x), \rho_\alpha) - \Psi(Y(x), \rho_{\alpha-1}) \right|$. When Ψ is an opening or a closing, the increments $Z_a(x)$ provide a granulometric spectrum, as used in various domains: binary textures [6, 32], rough surfaces [3], satellite imagery [29], to mention a few. In some specific situations, the components $I_\alpha(x)$ or $Z_a(x)$ are averaged in a local window $K(x)$ around x, to provide local granulometries [6, 7, 11] or the output of linear filters, like curvelet

transform [6, 7]. These descriptors are easily extended in a marginal way to the components of multispectral images.

4 Texture classification

The morphological descriptors generate a vector field on the domain $\mathcal{D}$, from which a classification of pixels in the various textures present in the image can be looked for. For this, a partition in classes C_β must be built in the high dimensional parameter space. A convenient methodology is based on multivariate factor analysis to reduce the dimension of the data and to remove noise: in [25, 26] use is made of Factor Correspondence analysis FCA, well suited to positive data, like multi spectral images or like probability distributions as encountered in granulometric spectra; for heterogeneous data, Principal Component Analysis PCA can be a good method to produce the dimensional reduction [7]. Each of these analysis makes use of some specific distance in the parameter space, which we will denote $\|Z(x_1) - Z(x_2)\|$ for the descriptors of the two pixels x_1 and x_2. Various distances can be chosen (for instance the chi-squared between distributions in the case of FCA). We will not discuss this choice here, which is highly application dependent.

A classification of pixels is then made in the parameter space or in its reduced version, after keeping the most prominent factors. This classification can be unsupervised, using random germs in the K-means algorithm or a hierarchical classification, as described in [4]. When the textures are documented by a set of representative pixels, supervised statistical learning methods can be implemented for later classification (see an extensive presentation in [12]). In [6] a Linear Discriminant Analysis LDA is used. It follows a PCA in [7].

5 Probabilistic texture segmentation

In this section we assume that in every pixel x in the image embedded in $\mathbb{R}^n$, multivariate information (like multispectral data, or transformed images as described in section 3) is stored in a vector $Z(x)$ with components $Z_\alpha(x)$. For any pair of pixels x_1 and x_2, a multivariate distance $\|Z(x_1) - Z(x_2)\|$ is defined in the parameter space $\mathbb{R}^p$.

5.1 Watershed texture segmentation

Considering points y in the neighborhood $B(x)$ of point x, a multivariate gradient can be defined as (Noyel et al. 2007; Noyel et al. 2008):

$$\text{grad}\left(Z(x)\right) = \vee_{y \in B(x)} \left\|Z(x) - Z(y)\right\| - \wedge_{y \in B(x)} \left\|Z(x) - Z(y)\right\|$$

The gradient image can be used as the starting point of the segmentation of the domain $\mathcal{D}$ into homogeneous regions A_i. In fact it is expected that a texture sensitive gradient will provide weak values in homogeneous regions, and high values on the boundary A_{ij} between two regions A_i and A_j. A separation of the domain $\mathcal{D}$ in homogeneous connected regions A_i is obtained by the construction of the watershed of the gradient image from markers generated by the minima of the gradient, as initially defined for scalar images [5, 23] and later extensively used for multispectral images [25, 26].

The main drawback of the watershed segmentation is its sensitivity to noise, resulting in systematic oversegmentation of the image. This is alleviated by means of a careful choice of markers, driven by some local content, like for instant chosen from a multivariate classification [27].

Another approach, the stochastic watershed [2], makes use of random markers replacing the usual markers, enabling us to estimate a local probability of boundaries at each point $x \in A_{ij}$. The main idea is to evaluate the strength of contours by their probability, estimated from Monte Carlo simulations in a first step, as developed in the scalar [2] and in the multispectral cases [25–27]. Simulations can be replaced by a direct

calculation of the probability of contour for each boundary A_{ij} between adjacent regions A_i and A_j [15, 24]. This was successfully applied for 3D multiscale segmentation of granular media using point markers [9] or oriented Poisson lines markers [10].

5.2 Probabilistic hierarchical segmentation

In what follows, we design a new probabilistic segmentation obtained by a hierarchical merging of regions from a fine partition of a domain $\mathcal{D}$ in regions A_i. This initial partition can be obtained in a first step from the watershed of a gradient image, or from some classification of pixels. Given two regions A_i and A_j, not necessarily connected or even adjacent, we will estimate for various criteria the probability p_{ij}:

$$p_{ij} = P\{A_i \text{ and } A_j \text{ contain different textures}\} \tag{1}$$

The probability p_{ij} will play the same role as a gradient (or a distance) between regions A_i and A_j. In a hierarchical approach, a progressive aggregation of regions is performed, starting from lower values of p_{ij} and updating the probability after fusion of regions containing similar textures. This approach was proposed for the case of random markers [15] and implemented in an iterative segmentation based on the stochastic watershed [9]. Using a probabilistic framework makes easier the combination of different criteria for the segmentation, as illustrated later.

In the context of texture classification, the probability $1 - p_{ij} = P\{A_i \text{ and } A_j \text{ contain the same texture}\}$ is also a similarity index between regions A_i and A_j.

5.2.1 Probabilistic distance

Consider two points x_1 and x_2 in a domain $\mathcal{D}$, and the multivariate distance $\|Z(x_1) - Z(x_2)\|$. The choice of a specific multivariate distance (not necessarily Euclidean), with appropriate scaling of variables, is quite standard in multivariate data analysis, and is not discussed in the present paper. When the two points are located randomly in $\mathcal{D}$, $\|Z(x_1) - Z(x_2)\|$ becomes a random variable, characterized by its cumulative distribution function $P\left\{\|Z(x_1) - Z(x_2)\| \geq d\right\} = T(x_1, x_2, d)$. We have the following property:

Proposition 1. *For any $d > 0$, the distribution function $T(x_1, x_2, d)$ is a distance in $\mathcal{D}$.*

Proof. We have $T(x_1, x_1, d) = 0$ and $T(x_1, x_2, d) = T(x_2, x_1, d)$. T satisfies the triangle inequality: for any triple (x_1, x_2, x_3),

$$\left\|Z(x_1) - Z(x_2)\right\| \leq \left\|Z(x_1) - Z(x_3)\right\| + \left\|Z(x_2) - Z(x_3)\right\|.$$

Therefore

$$\left\|Z(x_1) - Z(x_2)\right\| \geq d \Longrightarrow \left\|Z(x_1) - Z(x_3)\right\| + \left\|Z(x_2) - Z(x_3)\right\| \geq d$$

and

$$T(x_1, x_2, d) \leq P\left\{\left\|Z(x_1) - Z(x_3)\right\| + \left\|Z(x_2) - Z(x_3)\right\| \geq d\right\}$$
$$\leq T(x_1, x_3, d) + T(x_2, x_3, d).$$

$\square$

Definition 1. *Consider two regions A_i and A_j in $\mathcal{D}$, and two independent random points $x_i \in A_i$, $x_j \in A_j$. For any $d > 0$, the probability*

$$P(A_i, A_j, d) = P\left\{\left\|Z(x_i) - Z(x_j)\right\| \geq d\right\} \tag{2}$$

defines a probabilistic distance between A_i and A_j.

By construction, $P(A_i, A_j, d)$ is a pseudo-distance, since we have not for every d, $P(A_i, A_i, d) = 0$. For any triple $(A_i \subset \mathcal{D}, A_j \subset \mathcal{D}, A_k \subset \mathcal{D})$ and $(x_i \in A_i, x_j \in A_j, x_k \in A_k)$ we have $P(A_i, A_j, d) \leq P(A_i, A_k, d) + P(A_j, A_k, d)$ as a result of the triangle inequality satisfied by $T(x_i, x_j, d)$.

The probabilistic distance can also be used between classes C_β obtained for a partition in the parameter space as a result of a classification. In that case we will define for two classes C_α and C_β the probability

$$p_{\alpha\beta} = P\left\{ \left\| Z(x_\alpha) - Z(x_\beta) \right\| \geq d \right\}$$

estimated for independent random points $x_\alpha \in C_\alpha$ and $x_\beta \in C_\beta$. For a classification of textures in homogeneous classes, we expect that the diagonal of the matrix P, with elements $p_{\alpha\beta}$, is close to 0. This can drive the choice of the threshold d, based on the data used for the classification.

Remark 1. *For any pair of regions A_i and A_j in $\mathcal{D}$, where the proportions of pixels belonging to class C_α are p_α^i and p_α^j respectively, and for independent uniform random points $x_i \in A_i, x_j \in A_j$ we have $P\left\{ x_i \in C_\alpha, x_j \in C_\beta \right\} = p_\alpha^i p_\beta^j$. In that case we get $P(A_i, A_j, d) = \sum_{\alpha,\beta} p_\alpha^i p_\beta^j p_{\alpha\beta}$. The calculation of the probabilistic distance between A_i and A_j is made faster after a preliminary storage of the probability matrix P.*

5.2.2 Probabilistic distance and hierarchical segmentation

As mentioned before, we can make use of the probabilistic distance $P(A_i, A_j, d)$ to build a hierarchical segmentation, starting from the lowest probability. Let $P(A_i, A_j, d) < P(A_i, A_k, d)$ and $P(A_i, A_j, d) < P(A_j, A_k, d), \forall k \neq i, k \neq j$. By merging regions A_i and A_j with measures $|A_i|$ and $|A_j|$ (for instance area in $\mathbb{R}^2$ and volume in $\mathbb{R}^3$), we generate a new region $A_l = A_i \cup A_j$. For any k we get $P(A_k, A_l, d) = \frac{|A_i|}{|A_l|} P(A_k, A_i, d) + \frac{|A_j|}{|A_l|} P(A_k, A_j, d)$. Therefore we have

$$P(A_k, A_i, d) \vee P(A_k, A_j, d) \geq P(A_k, A_l, d) \geq P(A_k, A_i, d) \wedge P(A_k, A_j, d)$$
$$> P(A_i, A_j, d).$$

The probabilistic distance increases when merging two classes, so that it can be used as an index in the hierarchy. All remaining values $P(A_k, A_l, d)$ are updated after fusion of two regions, and the process can be iterated. Indeed, $\partial(A_i, A_j) = \text{Inf}\{p, A_i \text{ and } A_j \text{ are included in the same region } A_l\}$ is equivalent to the diameter of the smallest region of the hierarchy containing A_i and A_j, which satisfies the ultrametric inequality required to generate a hierarchy [4]. Alternatively we can use the probabilistic distance involved in every level of merging, to generate an ultrametric distance used to build the hierarchy [8]. The segmentation involved with the probabilistic distance is unsupervised in the general case.

Remark 2. *In the context of segmentation, a partition of the domain $\mathcal{D}$ is obtained by considering all subdomains obtained when cutting the hierarchy at a given level (probability) p. The choice of the threshold p can be driven by the results of the preliminary classification of pixels, using the values of the elements of the matrix P, or by the number of subdomains, that should correspond to the number of textures present in $\mathcal{D}$. An estimate of this number can be derived from the spectral analysis of the matrix of the graph Laplacian derived from the matrix with elements $1 - P(A_i, A_j, d)$, which is an adjacency matrix [12].*

Remark 3. *The proposed hierarchy generates non necessarily connected subdomains, since no adjacency condition is imposed in the choice of regions to be merged. This condition can be required, as is made for the construction of watersheds, by restricting the use of the probabilistic distances to adjacent regions. In addition to connectedness of segmented regions, it reduces the cost of calculations by limiting the number of pairs, instead of considering the full cross-product $A_i \times A_j$. Intermediate constructions can involve the probabilistic distances of iterated adjacent regions. This approach can be followed in the first steps of the segmentation, to reduce the number of regions, and released for the remaining steps of the process, in order to allow for the extraction of non connected regions with the same texture.*

Remark 4. *By integrating the probability $P(A_i, A_j, d)$ with respect to the threshold d, we obtain the average distance $\|Z(x_i) - Z(x_j)\|$ for independent random points $x_i \in A_i$, $x_j \in A_j$. This average can be used in a hierarchy where the aggregation is made according to the average distance criterion [4].*

5.2.3 Combination of probabilistic segmentations

It can be useful to enrich the probabilistic distance by other probability distributions concerning the comparison of the content of two regions, in order to combine them for the segmentation. We will have to restrict the choice of probability distributions on the cross-product $A_i \times A_j$ according to the following definition.

Definition 2. *A probability $P(A_i, A_j)$ is said to be increasing with respect to the fusion of regions, when it satisfies $P(A_k, A_l) \geq P(A_k, A_i) \wedge P(A_k, A_j)$ for any i, j, k, with $A_l = A_i \cup A_j$.*

As shown before, the probabilistic distance satisfies the property given in definition 2. Other probability distributions with the same property will be introduced later.

The property given in definition 2 is satisfied when $P(A_i, A_j)$ is increasing with respect to $\subset$, which means that $P(A_k, A_l) \geq P(A_k, A_i)$ when $A_i \subset A_l$. However, this is not a necessary condition.

We start from two probabilistic segmentations, based on separate aggregation conditions, involving the probability of separation of regions A_i and A_j, $P^1(A_i, A_j)$ and $P^2(A_i, A_j)$. P^1 and P^2 are assumed to own the fusion property of definition 2. These probabilities can be combined according to different rules. For instance:

1. probabilistic independence: $P(A_i, A_j) = P^1(A_i, A_j)P^2(A_i, A_j)$
2. more reliable event: $P(A_i, A_j) = P^1(A_i, A_j) \vee P^2(A_i, A_j)$
3. least reliable event: $P(A_i, A_j) = P^1(A_i, A_j) \wedge P^2(A_i, A_j)$
4. weighting between the two events (with probabilities λ_1 and λ_2): $P(A_i, A_j) = \lambda_1 P^1(A_i, A_j) + \lambda_2 P^2(A_i, A_j)$
5. any combination $P(A_i, A_j) = \Phi(P_1, P_2)(A_i, A_j)$, where P is a probability increasing with respect to the fusion of regions according to definition 2.

These rules are easily extended to more than two conditions of aggregation. We have the following result.

Proposition 2. *The previous rules of combination of the probability of separation of regions satisfy the property given in definition 2.*

Proof. We start from $P(A_i, A_j) = \Phi(P_1, P_2)(A_i, A_j)$. As before, consider $A_l = A_i \cup A_j$ and the condition: $P(A_i, A_j) < P(A_i, A_k)$ and $P(A_i, A_j) < P(A_j, A_k)$, $\forall k \neq i$, $k \neq j$. We have for any region A_k

$$P(A_k, A_l) = \frac{|A_i|}{|A_l|}P(A_k, A_i) + \frac{|A_j|}{|A_l|}P(A_k, A_j).$$

and

$$P(A_k, A_i) \vee P(A_k, A_j) \geq P(A_k, A_l) \geq P(A_k, A_i) \wedge P(A_k, A_j) > P(A_i, A_j)$$

□

5.2.4 Local probability distributions

The regions of the fine partition (or obtained after some steps of aggregation) can be characterized by some local probability distributions.

5.2.4.1 Probabilistic classification
If pixels x_i in region A_i can be attributed to various classes of textures C_α by a probabilistic classification, the probability $p_\alpha^i = P\{x_i \in C_\alpha\}$ can be used as a probabilistic descriptor of A_i. Considering now independent

uniform random points $x_i \in A_i$ and $x_j \in A_j$, the probability (1) can be written

$$P(A_i, A_j) = 1 - P\{x_i \text{ and } x_j \text{ belong to the same texture}\} \tag{3}$$

$$= 1 - p_{ij} = 1 - \sum_\alpha p_\alpha^i p_\alpha^j \tag{4}$$

Denoting $I_\alpha(x)$ the indicator function of class C_α, and $I(x)$ the vector with components $I_\alpha(x)$, we have $P\{\|I_\alpha(x_i) - I_\alpha(x_j)\| = 0\} = P\{x_i \in C_\alpha, x_j \in C_\alpha\} = p_\alpha^i p_\alpha^j$ and therefore $P(A_i, A_j) = P\{\|I(x_i) - I(x_j)\| > 0\}$, so that $P(A_i, A_j)$ defined by (3) is a probabilistic distance corresponding to definition 1. It satisfies the property given in definition 2. This can be checked as follows:using $A_l = A_i \cup A_j$ we obtain

$$p_\alpha^l = \frac{|A_i|}{|A_l|} p_\alpha^i + \frac{|A_j|}{|A_l|} p_\alpha^j$$

and

$$1 - P(A_k, A_l) = \sum_\alpha p_\alpha^k p_\alpha^l = \frac{|A_i|}{|A_l|} \sum_\alpha p_\alpha^i p_\alpha^k + \frac{|A_j|}{|A_l|} \sum_\alpha p_\alpha^j p_\alpha^k$$

$$= \frac{|A_i|}{|A_l|} \left(1 - P(A_i, A_k) \right) + \frac{|A_j|}{|A_l|} \left(1 - P(A_j, A_k) \right)$$

Therefore we get

$$P(A_k, A_l) = \frac{|A_i|}{|A_l|} P(A_i, A_k) + \frac{|A_j|}{|A_l|} P(A_j, A_k)$$

and

$$P(A_k, A_i) \vee P(A_k, A_j) \geq P(A_k, A_l) \geq P(A_k, A_i) \wedge P(A_k, A_j)$$

We now illustrate this situation in the binary case of two textures ($\alpha = 1, 2$). We have

$$p_{ij} = p_1^i p_1^j + (1 - p_1^i)(1 - p_1^j)$$

Note that when $p_1^j = \frac{1}{2}$ (meaning that the region A_i is randomly allocated to texture 1 or to texture 2), we also have $p_{ij} = \frac{1}{2}$.

A simple practical example is obtained by starting with a partition in seven regions A_i with the same measure $|A_i|$ and with the following probabilities p_1^i:

$$\begin{bmatrix} 0.1 & 0.9 & 0.25 & 0.75 & 0.4 & 0.6 & 0.5 \end{bmatrix}$$

The symmetrical probability matrix p_{ij} used in equation (3), with lines ($i = 1$ to $i = 6$) and columns ($j = 2$, to $j = 7$) is given by:

$$\begin{bmatrix} 0.18 & 0.7 & 0.3 & 0.58 & 0.42 & 0.5 \\ & 0.3 & 0.7 & 0.42 & 0.58 & 0.5 \\ & & 0.375 & 0.55 & 0.45 & 0.5 \\ & & & 0.45 & 0.55 & 0.5 \\ & & & & 0.48 & 0.5 \\ & & & & & 0.5 \end{bmatrix}$$

The two regions with the highest probability p_{ij} (namely 0.7 for p_{13} or p_{24}) are merged, and the p_{ij} are updated. Using for instance $A_{1'} = A_1 \cup A_3$, the probability matrix becomes:

$$\begin{bmatrix} 0.24 & 0.338 & 0.565 & 0.435 & 0.5 \\ & 0.7 & 0.42 & 0.58 & 0.5 \\ & & 0.55 & 0.45 & 0.5 \\ & & & 0.48 & 0.5 \\ & & & & 0.5 \end{bmatrix}$$

Setting now $A_{2'} = A_2 \cup A_4$, p_{ij} becomes:

$$\begin{bmatrix} 0.289 & 0.565 & 0.435 & 0.5 \\ & 0.485 & 0.515 & 0.5 \\ & & 0.48 & 0.5 \\ & & & 0.5 \end{bmatrix}$$

In the next step regions $A_{1'}$ and A_5 are merged (since $p_{1'5}$ is the highest probability) to obtain region $A_{1''} = A_1 \cup A_3 \cup A_5$ and the probability matrix for regions $A_{1''}$, $A_{2'}$, A_6 and A_7 becomes:

$$\begin{bmatrix} 0.354 & 0.45 & 0.5 \\ & 0.515 & 0.5 \\ & & 0.5 \end{bmatrix}$$

Regions $A_{2'}$ and A_6 are now merged to form $A_{2''}$ and we are left with a segmentation of $\mathcal{D}$ in two regions $A_{1''}$ and $A_{2''}$ corresponding to textures 2 and 1 respectively. Region A_7 can be left unclassified or can indifferently be merged with $A_{1''}$ or with $A_{2''}$. The obtained segmentation is what is expected in this very simple case, where it would have been obvious to immediately merge regions with $p_1^i > 0.5$ (corresponding to texture 1) and regions with $p_1^i < 0.5$ (corresponding to texture 2).

5.2.4.2   Local granulometric spectrum

As indicated in section 3, granulometric information can be provided after transformation of the image by morphological opening or closing operations. A local granulometric spectrum can be obtained in each region A_i by averaging the components $I_\alpha(x_i)$ or $Z_\alpha(x_i)$ over A_i. After normalization, we obtain local granulometric spectra in classes of sizes C_α where for size α, $p_\alpha^i = P\{x_i \in C_\alpha\}$. This local classification with respect to size can be introduced in the probability (3) to build a hierarchy. We can alternatively combine different granulometries, like opening and closing, for instance by linear combination as proposed in section 5.2.3, to generate a composite hierarchy.

5.2.4.3   Local orientation

Some textures present local orientation [16] which it is convenient to study from a vector field $\vec{V}(x)$, like for instance the gradient vector. To remove the effect of noise in the gradient, a local orientation can be extracted as follows: in every window $K(x)$ is considered the cloud of points M_γ generated by connecting the origin to vector $\vec{V}(x_\gamma)$. The principal axes of inertia of the points M_γ are then extracted [16]. A confidence degree of the orientation is provided by the ratio of the largest eigenvalue of the inertia matrix to its trace. Working with a partition of the domain $\mathcal{D}$, the same approach can be followed with the cloud of points generated by $\vec{V}(x_i)$, $x_i \in A_i$. In a second step, we can characterize the disorientation between A_i and A_j from the scalar product between the two main eigenvectors $\vec{V}_i$ and $\vec{V}_j$. We have, α_{ij} being the angle between $\vec{V}_i$ and $\vec{V}_j$ with Euclidean norms $\left\|\vec{V}_i\right\|$ and $\left\|\vec{V}_j\right\|$, for $\left\|\vec{V}_i\right\| \neq 0$ and $\left\|\vec{V}_j\right\| \neq 0$:

$$\cos^2 \alpha_{ij} = \frac{\left(\vec{V}_i . \vec{V}_j\right)^2}{\left\|\vec{V}_i\right\|^2 \left\|\vec{V}_j\right\|^2}$$

The value $\cos^2 \alpha_{ij}$ is a proximity index between orientations of regions A_i and A_j, while $1 - \cos^2 \alpha_{ij}$ is an angular distance, which can be used in a similar way as $P(A_i, A_j)$ to build an orientation based segmentation, starting from a partition.

An alternative way to account for the disorientation between two regions is to use the probability distribution of the quantity $(1 - \cos^2 \alpha_{ij})$ obtained for $\vec{V}(x_i)$ and $\vec{V}(x_j)$, when x_i and x_j are independent uniform

points in A_i and A_j. We can define an orientational probabilistic distance $P(A_i, A_j, d) = P\left\{(1 - \cos^2 \alpha_{ij}) \geq d\right\}$. In the presence of pixels with a weak gradient $\overrightarrow{V}$ it may be wise to exclude them from the calculation of the orientational probabilistic distance, since the computation of $\cos^2 \alpha_{ij}$ is ill-defined in that case. Alternatively, we can use the probability distribution of the random variable $\left(\overrightarrow{V_i}.\overrightarrow{V_j}\right)^2$.

5.3 Higher order probabilistic segmentation

In the previous section, the fusion construction of a hierarchy was limited to the fusion of pairs of regions, restricting to second order probabilities $P(A_i, A_j)$ and generating binary trees. This can be easily extended to higher order probabilities, to produce more general trees and hierarchies. The resulting aggregation process can be accelerated, as compared to the binary case, at a marginal computational cost.

Consider m regions $A_{i_1}, A_{i_2}, ..., A_{i_m}$. For every pair A_{i_k}, A_{i_l} independent uniform random points x_{i_k}, x_{i_l} and the independent random variables $\left\|Z(x_{i_k}) - Z(x_{i_l})\right\|$ are generated. We decide to merge the m regions when

$$\left(\vee_{i_k, i_l} \left\|Z(x_{i_k}) - Z(x_{i_l})\right\|\right) < d.$$

The probability of this event is given by:

$$P\left\{\vee_{i_k, i_l} \left\|Z(x_{i_k}) - Z(x_{i_l})\right\| < d\right\} = \prod_{i_k, i_l} P\left\{\left\|Z(x_{i_k}) - Z(x_{j_l})\right\| < d\right\} \tag{5}$$

$$= \prod_{i_k, i_l} \left(1 - P(A_{i_k}, A_{i_l}, d)\right)$$

An indexed hierarchy is obtained by sorting the probabilities (5) with decreasing order. Alternatively we can use the m order version of the probabilistic distance sorted with increasing order:

$$P(A_{i_1}, A_{i_2}, ..., A_{i_m}, d) = P\left\{\vee_{i_k, i_l} \left\|Z(x_{i_k}) - Z(x_{i_l})\right\| \geq d\right\} \tag{6}$$

$$= 1 - \prod_{i_k, i_l} \left(1 - P(A_{i_k}, A_{i_l}, d)\right)$$

For instance in the case of a ternary hierarchy the probabilities (5) and (6) become:

$$P\left\{\left\|Z(x_i) - Z(x_j)\right\| \vee \left\|Z(x_i) - Z(x_k)\right\| \vee \left\|Z(x_j) - Z(x_k)\right\| < d\right\} \tag{7}$$

$$= \left(1 - P(A_i, A_j, d)\right)\left(1 - P(A_i, A_k, d)\right)\left(1 - P(A_j, A_k, d)\right)$$

and

$$P(A_i, A_j, A_k, d) = 1 - \left(1 - P(A_i, A_j, d)\right)\left(1 - P(A_i, A_k, d)\right)\left(1 - P(A_j, A_k, d)\right) \tag{8}$$

By construction, the m order probabilities (5) are always lower than the corresponding second order probabilities, and the $P(A_{i_1}, A_{i_2}, ..., A_{i_m}, d)$ are larger than the $P(A_i, A_j, d)$. Once the binary probabilities $P(A_i, A_j, d)$ are available, it is easy to work out higher order probabilities (5, 6, 7, 8). When starting from a partition with r regions, computing the $C_r^m = \frac{r!}{m!(r-m)!}$ order m probabilities may be expensive. However this can be easily performed when restricting the fusion to adjacent regions.

When merging the m regions with the lowest probability

$$P(A_{i_1}, A_{i_2}, ..., A_{i_m}, d)$$

a new region $A_l = A_{i_1} \cup A_{i_2} \cup ... \cup A$ is generated. In a next step, all possible fusions of $m - 1$ regions with A_l have to be considered. We get

$$P(A_{k1}, A_{k_2}, ..., A_{k_{m-1}}, A_l, d) = \frac{|A_{i_1}|}{|A_l|} P(A_{k1}, A_{k_2}, ..., A_{k_{m-1}}, A_{i_1}, d) + ...$$

$$+ \frac{|A_{i_{m-1}}|}{|A_l|} P(A_{k1}, A_{k_2}, ..., A_{k_{m-1}}, A_{i_{m-1}}, d)$$

so that

$$\vee_{i=i_1,\dots i_{m-1}} P(A_{k1}, A_{k_2}, \dots, A_{k_{m-1}}, A_i, d) \geq P(A_{k1}, A_{k_2}, \dots, A_{k_{m-1}}, A_l, d)$$
$$\geq \wedge_{i=i_1,\dots i_{m-1}} P(A_{k1}, A_{k_2}, \dots, A_{k_{m-1}}, A_i, d)$$
$$> P(A_{i_1}, A_{i_2}, \dots, A_{i_m}, d)$$

The probability $P(A_{i_1}, A_{i_2}, \dots, A_{i_m}, d)$ increases when merging m classes, and therefore satisfies the property given in definition 2.

The hierarchy can also be built by combining distance probabilities of various orders m. To illustrate this point, consider the combination of $m = 2$ and $m = 3$. At a given step of the hierarchy the three regions A_i, A_j, and A_k are merged into $A_l = A_i \cup A_j \cup A_k$ if for any p we have

$$P(A_i, A_j, A_k, d) < P(A_i, A_p, d)$$
$$P(A_i, A_j, A_k, d) < P(A_j, A_p, d)$$
$$P(A_i, A_j, A_k, d) < P(A_k, A_p, d)$$

It comes for p,

$$P(A_l, A_p, d) = \frac{|A_i|}{|A_l|} P(A_i, A_p, d) + \frac{|A_j|}{|A_l|} P(A_j, A_p, d) + \frac{|A_k|}{|A_l|} P(A_k, A_p, d)$$

and therefore

$$P(A_l, A_p, d) > P(A_i, A_p, d) \wedge P(A_j, A_p, d) \wedge P(A_k, A_p, d)$$
$$> P(A_i, A_j, A_k, d)$$

so that the index of the hierarchy increases after the fusion of the regions A_i, A_j, and A_k.

Remark 5. *There is a connection between the m order probabilistic distance and some image transformations in the case of a digitized image. Consider a neighborhood B(x) with m pixels. The criterion $\left(\vee_{i_k,i_l} \|Z(x_{i_k}) - Z(x_{i_l})\|\right)$ becomes*

$$Y(x) = \vee_{x_1 \in B(x), x_2 \in B(x)} \|Z(x_1) - Z(x_2)\|$$

In the case of a scalar image $Z(x)$ we have $Y(x) = \vee_{y \in B} Z(y) - \wedge_{y \in B} Z(y) = Z(x) \oplus B - Z(x) \ominus B$, and $Y(x)$ is a standard morphological gradient.

For a probabilistic classification of textures, the result (3) is easily extended to higher orders:

$$P(A_{i_1}, A_{i_2}, \dots, A_{im}) = 1 - P\{x_{i_1}, x_{i_2}, \dots, x_{i_m} \text{ belong to the same texture}\} \qquad (9)$$
$$= 1 - \sum_\alpha \prod p_\alpha^{i_1} p_\alpha^{i_2} \dots p_\alpha^{i_m}$$

This is illustrated by a short numerical example for $r = 5$ regions and two textures, as earlier, starting from the following probabilities p_1^i:

$$\begin{bmatrix} 0.1 & 0.9 & 0.25 & 0.75 & 0.5 \end{bmatrix}$$

To the second order, the symmetrical probability matrix p_{ij} with lines ($i = 1$ to $i = 4$) and columns ($j = 2$, to $j = 5$) is given by:

$$\begin{bmatrix} 0.18 & 0.7 & 0.3 & 0.5 \\ & 0.3 & 0.7 & 0.5 \\ & & 0.375 & 0.5 \\ & & & 0.5 \end{bmatrix}$$

We can decide to merge regions of highest probabilities $p_{ij} = 0.7$, namely regions A_1 and A_3 (with the dominant texture 2) and regions A_2 and A_4 (with the dominant texture 1), region A_5 being left, or indifferently merged with the two groups.

To the third order, it is easy to compute the 10 following probabilities p_{ijk}:

$$
\begin{bmatrix}
p_{123} & p_{124} & p_{125} & p_{134} & p_{135} & p_{145} & p_{234} & p_{235} & p_{245} & p_{345} \\
0.09 & 0.09 & 0.09 & 0.1875 & 0.35 & 0.15 & 0.1875 & 0.17 & 0.35 & 0.1875
\end{bmatrix}
$$

From the highest probability $p_{135} = p_{245} = 0.35$ we can decide to merge regions A_1, A_3 and A_5 or regions A_2, A_4 and A_5, which is consistent to what is expected from the probabilities p_1^i and from the second order probabilities p_{ij}.

5.4 Probabilistic distances between sets

As a particular case of a probabilistic distance, it is interesting to consider the distribution of the Euclidean distance $d(x, y)$ between two random points x and y.

Proposition 3. *Consider two uniform random points x and y located in the set $A_i \subset \mathbb{R}^n$. The distribution of distances $d(x, y)$ is given by*

$$
P\left\{d(x, y) < d\right\} = \int_{A_i} \frac{\left|B_x(d) \cap A_i\right|}{\left|A_i\right|^2} dx \tag{10}
$$

where $B_x(d)$ is the ball with center x and radius d.

Proof. Given point x in A_i, we have $d(x, y) < d \iff y \in B_x(d)$ and $y \in A_i$. The probability of this event is given by $\frac{|B_x(d) \cap A_i|}{|A_i|}$. After deconditioning with respect to the uniform location of x in A_i, we obtain the result (10). □

Proposition 4. *Consider two uniform random points $x \in A_i$ and $y \in A_j$. The distribution of distances $d(x, y)$ is given by*

$$
P\left\{d(x, y) < d\right\} = \int_{A_i} \frac{\left|B_x(d) \cap A_j\right|}{\left|A_i\right|\left|A_j\right|} dx = \int_{A_j} \frac{\left|B_y(d) \cap A_i\right|}{\left|A_i\right|\left|A_j\right|} dy \tag{11}
$$

where $B_x(d)$ is the ball with center x and radius d. By construction, it satisfies the criterion of fusion given in definition 2.

Proof. Given a random point x in A_i, we have $d(x, y < d \iff y \in B_x(d)$ and $y \in A_j$. The probability of this event is given by $\frac{|B_x(d) \cap A_j|}{|A_j|}$. After deconditioning with respect to the uniform location of x in A_i, we obtain the result (11). Alternatively we can compute this probability, given a random point y in A_j. The two expressions are equal since, noting $1_{A_i}(x)$, $1_{A_j}(y)$ and $1_{B_x(d)}(y)$ the indicator functions of the sets A_i, A_j and $B_x(d)$, we get:

$$
\int_{A_i} \frac{\left|B_x(d) \cap A_j\right|}{\left|A_i\right|\left|A_j\right|} dx = \frac{1}{\left|A_i\right|\left|A_j\right|} \int_{\mathbb{R}^n} 1_{A_i}(x) 1_{A_j}(y) 1_{B_x(d)}(y) dy dx
$$

$$
= \frac{1}{\left|A_i\right|\left|A_j\right|} \int_{\mathbb{R}^n} 1_{A_i}(x) 1_{A_j}(y) 1_{B_0(d)}(y - x) dy dx
$$

$$
= \frac{1}{\left|A_i\right|\left|A_j\right|} \int_{\mathbb{R}^n} 1_{A_i}(x) 1_{A_j}(y) 1_{B_y(d)}(x) dy dx \int_{A_j} \frac{\left|B_y(d) \cap A_i\right|}{\left|A_i\right|\left|A_j\right|} dy
$$

□

The probability distributions (10) and (11) involve average values of the measures $|B_x(d) \cap A_i|$, $|B_x(d) \cap A_j|$ and $|B_y(d) \cap A_i|$. In practice, they can be estimated from sampling the locations of the ball $B_x(d)$. These probabilities of distances can be used for some morphological characterization of each set A_i and for each pair A_i, A_j. When considering m regions in a fine partition, they allow us the calculation of m order probabilities, after introduction in the probability given in (5) to calculate $P\left\{\vee_{i_k, i_l} d(x_{i_k}, x_{i_l}) < d\right\}$.

These probabilities of distance between sets carry some morphological content on the partition, which can be introduced in a hierarchical segmentation based on the shape and spatial distribution of regions A_i. For a given pair of regions A_i and A_j, the probability of fusion decreases when their spatial separation increases, so that the process promotes the fusion of neighbour regions. This information can be combined with the previously defined probability distances with a textural content.

Another way to introduce morphological data on regions of the partition is obtained by the introduction of random markers.

5.5 Use of random markers

Following the approach proposed for the stochastic watershed [2] we can use random markers to randomly select regions for which the previous probabilistic segmentation will be performed. Doing this, some morphological content on the regions of the hierarchy and on their location in $\mathcal{D}$ is accounted for, in addition to the previous probabilistic textural information. The aim of this section is to calculate the probability $P_R(A_i, A_j)$ of selection of two regions (A_i, A_j) by random markers. We follow the results introduced in [15].

5.5.1 Reminder on random allocation of germs

We use n_g random points (or germs) x_k with independent uniform distributions in the domain $\mathcal{D}$ containing r regions. The probability p_i for a germ to fall in A_i is given by

$$p_i = \frac{|A_i|}{|\mathcal{D}|}, \text{ with } \sum_{i=1}^{r} p_i = 1.$$

By construction, the allocation of germs in the regions of a partition follows a multinomial distribution (N_i being the random number of germs in A_i) with multivariate generating function:

$$G(s_1, s_2, \ldots, s_r) = E\left\{s_1^{N_1} s_2^{N_2} \ldots s_r^{N_r}\right\}$$

Starting with $n_g = 1$ (use of a single germ):

$$G_1(s_1, s_2, \ldots, s_r) = p_1 s_1 + p_2 s_2 + \ldots + p_r s_r$$

For $n_g \geq 1$, the numbers N_i are the sum of n independent binary random variables, and

$$G_{n_g}(s_1, s_2, \ldots, s_r) = G_1(s_1, s_2, \ldots, s_r)^{n_g} = (p_1 s_1 + p_2 s_2 + \ldots + p_r s_r)^{n_g}$$

so that

$$P\{N_1 = k_1, N_2 = k_2, \ldots, N_r = k_r\} = \frac{n_g!}{k_1! k_2! \ldots k_r!} p_1^{k_1} p_2^{k_2} \ldots p_r^{k_r} \tag{12}$$

$$\text{with } k_1 + k_2 + \ldots + k_r = n_g$$

An interesting case is asymptotically obtained when $|\mathcal{D}| \to \infty$ and $n_g \to \infty$, with $\frac{n_g}{|\mathcal{D}|} \to \theta$. For these conditions, the multinomial distribution converges towards the multivariate Poisson distribution. We have:

$$\log G_{n_g}(s_1, s_2, \ldots, s_r) = n_g \log(p_1 s_1 + p_2 s_2 + \ldots + p_r s_r)$$
$$= n_g \log(1 + p_1(s_1 - 1) + p_2(s_2 - 1) + \ldots + p_r(s_r - 1))$$

and

$$\log G_{n_g}(s_1, s_2, ..., s_r) \to \theta\,|A_1|\,(s_1 - 1) + \theta\,|A_2|\,(s_2 - 1) + ... + \theta\,|A_r|\,(s_r - 1)$$

so that:

$$\lim_{n_g \to \infty} G_{n_g}(s_1, s_2, ..., s_r) = \pi_{i=1}^{i=r} \exp(\theta\,|A_i|\,(s_i - 1))$$

The random numbers $N_1, N_2, ..., N_r$ are independent Poisson random variables with intensities $\theta_i = \theta\,|A_i|$:

$$P\{N_i = k\} = \frac{\theta_i^k}{k!} \exp -(\theta_i) \tag{13}$$

Using Poisson points as markers, the number of germs for each realization follows a Poisson distribution with parameter $\theta\,|\mathcal{D}|$.

5.5.2 Calculation of the probability $P_R(A_i, A_j)$ for point markers

Random markers are used to select regions of a partition by reconstruction. With this process, the reconstructed regions for any realization of the random germs are left intact, while regions without germs are merged. Considering many realizations of the germs, we can compute the probability $P_R(A_i, A_j)$ for the two regions to remain separate.

Proposition 5. *For n_g independent uniformly distributed random germs, the probability $P_R(A_i, A_j)$ for the two regions A_i, A_j to remain separate is given by:*

$$P_R(A_i, A_j) = 1 - \left(1 - p_i - p_j\right)^{n_g} \tag{14}$$

Proof. The pair (A_i, A_j) is merged $\iff \{N_i = 0 \text{ and } N_j = 0\} \iff N_i + N_j = 0 \iff N(A_i \cup A_j) = 0$.   □

Working on images, the probabilities $P_R(A_i, A_j)$ computed for all pairs (A_i, A_j) are easily ranked in increasing order. A hierarchical fusion of regions is obtained by starting with the lowest probability $P_R(A_i, A_j)$. After fusion of two regions with $A_l = A_i \cup A_j$ the probabilities $P_R(A_k, A_l)$ are updated. The pair (A_k, A_l) is merged $\iff \{N_k = 0 \text{ and } N_l = 0\} \iff N_k + N_l = 0 \iff N(A_k \cup A_l) = 0 \iff N(A_i \cup A_j \cup A_k) = 0$. We get:

$$P_R(A_k, A_l) = 1 - \left(1 - p_i - p_j - p_k\right)^{n_g} > P_R(A_i, A_j)$$

and the probability $P_R(A_k, A_l)$ is increasing with respect to the fusion of regions as in definition 2.

In general no conditions of connectivity or of adjacency of regions are required for the fusion process. It is easy to force the connectivity by working on connected components of regions, or to limit the fusion to adjacent regions.

The random germs can be generated by a Poisson point process.

Proposition 6. *For Poisson point germs with intensity θ, the probability $P_R(A_i, A_j)$ for the two regions A_i, A_j to remain separate is given by:*

$$P_R(A_i, A_j) = 1 - \exp\left[-\theta\left(|A_i| + |A_j|\right)\right] \tag{15}$$

and $P_R(A_k, A_l)$ is increasing with respect to the fusion of regions as in definition 2.

The morphological content in the probabilities (14, 15) only depends on the Lebesgue measure (area in $\mathbb{R}^2$ and volume in $\mathbb{R}^3$) of regions. It increases with the measure of regions, larger regions resisting more to fusion. For a pair of regions, $P_R(A_i, A_j)$ is maximal when $|A_i| = |A_j|$, so that the random markers hierarchy tends to generate by fusion regions with homogeneous sizes, the regions with lower measure disappearing first.

5.5.3 Calculation of the probability $P_R(A_i, A_j)$ for Poisson lines and Poisson flats markers

It can be interesting to obtain other weightings of regions with a probabilistic meaning, like the perimeter in $\mathbb{R}^2$ or the surface area in $\mathbb{R}^3$. Restricting to the Poisson case, it is easy to make this extension, provided use is made of appropriate markers. For this purpose, we will consider now isotropic Poisson lines in $\mathbb{R}^2$, isotropic Poisson planes and Poisson lines in $\mathbb{R}^3$ [14, 22]. Oriented Poisson lines in $\mathbb{R}^3$ were used as markers in the context of the stochastic watershed (and so with another type of probability), and applied to the segmentation of granular structures [10].

Proposition 7. *Consider stationary isotropic Poisson lines with intensity λ as random markers in $\mathbb{R}^2$. The probability $P_R(A_i, A_j)$ for the two regions A_i, A_j to remain separate is expressed as a function of the average projected length l of the projection of $A_i \cup A_j$ in directions ω:*

$$P_R(A_i, A_j) = 1 - \exp\left[-\lambda \int_0^\pi l(A_i(\omega) \cup A_j(\omega))d\omega\right] \tag{16}$$

when $A_i \cup A_j$ is a connected set, $P_R(A_i, A_j)$ is given by:

$$P_R(A_i, A_j) = 1 - \exp\left[-\lambda L\left(\mathcal{C}(A_i \cup A_j)\right)\right] \tag{17}$$

where L is the perimeter and $\mathcal{C}(A_i \cup A_j)$ is the convex hull of $A_i \cup A_j$.

Proposition 8. *Consider stationary isotropic Poisson lines with intensity λ as random markers in $\mathbb{R}^3$. The probability $P_R(A_i, A_j)$ for the two regions A_i, A_j to remain separate is expressed as a function of the average projected area A of the projection of $A_i \cup A_j$ in directions ω:*

$$P_R(A_i, A_j) = 1 - \exp\left[-\lambda \int_0^{2\pi ster} A(A_i(\omega) \cup A_j(\omega))d\omega\right] \tag{18}$$

when $A_i \cup A_j$ is a connected set, $P_R(A_i, A_j)$ is given by:

$$P_R(A_i, A_j) = 1 - \exp\left[-\lambda \frac{\pi}{4} S(\mathcal{C}(A_i \cup A_j))\right] \tag{19}$$

where S is the surface area and $\mathcal{C}(A_i \cup A_j)$ is the convex hull of $A_i \cup A_j$. For random markers in $\mathbb{R}^3$ made of stationary isotropic Poisson planes with intensity λ, the probability $P_R(A_i, A_j)$ for the two regions A_i, A_j to remain separate is expressed as a function of the average projected length l of the projection of $A_i \cup A_j$ in directions ω:

$$P_R(A_i, A_j) = 1 - \exp\left[-\lambda \int_0^{2\pi ster} l(A_i(\omega) \cup A_j(\omega))d\omega\right] \tag{20}$$

when $A_i \cup A_j$ is a connected set, the probability $P_R(A_i, A_j)$ is given as a function of the integral of mean curvature $\mathcal{A}$ by:

$$P_R(A_i, A_j) = 1 - \exp\left[-\lambda \mathcal{A}(\mathcal{C}(A_i \cup A_j))\right] \tag{21}$$

It is possible to combine various types of Poisson markers (points and lines in $\mathbb{R}^2$, points, planes and lines in $\mathbb{R}^3$) with their own intensities. For instance, when $A_i \cup A_j$ is a connected set in $\mathbb{R}^2$, we obtain:

$$P_R(A_i, A_j) = 1 - \exp\left[-\left\{\theta\left(|A_i| + |A_j|\right) + \lambda L\left(\mathcal{C}(A_i \cup A_j)\right)\right\}\right] \tag{22}$$

where a weighting by the area and the perimeter of the regions acts for the segmentation. Similarly in $\mathbb{R}^3$ is introduced a weighting of the volume, and the surface area and integral of mean curvature of $\mathcal{C}(A_i \cup A_j)$ in the process of segmentation.

5.5.4 Calculation of the probability $P_R(A_i, A_j)$ for compact markers

Further morphological information on the regions can be accounted for when introducing compact random markers (not necessarily connected). In the process of selection of regions of a partition by reconstruction, point markers are replaced by a compact grain A' located on Poisson points, and generating a Boolean model A. We have:

Proposition 9. *For compact markers A' generating a Boolean model with intensity θ, the probability $P_R(A_i, A_j)$ for the two regions A_i, A_j to remain separate is given by:*

$$P_R(A_i, A_j) = 1 - \exp\left[-\theta\left|(\check{A}_i \oplus A') \cup (\check{A}_j \oplus A')\right|\right] = 1 - \exp\left[-\theta\left(\left|\check{A}_i \oplus A'\right| + \left|\check{A}_j \oplus A'\right| - \left|(\check{A}_i \oplus A') \cap (\check{A}_j \oplus A')\right|\right)\right] \tag{23}$$

Proof. The pair (A_i, A_j) is merged $\iff A_i \cup A_j$ is outside the Boolean model with primary grain A'. The expression (23) is the Choquet capacity $T(K)$ of the Boolean model when $K = A_i \cup A_j$. □

The compact markers can be random sets (for instance spheres wit a random radius). In that case, the measures $||$ are replaced by their mathematical expectations with respect to the random set A'. Using for A' a ball with radius ρ, $P_R(A_i, A_j)$ increases until a constant value when the distance between A_i and A_j increases from 0 to 2ρ: the probability to merge two regions is higher when their distance is lower.

5.5.5 Combination of textural and of morphological information

We can now combine the use of random markers, conveying morphological content on the partition and on its evolution in the hierarchy, to the previous textural content (probabilistic distance, or local probability information. For instance, we can decide to merge two regions when they are not reconstructed by markers (with a marker dependent probability $1 - P_R(A_i, A_j)$) and the textures they enclose are similar (with a probability $1 - P(A_i, A_j, d)$). In this context the probability p_{ij} (1) becomes

$$P(A_i, A_j) = \Phi(P_R, P_d)(A_i, A_j) \tag{24}$$
$$= P_R(A_i, A_j) + P(A_i, A_j, d) - P_R(A_i, A_j)P(A_i, A_j, d)$$

By construction, this composite probability is increasing with respect to the fusion of regions as in definition 2, and will generate a hierarchy for the segmentation, by updating each terms of (24) according to the previous rules.

Alternatively, the probabilities of distance between sets introduced in section 5.4 can replace $P_R(A_i, A_j)$ in the formulation (24).

Other textural information can be introduced in the probability (24), such as orientational information, as discussed in section 5.2.4.3, or extinction values: in the case of watershed segmentation, the flooding algorithm relies on the minimal values z_{ij} of the function to be flooded, for instance grad $(Z(x))$, on boundaries A_{ij} between adjacent regions A_i and A_j. Let $F(z)$ be the cumulative distribution function of the extinction values. It is equivalent to sort the values z_{ij} or $F(z_{ij})$, since the distribution function is a monotonous transformation of data. Using $F(z_{ij})$ instead of z_{ij} gives a probabilistic content to the flooding. In this context, $1 - F(z_{ij})$ is a probabilistic distance between A_i and A_j. Consider the regions A_i, A_j and A_k. A_j and A_k are merged by flooding, if $z_{jk} < z_{ij}$ and $z_{jk} < z_{ik}$, and consequently $z_{jk} < z_{ij} \wedge z_{ik}$. After fusion, we obtain $A_l = A_j \cup A_k$ and $z_{il} = z_{ij} \wedge z_{ik}$. Therefore, the probabilistic distance $1 - F(z_{il})$ increases by fusion, as required in definition 2. Combining this probabilistic distance to the various probabilities $P_R(A_i, A_j)$ generated by random markers provides a hierarchical segmentation, even outside of the field of textures. This hierarchical segmentation resulting from a progressive fusion of adjacent regions is close to the segmentation given by the stochastic watershed, but remains different.

5.6 Random markers and higher order fusion of regions

We now consider random germs and decide that regions without germs are merged. For the fusion of m regions, the previous second order results are easily extended.

Proposition 10. *For n_g independent uniformly distributed random germs, the probability $P_R(A_{i_1}, A_{i_2}, ..., A_{i_m})$ for the m regions $A_{i_1}, A_{i_2}, ..., A_{i_m}$ to remain separate is given by:*

$$P_R(A_{i_1}, A_{i_2}, ..., A_{i_m}) = 1 - \left(1 - p_{i_1} - p_{i_2} - .... - p_{i_m}\right)^{n_g} \tag{25}$$

Proof. The regions $A_{i_1}, A_{i_2}, ..., A_{i_m}$ are merged $\Longleftrightarrow$

$$\left\{N_{i_1} = 0 \text{ and } N_{i_2} = 0 \text{ and } .... N_{i_m} = 0\right\}$$
$$\Longleftrightarrow N_{i_1} + N_{i_2} + ... + N_{i_m} = 0 \Longleftrightarrow N(A_{i_1} \cup A_{i_2} \cup ... \cup A_{i_m}) = 0.$$

$\square$

A hierarchical fusion of regions is obtained by starting with the lowest probability $P_R(A_{i_1}, A_{i_2}, ..., A_{i_m})$. After fusion of m regions with $A_l = A_{i_1} \cup A_{i_2} \cup ... \cup A_{i_m}$ the probabilities $P_R(A_{i_1}, A_{i_2}, ..., A_{i_m})$ are updated. The regions $(A_{k1}, A_{k_2}, ..., A_{k_{m-1}}, A_l)$ are merged

$$\Longleftrightarrow \left\{N_{k_1} = 0, N_{k_2} = 0,..., N_{k_{m-1}} = 0 \text{ and } N_l = 0\right\}$$
$$\Longleftrightarrow N_{k_1} + N_{k_2} + ... + N_{k_{m-1}} + N_l = 0$$
$$\Longleftrightarrow N(A_{k1} \cup A_{k_2}, \cup .... \cup A_{k_{m-1}} \cup A_l) = 0.$$

We get:

$$P_R(A_{k1}, A_{k_2}, ..., A_{k_{m-1}}, A_l) = 1 - \left(1 - p_{i_1} - p_{i_2} - .... - p_{i_m} - p_{k_1} - p_{k_2} - ... - p_{k_{m-1}}\right)^{n_g}$$
$$> P_R(A_{i_1}, A_{i_2}, ..., A_{i_m})$$

and the probability $P_R(A_{i_1}, A_{i_2}, ..., A_{i_m})$ is increasing with respect to the fusion of regions as in definition 2.

The extension of previous results to various markers (Poisson points, compact markers, Poisson lines or Poisson planes) is straightforward. We get the following results.

Proposition 11. *For Poisson point germs with intensity θ, the probability*

$$P_R(A_{i_1}, A_{i_2}, ..., A_{i_m})$$

for the m regions $A_{i_1}, A_{i_2}, ..., A_{i_m}$ to remain separate is given by:

$$P_R(A_{i_1}, A_{i_2}, ..., A_{i_m}) = 1 - \exp\left[-\theta\left(\left|A_{i_1}\right| + \left|A_{i_2}\right| + ... + \left|A_{i_m}\right|\right)\right] \tag{26}$$

and $P_R(A_{i_1}, A_{i_2}, ..., A_{i_m})$ is increasing with respect to the fusion of regions as in definition 2.

Proposition 12. *For compact markers A' generating a Boolean model with intensity θ, the probability*

$$P_R(A_{i_1}, A_{i_2}, ..., A_{i_m})$$

for the m regions $A_{i_1}, A_{i_2}, ..., A_{i_m}$ to remain separate is given by:

$$P_R(A_{i_1}, A_{i_2}, ..., A_{i_m}) = 1 - \exp\left[-\theta\left|(\check{A}_{i_1} \oplus A') \cup (\check{A}_{i_2} \oplus A') \cup ... \cup (\check{A}_{i_m} \oplus A')\right|\right] \tag{27}$$

Proposition 13. *Consider stationary isotropic Poisson lines with intensity λ as random markers in $\mathbb{R}^2$. The probability*

$$P_R(A_{i_1}, A_{i_2}, ..., A_{i_m})$$

for the m regions $A_{i_1}, A_{i_2}, ..., A_{i_m}$ to remain separate is expressed as a function of the average projected length l of the projection of $A_{i_1} \cup A_{i_2} \cup ... \cup A_{i_m}$ in directions ω:

$$P_R(A_{i_1}, A_{i_2}, ..., A_{i_m}) = 1 - \exp\left[-\lambda \int_0^\pi l(A_{i_1}(\omega) \cup A_{i_2}(\omega) \cup ... \cup A_{i_m}(\omega))d\omega\right] \tag{28}$$

when $A_{i_1} \cup A_{i_2} \cup ... \cup A_{i_m}$ is a connected set, it is given by:

$$P_R(A_{i_1}, A_{i_2}, ..., A_{i_m}) = 1 - \exp\left[-\lambda L\left(\mathcal{C}(A_{i_1} \cup A_{i_2} \cup ... \cup A_{i_m})\right)\right] \tag{29}$$

where L is the perimeter and $\mathcal{C}(A_{i_1} \cup A_{i_2} \cup ... \cup A_{i_m})$ is the convex hull of $A_{i_1} \cup A_{i_2} \cup ... \cup A_{i_m}$.

Proposition 14. *Consider stationary isotropic Poisson lines with intensity λ as random markers in $\mathbb{R}^3$. The probability*

$$P_R(A_{i_1}, A_{i_2}, ..., A_{i_m})$$

for the m regions $A_{i_1}, A_{i_2}, ..., A_{i_m}$ to remain separate is expressed as a function of the average projected area A of the projection of $A_{i_1} \cup A_{i_2} \cup ... \cup A_{i_m}$ in directions ω:

$$P_R(A_{i_1}, A_{i_2}, ..., A_{i_m}) = 1 - \exp\left[-\lambda \int_0^{2\pi ster} A(A_{i_1}(\omega) \cup A_{i_2}(\omega) \cup ... \cup A_{i_m}(\omega))d\omega\right] \tag{30}$$

when $A_{i_1} \cup A_{i_2} \cup ... \cup A_{i_m}$ is a connected set, it is given by:

$$P_R(A_{i_1}, A_{i_2}, ..., A_{i_m}) = 1 - \exp\left[-\lambda\frac{\pi}{4}S(\mathcal{C}(A_{i_1} \cup A_{i_2} \cup ... \cup A_{i_m}))\right] \tag{31}$$

For random markers in $\mathbb{R}^3$ made of stationary isotropic Poisson planes with intensity λ, the probability

$$P_R(A_{i_1}, A_{i_2}, ..., A_{i_m})$$

for the m regions $A_{i_1}, A_{i_2}, ..., A_{i_m}$ to remain separate is expressed as a function of the average projected length l of the projection of $A_{i_1} \cup A_{i_2} \cup ... \cup A_{i_m}$ in directions ω:

$$P_R(A_{i_1}, A_{i_2}, ..., A_{i_m}) = 1 - \exp\left[-\lambda \int_0^{2\pi ster} l(A_{i_1}(\omega) \cup A_{i_2}(\omega) \cup ... \cup A_{i_m}(\omega))d\omega\right] \tag{32}$$

when $A_{i_1} \cup A_{i_2} \cup ... \cup A_{i_m}$ is a connected set, it is given by:

$$P_R(A_{i_1}, A_{i_2}, ..., A_{i_m}) = 1 - \exp\left[-\lambda\mathcal{A}(\mathcal{C}(A_{i_1} \cup A_{i_2} \cup ... \cup A_{i_m}))\right] \tag{33}$$

As previously (cf. the combination(24)), the morphological content carried by random markers can be combined to the textural content given by the probability (6) in the construction of a hierarchy based on m order probabilities. Alternatively, m order probabilities of distance between regions (section 5.4) can play the same role.

6 Conclusion

The probabilistic hierarchical segmentation tools introduced in this work are flexible enough to handle various types of textures (scalar or multivariate) and their spatial distribution, by progressively merging regions of a fine partition. Combining appropriate morphological operations and texture classification, successfully implemented in previous studies on texture segmentation mentioned in the references, supervised or unsupervised texture segmentations can be obtained. A probabilistic distance between regions, carrying statistical

information on textures, is defined. From this distance, hierarchies involving progressive binary or multiple fusions of regions with similar textures are built. Additionally, morphological information on the regions of the fine partition and on merged regions of the hierarchy can be accounted for in the process, through the use of a probabilistic distance between regions of the initial partition, or of various kinds of random markers. In each situation, the required probabilities are computed in a closed form by simple algebra, from the set of probability distributions (2) estimated on the initial partition, as a function of the content and of the location of the regions of the partition evolving during the construction of the hierarchy.

References

[1] Alpert S, Galun M, Basri R, Brandt A. (2007) Image segmentation by probabilistic bottom-up aggregation and cue integration. In Computer Vision and Pattern Recognition, CVPR'07. IEEE Conference, 1-8

[2] Angulo J, Jeulin D (2007) Stochastic watershed segmentation, Proc. ISMM'2007, 8th International Symposium on Mathematical Morphology, Rio de Janeiro, Brazil, October 10-13 2007, Banon G., Barrera J., Braga-Neto U. (eds), ISBN 978-85-17-00032-4, 265-276

[3] Aubert A, Jeulin D (2000) Classification morphologique de surfaces rugueuses, Revue de Métallurgie - CIT/Sience et Génie des Matériaux Feb 2000 (2000), 253–262

[4] Benzecri JP (1973) L'analyse des données, Dunod, Paris, TIB n°3, §3-4, pp. 133-149; TIB n°4, §2.3, pp. 180-183

[5] Beucher S, Lantuéjoul Ch (1979) Use of watersheds in contour detection. In: International workshop on image processing, real-time edge and motion detection

[6] Cord A, Jeulin D, Bach F (2007) Segmentation of random textures by morphological and linear operators, Proc. ISMM'2007, 8th International Symposium on Mathematical Morphology, Rio de Janeiro, Brazil, October 10-13 2007, Banon G., Barrera J., Braga-Neto U. (eds), ISBN 978-85-17-00032-4, 387-398

[7] Cord A, Bach F, Jeulin D (2010) Texture classification by statistical learning from morphological image processing. Application to metallic surfaces, Journal of Microscopy, 239, 159-166

[8] Duda RO, Hart PE (1973) Pattern recognition and scene analysis, Wiley, New York, 236-237

[9] Gillibert L, Jeulin D (2011) Stochastic Multiscale Segmentation Constrained by Image Content, Proc. ISSM 2011, Soille P., Pesaresi M., Ouzounis G.K. (eds), Lecture Notes in Computer Science 6671, Springer Verlag, 132-142

[10] Gillibert L, Peyrega Ch, Jeulin D, Guipont V, Jeandin M (2012) 3D Multiscale Segmentation and Morphological Analysis of X-ray Microtomography from Cold-sprayed Coatings, Journal of Microscopy, vol. 248, Part 2, 187-199. doi: 10.1111/j.1365-2818.2012.03655.x

[11] Gratin C, Vitria J, Moreso F, Seron D (1994) Texture classification using neural networks and local granulometries. In: Mathematical Morphology and its Applications to Image Processing, J. Serra and P. Soille (eds), Kluwer Academic Publishers, 309-316

[12] Hastie T, Tibshirani R, Friedman J (2001) The Elements of Statistical Learning, Data Mining, Inference, and Prediction, Springer pp. 544-547

[13] Hummel RA, Zucker SW (1983) On the Foundations of Relaxation Labeling Processes, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. PAMI-5, N° 3, 267-287

[14] Jeulin D (1991) Modèles Morphologiques de Structures Aléatoires et de Changement d'Echelle. Thèse de Doctorat d'Etat ès Sciences Physiques, Université de Caen

[15] Jeulin D (2008) Remarques sur la segmentation probabiliste, N-10/08/MM, Internal report, CMM, Mines ParisTech

[16] Jeulin D, Moreaud M (2008) Segmentation of 2D and 3D textures from estimates of the local orientation, Image Analysis and Stereology, Vol. 27, 183-192

[17] Jeulin D (2015) Probabilistic Hierarchical Morphological Segmentation of Textures. In: International Symposium on Mathematical Morphology and Its Applications to Signal and Image Processing, J. A. Benediktsson, J. Chanussot, L. Najman, H. Talbot (eds), Springer International Publishing, 313-324

[18] Karoui I, Fablet R, Boucher J M, Augustin J M (2006) Region-based image segmentation using texture statistics and level-set methods. In Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference, Vol. 2

[19] Karoui I, Fablet R, Boucher J M, Pieczynski W, Augustin J M (2008) Fusion of textural statistics using a similarity measure: application to texture recognition and segmentation, Pattern Analysis and Applications, 11(3-4), 425-434

[20] Manjunath BS, Simchony T, Chellappa R (1990) Stochastic and Deterministic Networks for Texture Segmentation, IEEE Transactions on Acoustics, Speech, and Signal ProcessingVol. 38, N° 6, 1039-1049

[21] Matheron G (1967) Eléments pour une théorie des milieux poreux, Masson, Paris

[22] Matheron G (1975) Random sets and integral geometry, Wiley, New York

[23] Meyer F, Beucher S (1990) Morphological segmentation, Journal of Visual Communication and Image Representation, Volume 1, Issue 1, 21–46

[24] Meyer F, Stawiaski J (2010) A stochastic evaluation of the contour strength, DAGM-Symposium'10, 513-522

[25] Noyel G, Angulo J, Jeulin D (2007) Morphological segmentation of hyperspectral images, Image Analysis and Stereology, Vol. 26, 101-109

[26] Noyel, G, Angulo J, Jeulin D (2008) Classification-driven stochastic watershed. Application to multispectral segmentation, Proc. Fourth European Conference on Color in Graphics, Imaging and Vision (CGIV 2008), 471 - 476

[27] Noyel G, Angulo J., Jeulin D (2010) A new spatio-spectral morphological segmentation for multispectral remote sensing images, International Journal of Remote Sensing, Vol. 31, Issue 22, 5895-5920

[28] Peleg S (1980) A New Probabilistic Relaxation Scheme, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. PAMI-2, N° 4, 362–369

[29] Pesaresi M, Benediktsson J (2001) A new approach for the morphological segmentation of high resolution satellite imagery, Geoscience and Remote Sensing, Vol. 39 , Issue 2, 309–320

[30] Rosenfeld A, Hummel RA, Zucker SW (1976) Scene Labeling by Relaxation Operations, IEEE Transactions on Systems, Man, and Cybernetics, Vol. SMC-6, N° 6, 420-433

[31] Serra J (1982) Image Analysis and Mathematical Morphology, Academic Press, London

[32] Sivakumar K, Goutsias J (1994) Monte Carlo estimation of morphological granulometric discrete size distributions. In: Mathematical Morphology and its Applications to Image Processing, J Serra and P Soille (eds), Kluwer Academic Publishers, 233-240