Introduction to the Multifractal Analysis of Images

Jacques Lévy Véhel

Introduction to the multifractal analysis of images

Jacques LÉVY VÉHEL

INRIA, 78153 Le Chesnay Cedex, FRANCE
e-mail : jlv@bora.inria.fr
http://www-rocq.inria.fr/fractales

Abstract

After a brief review of some classical approaches in image segmentation, the basics of multifractal theory and its application to image analysis are presented. Practical methods for multifractal spectrum estimation are discussed and some experimental results are given.

1 Introduction

Image Analysis deals with teaching a computer how to "see". It has applications in many fields, including robotics, medical imaging (e.g. automatic detection of diseases from images), satellite imaging (e.g. monitoring vegetation evolution), and a lot more. In most cases, the aim is either to let the computer process data which are too numerous to be analyzed by human beings, or to be able to detect features in images that are not easily found with the eye.

A first step towards these goals is to detect and localize objects and features in images. A considerable body of research has been devoted to this

problem in the past twenty years. It has raised difficult problems, some of which have lead to elegant mathematical developments, even though several issues are still unsolved.

In this paper, we describe the multifractal approach to image analysis. The essential difference between this approach and "classical" ones lies in the way they handle irregularities. In order to clearly illustrate this and other important differences, we briefly recall in section 2 some of these classical methods. Section 3 presents the principles of the multifractal theory, whose application to image analysis is detailed in section 4. Practical issues and experiments are described at the end of the paper.

2 A quick review on some "classical" methods

2.1 Canny's edge detector

Canny's detector is based on a gradient approach [7]. The image is modeled by a function I from $\mathbb{R}^2$ to $\mathbb{R}^+$. Edges are assumed to correspond to sharp variations of the grey levels, i.e. of I. These sharp variations may be detected by looking at the maxima of the gradient of I. Two problems arise:

- Because the data are discrete, it is not straightforward to compute a gradient.
- In many cases, noise is present in the image, which induces spurious edges.

The basic idea is to filter the image in order to get rid of these problems. More precisely, a linear filter f is defined and edges are detected as follows:

- Filter the image I with f: $I * f$.
- Compute the gradient of the filtered image: $(I * f)' = I * f'$.
- Compute the norm of the gradient $|(I * f)'|$.
- Select the points where the norm of the gradient is maximum in the gradient's direction.

The problem of edge detection is thus reduced to the one of finding a linear filter f. f should satisfy the following conditions:

- Detection: the edge detector should react to edges.

- Localization: edges should be precisely localized.
- Unicity of response: an edge should trigger the detector only once.

Different types of edges may be considered:

- Step edge or $0^{-\text{th}}$ order discontinuity (figure 1).
- Roof edge or $1^{-\text{st}}$ order discontinuity (figure 2) .

FIG. 1 - *Model of a step edge.*

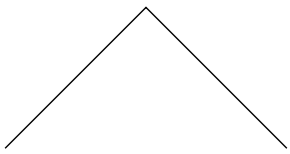

FIG. 2 - *Model of a roof edge.*

Other types of edges include corners, lines, ... Some assumptions are then made about the nature of the noise in the image (additive/multiplicative, Gaussian/Rayleigh/K-distributed, ...), and, for some simple cases, it is possible to derive an explicit optimal filter f w.r.t. the three criteria above. For instance, for an ideal step edge with additive Gaussian noise, the optimal filter is:

$$\begin{aligned} f(x) &= a_1 e^{\alpha x}\sin wx + a_2 e^{\alpha x}\cos wx \\ &\quad + a_3 e^{-\alpha x}\sin wx + a_4 e^{-\alpha x}\cos wx \qquad \text{(Canny)} \end{aligned}$$

Usual filters in this approach are:

$$\begin{aligned} f(x) &= c e^{-\alpha|x|} && \text{(Shen–Castan)} \\ f(x) &= (1 + c|x|)e^{-\alpha|x|} && \text{(Deriche)} \end{aligned}$$

where α is a parameter that tunes the smoothing of the signal. It allows to adjust the trade-off between localization and robustness.

2.2 Mathematical Morphology

Mathematical Morphology was first introduced by G. Matheron in the 60's and then developed by a number of authors ([35], [34])

The basic idea underlying Mathematical Morphology is that the notion of a geometrical structure or texture is not purely objective. It does not lie in the phenomenon itself, nor in the observer, but somewhere between the two. Mathematical Morphology quantifies this intuition by introducing the concept of structuring elements. The structuring elements are chosen by the analyst, and they interact with the studied object to modify its shape and transform it into a simpler one. The result thus reflects information both on the structuring elements and on the object. An important issue is then the choice of the structural elements. This choice is based on several principles which can be loosely stated as:

- "We only see what we want to look at."
- "To perceive an image, is to transform it."

More precisely, given a set X, a morphological operation on X is by definition the composition of a transformation Ψ of X into a new set $\Psi(X)$ followed by a measure μ of $\Psi(X)$. Thus, if we note $E = \mathbb{R}^n$ and $\mathcal{P}(E)$ a family of subsets of E satisfying certain technical conditions, we have:

$$\begin{array}{ccccc} \mathcal{P}(E) & \longrightarrow & \mathcal{P}(E) & \longrightarrow & \mathbb{R}^+ \\ X & \longmapsto & \Psi(X) & \longrightarrow & \mu(\Psi(X)) \end{array}$$

A morphological operation must satisfy four constraints:

1. Compatibility under translation:

Let 0 be an "origin" in E and $X_h = \{x + h, x \in X\}$. Then we must have:

$$\Psi^0(X_h) = [\Psi^{-h}(X)]_h$$

where Ψ^0 is the transformation applied when 0 is the origin.

This means that it is equivalent to apply Ψ^0 to a shifted version of X or to apply Ψ^{-h} to X. In other words, the result of the transformation should not depend on the origin.

2. Compatibility under change of scale:

Let $X_\lambda = \{\lambda x, x \in X\}$. Then, we must have

$$\Psi_\lambda(X) = \lambda\Psi\left(X_{\frac{1}{\lambda}}\right)$$

where Ψ_λ is the transformation applied when the coordinate axes are scaled by λ. In other words, the transformation must be independent of an arbitrary magnification/reduction of the object.

3. Local knowledge:

$\forall Z'$ bounded, $\exists Z$ bounded such that:

$$(\Psi(X \cap Z)) \cap Z' = \Psi(X) \cap Z'$$

In other words, for any bounded set Z' in which we want to know $\Psi(X)$, we can find a bounded set Z in which the knowledge of X is sufficient to locally perform the transformation. This requirement stems from the fact that we generally do not see the whole object X but only a part of it, $X \cap Z$.

4. Semi Continuity:

This principle insures that, if we have a decreasing sequence of closed sets $(X_n)_{n\geq 1}$ converging to $X \neq \emptyset$, then $(\Psi(X_n))_{n\geq 1}$ should converge to $\Psi(X)$, as soon as Ψ is increasing (i.e. $A \subset B \Longrightarrow \Psi(A) \subset \Psi(B)$).

To be more precise, we have to define a fundamental tool of Mathematical Morphology, the "Hit or Miss" topology. Let $\mathcal{F}$ be the space of all closed subsets of $\mathbb{R}^n$. Given two finite sequences $(G_1, \ldots, G_m)$ of open sets and $(K_1, \ldots, K_p)$ of compact sets of $\mathbb{R}^n$, the class of all the closed sets that hit every G_i and miss every K_j defines an open neighborhood in $\mathcal{F}$. The induced topology is called the Hit or Miss topology (see figure 3).

The following result is fundamental.

Theorem: *The space $\mathcal{F}$ equipped with the Hit or Miss topology is compact and has a countable base.*

The fourth principle now reads: Ψ satisfies the fourth principle iff it is semi continuous w.r.t. the Hit or Miss topology.

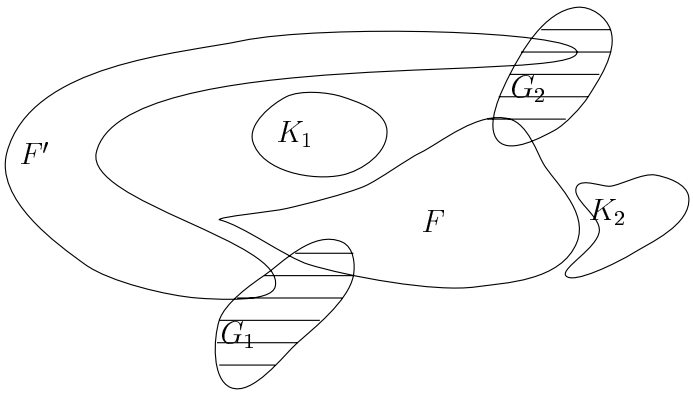

FIG. 3 - *F and F' belongs to the same neighborhood defined by K_1, K_2 and G_1, G_2.*

This last principle implies that morphological operations are ideally performed in the continuous space, and not in the discrete space of the pixels.

Basic morphological transformations include:

- Erosion

  Let B be a structuring elements. The eroded set of X is defined as:

  $$Y = \{x / B_x \subset X\} = \cap_{y \in B_0} X_{-y}$$

  Let:

  $$\check{B} = \cup_{y \in B} \{-y\} \text{ and } X \ominus B = \cap_{b \in B} X_b$$

  Then: $Y = X \ominus \check{B}$ ($X \ominus B$ is the classical Minkowski substraction).

  Thus Y is the set of points where we can center the structuring element with the constraint that it is included in X.

- Dilation

  It is the dual operation of erosion and is defined through Minkowski addition:

  $$X \oplus \check{B} = \{x / B_x \cap X \neq \emptyset\} = \cup_{y \in \check{B}} X_y$$

  The dilated set is thus the set of points where we can center B with the constraint that it intersects (or hits) X.

Dilation is associative and commutative, and $\mathcal{P}(\mathbb{R}^2)$ with the dilation is an Abelian semi-group. The origin is the unit element of the semi-group, and dilation distribute the union. Erosion and dilation are increasing transformations.

- Opening and closing

The opening of X w.r.t. B is:

$$X_B = (X \ominus \check{B}) \oplus B$$

An opening thus consists in eroding X and then performing a dilation on the result. X_B will in general be different from X since it will be "simpler" (it will have less details). Similarly the closing of X w.r.t. B is:

$$X^B = (X \oplus \check{B}) \ominus B$$

i.e. a dilation followed by an erosion.

Intuitively, an opening smoothes the contours and suppresses small "islands". Conversely, a closing blocks up small "lakes".

The main properties of opening and closing are:

- $(X^c)_B = (X^B)^c$ and $(X_B)^c = (X_c)^B$

  (the opening of the complement of X is the complement of the closing of X, and conversely: opening and closing are dual w.r.t. the complementation.)

- The morphological opening transform satisfies the three properties defining an opening in algebra. It is:
  - anti-extensive: $X_B \subset X$,
  - increasing: $X \subset X' \Longrightarrow X_B \subset X'_B$,
  - idempotent: $(X_B)_B = X_B$.

- The closing transform verifies:

$$X^B \supset X$$

$$X \subset X' \Longrightarrow X^B \subset X'^B$$

$$(X^B)^B = X^B$$

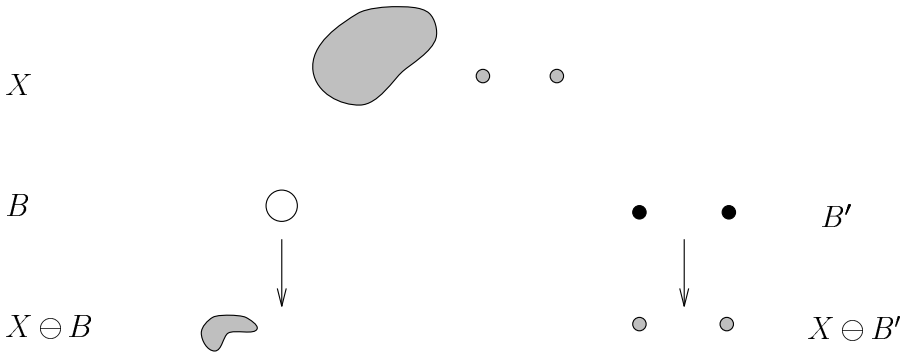

FIG. 4 - *Erosion with two different structuring elements.*

To conclude, let us say a few words about the choice of the structuring element. Different structuring elements lead to different information on the studied set: this is illustrated on figure 4, where B has given some information relative to the size of X, as B' has given some information about the spatial distribution of X.

As noted above, it thus makes no sense to speak of the characteristic features of X, because we "only see what we want to look at": the information we get depends on the structuring element used. Some basic rules to choose the structuring elements are:

- simplicity: B should have in any case a much simpler geometric structure than X,
- B should be bounded,
- B should have "extreme" properties: if an isotropic filtering is needed, B should consist in a union of circular annuli with same center. If we want to perform an anisotropic filtering, B should be a union of aligned segments rather than elongated ellipses.
- Convex structures should be used to investigate size distributions, and clusters of isolated points should be used to measure set covariances.

2.3 Image Multiscale Analysis

This approach formalizes some basic assumptions that are usually made in image analysis and shows that, under these assumptions, the analysis must

obey a simple equation. The merits of this approach are:

- to clearly state which assumptions are made,
- to discuss the validity of these assumptions and their relative importance,
- to derive rigorously the consequences of these assumptions and to obtain the unexpected result that in general only one equation is compatible with them.

This approach has been extensively studied in recent years. Useful references are [24], [3], [32]. Here, an image is a "brightness" function $u_0(x)$, for x in a subset of $\mathbb{R}^2$. u_0 is assumed to be continuous and to verify:

$$\exists N, C / \forall x, \|(1+|x|)^{-N} u_0(x)\| \leq C.$$

The set of all such functions will be denoted by $\mathcal{F}$. u_0 has no "absolute" meaning, but rather is an element of an equivalence class. More precisely, an image is an equivalence class of functions $g(u_0(Ax))$ where:

- g is a continuous non decreasing function: the perceived image does not depend on the sensibility of the sensor nor on any increasing modification of the contrast and/or the brightness.
- A is an affine map: the perceived image does not depend on any affine transformation of the plane.

An analysis of the image $u_0(x)$ is a "multiscale" image $u(t,x)$ where t is a "time" parameter measuring how much the original image has been "smoothed". Here, smoothing means that we filter out small "details" and that we take into account more and more global information. Thus t also measures the size of the neighborhood that influences the value of $u(t,.)$ at x. The parameter t is related to the "scale" of analysis.

The series of operation leading from $u_0(x) = u(0,x)$ to $u(\infty,x)$ is sometimes referred to as the Visual Pyramid. The basic principles that a visual pyramid must obey are the following ones:

- Causality:

This principle states that what happens at scale t cannot influence what happens at scale $t' < t$. Furthermore, the output at scale t can be computed from the output at scale $t-h$, $h > 0$.

Let:

$$\begin{array}{lcccc} T_t & : & \mathcal{F} & \rightarrow & \mathcal{F} \\ & & u_0 & \longmapsto & u(t,.) \end{array}$$

Then

$$\begin{array}{lcccc} \exists\ T_{t+h,t} & : & \mathcal{F} & \longrightarrow & \mathcal{F} \\ & & u(t,.) & \longmapsto & u(t+h,.) \end{array}$$

such that:

$$T_{t+h} = T_{t+h,t} T_t, \qquad \text{with} \qquad T_0 = Id$$

- Local Comparison Principle

  This principle states that $T_{t+h,t}$ acts "locally", i.e. that $T_{t+h,t}u(x)$ should depend upon the values of u only in a neighborhood of x. Let u and v be two images in $\mathcal{F}$. If for some $\varepsilon > 0$, $u(y) > v(y)$ for $y \in B(x,\varepsilon), y \neq x$, then for all h small enough,

  $$(T_{t+h,t}u)(x) \geq (T_{t+h,t}v)(x)$$

  This implies in particular that no new boundaries are created when the scale increases.

- Regularity:

  This principle states that a smooth image should evolve in a smooth way: let u be a quadratic form of $\mathbb{R}^2$. Then there exists a function $F(u,x,t)$ continuous w.r.t. u such that:

  $$\frac{(T_{t+h,t}u - u)(x)}{h} \longrightarrow F(u,x,t) \text{ when } h \longrightarrow 0$$

The last principles deal with the fact that an image is really an equivalence class. The operators should thus commute with some perturbations of the data:

- Morphological Invariance:

  Let g be a continuous non decreasing function.

  Then:

  $$gT_{t+h,t} = T_{t+h,t}g$$

- Euclidean Invariance:

  Let A be an isometry. Then:

  $$AT_{t+h,t} = T_{t+h,t}A$$

- Affine Invariance:

  Here we only impose a weak commutation property, because A may reduce or enlarge the image. For instance, if A is a zoom by a factor $\lambda > 0$:

  $Au(x) = u(\lambda x)$

  The affine invariance principle is stated as follows: there exists a C^1 function $t'(t, A) \geq 0$ such that:

  $$AT_{t'(t,A),t'(s,A)} = T_{t,s}A$$

  Moreover, $\dfrac{\partial t'}{\partial \lambda}(t, \lambda Id)$ is positive for $t > 0$.

  This means that the result of the multiscale analysis T_t is independent of the size and the position in space of the analyzed features.

Although the most general invariance should be projective, the affine approximation is valid if the studied objects are sufficiently small or sufficiently far away from the sensor. The condition on $\dfrac{\partial t'}{\partial \lambda}$ means that the analysis scale increases with the size of the picture. It may be shown that, if $t \longrightarrow T_t$ is a one-to-one family of operators satisfying the affine invariance principle and the causality principle, then:

- $t'(t, B)$ only depends on t and $|detB|$
- t' is increasing w.r.t. t
- there exists an increasing differentiable function

  $$\sigma : [0, \infty] \longmapsto [0, +\infty]$$

  such that:

  $$t'(t, B) = \sigma^{-1}(\sigma(t)|detB|^{1/2})$$

σ acts as a rescaling function: if $S_t = T_{\sigma^{-1}(t)}$ then $t'(t, B) = t|detB|^{1/2}$ for the rescaled analysis S_t.

The following results provide a complete solution to the problem of image analysis under the principles and assumptions presented above:

- If a multiscale analysis is causal, local, Euclidean invariant and linear, then $u(t, x) = (T_t u)(x)$ obeys the heat equation:

$$\frac{\partial u}{\partial t} = \Delta u$$

- If a multiscale analysis is causal, local and regular, then $u(t, x)$ is a viscosity solution of:

$$\frac{\partial u}{\partial t} = F(D^2 u, Du, u, x, t)$$

  Where Du is the gradient of u, F is defined in the regularity principle and is non decreasing w.r.t. to $D^2 u$. If u_0 is a bounded uniformly continuous image, this equation has a unique viscosity solution.

- If a multiscale image analysis is causal, local, regular, morphological and Euclidean invariant, then:

$$\frac{\partial u}{\partial t} = |Du| F\left(div\left(\frac{Du}{|Du|}\right), t\right)$$

  where F is non decreasing w.r.t. its first argument. The quantity $div\left(\frac{Du}{|Du|}\right)(x)$ can be interpreted as the curvature of the level line of the image $u(t, x)$ passing by x. In the particular case where if $F = +1$ is constant, we get:

$$\frac{\partial u}{\partial t} = |Du|$$

  This corresponds to a morphological dilation with a ball $B(x, t)$ as a structuring element. Likewise, if $F = -1$, we obtain the morphological erosion.

The "most invariant" model is the following.

Theorem: *There is a single causal, local, regular morphological and affine invariant multiscale analysis, whose equation is:*

$$\frac{\partial u}{\partial t} = |Du| \left(t.div \left(\frac{Du}{|Du|} \right) \right)^{1/3}$$

3 The multifractal approach

3.1 Overview

In this approach, the image is not modeled by a function but by a measure μ. This modeling allows to emphasize the fundamental role played by resolution. In this framework, the basic assumptions are:

- The relevant information for the analysis can be extracted from the Hölder regularity of μ.
- Three levels contribute to the perception of the image:
  - the pointwise Hölder regularity of μ at each point,
  - the variation of the Hölder regularity of μ in local neighborhoods.
  - the global distribution of the regularity in the whole scene.
- The analysis should be translation and scale invariant.

Compared to the approaches described above, the multifractal one has the following specific features:

- As in mathematical morphology (MM) and image multiscale analysis (IMA), translation and scale invariance principles are fulfilled.
- There is no "local comparison principle" or "local knowledge principle". On the contrary, information about whole parts of the image is considered essential to analyze each point (see section 4.1).
- The most important difference between classical and multifractal methods lies in the way they deal with regularity. While the former aim at obtaining smoother versions of the image (possibly at different scales) in order to get rid of irregularities, the latter tries to extract information

directly from the singularities. Edges, for instance, are not considered as points where large variations of the signal still exist after smoothing, but as points whose regularity is different from the "background's" regularity in the raw data. Such an approach makes sense for "complex" images, in which the relevant structures are irregular in nature. Indeed, an implicit assumption of MM and IMA is that the useful information lies at boundaries between originally smooth regions, so that it is natural to filter the image. But there are cases (e.g. in medical imaging, satellite or radar imaging) where the meaningful features are essentially singular.

- As in MM, and contrarily to IMA, the multifractal approach does not assume that there is a universal scheme for image analysis. Rather, depending on what one is looking for, different measures may be used.
- Both MM and IMA consider the relative values of the grey levels as the basic information ("morphological invariance"). Here the Hölder regularity is considered instead. This again is justified in situations where the important information lies in the singularity structure of the image. As an example, figure 5 displays the lumping of two 2D Weierstrass functions with different exponents. On this image, the use of classical methods with standard parameters would result in the detection of several edges. The multifractal approach, on the other hand, will detect only one vertical line in the middle. This case of a fractal texture is somewhat extreme. Experiments on real images are shown at the end of the paper.

The natural theoretical framework that allows to perform an analysis based on the above ideas is an extension of multifractal analysis that deals with multifractal correlations and sequences of Choquet capacities. A complete presentation of these extensions goes beyond the scope of this paper, and only the basics principles are presented below.

Before concluding this paragraph, let us note that the multifractal approach does not assume that the image is a "fractal" object or a "multifractal" measure, whatever meaning is given to these words. Instead, a multifractal analysis of the image is performed, without making any assumptions on its structure or regularity.

3.2 Choquet capacities

Let E be a set, $\mathcal{P}(E)$ the power set of E. A paving on E is a set $\mathcal{E}$ of subsets of E containing the empty set and stable under finite union and finite

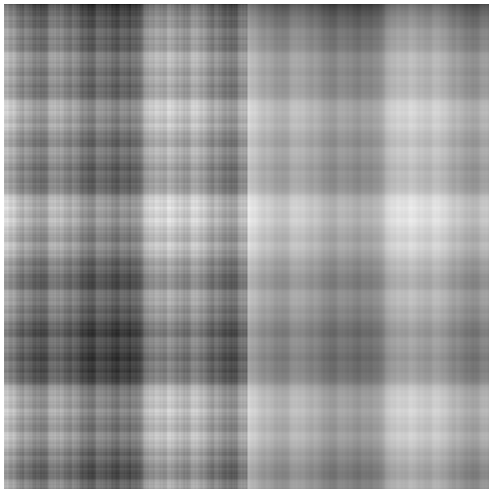

FIG. 5 - *Lumping of two 2D Weierstrass functions.*

intersection. The pair $(E, \mathcal{E})$ is called a paved space. A Choquet $\mathcal{E}$-capacity c is a function:

$$\mathcal{P}(E) \to \overline{\mathbb{R}}$$

such that:

- c is non decreasing,
- if (A_n) is an increasing sequence of subsets of E, then:

$$c\left(\cup_n A_n\right) = \sup_n c(A_n),$$

- if (A_n) is a decreasing sequence of elements of $\mathcal{E}$, then:

$$c\left(\cap_n A_n\right) = \inf_n c(A_n)$$

In image analysis, it is useful to consider sequences of Choquet capacities. Let us give a few examples. Let $E := [0, 1)$, $\mathcal{P} := ((I_k^n)_{0 \le k < \nu_n})_{n \ge 1}$ a sequence

of partitions of $[0,1)$, $x_k^n := \inf(I_k^n)$ and for all n, $\zeta_n : \mathcal{P}(E) \longrightarrow [0,1]$. The following set functions defined on $\mathcal{P}(E)$ are Choquet capacities:

$$c_n(A) := \max\{\zeta_n(I_k^n); x_k^n \in A\}$$

$$\forall p \geq 1 \quad c_n^{(p)}(A) := \left(\sum_{x_k^n \in A} \zeta_n(I_k^n)^p \right)^{1/p}$$

$$c_n(A) := \max_t \ \# \ \{k \ / \ \zeta_n(I_k^n) = t \, ; x_k^n \in A\}$$

Remark: if $A \subseteq I_k^n$, then

$$c_n(A) = 0 \quad \text{or} \quad c_n(A) = c_n(I_k^n) = \zeta_n(I_k^n)$$

Such capacities are called myopic capacities.

3.3 Multifractal analysis of sequences of Choquet capacities

Multifractal analysis has recently drawn much attention as a tool for studying the structure of singular measures, both in theory and in applications ([19, 29, 16]). Efforts have mainly been focused on special cases such as self-similar and self-affine measures, both in the deterministic [33, 31, 6] and in the random case [28, 23, 20, 14, 4, 30]. Multifractal analysis has been extended to deal with Choquet capacities in [27]. Studies have also been conducted to extend the analysis to point functions [21, 22].

In the multifractal scheme, the pointwise structure of a singular measure is analyzed through the so-called "multifractal spectrum", which gives either geometrical or probabilistic information about the distribution of points having the same singularity. Several definitions of a multifractal spectrum exist. Some are related to measure theory (Hausdorff and packing measures), other to large deviation theory (Rényi exponents and Legendre spectrum). The so-called "multifractal formalism" assesses that, in some situations, all the spectra coincide. The motivation for studying the multifractal formalism stems from the fact that both Hausdorff and packing dimensions are in general difficult to compute. The other definitions are more suited to applications, as they are easier to evaluate. The multifractal formalism is known to hold for the class of multiplicative measures, but it fails in general.

Let $c := (c_n)$ be a sequence of Choquet capacities defined on $[0,1)$, with values in $[0,1]$. Let μ be a non-atomic probability measure, called the "reference measure". Let $\mathcal{P} := ((I_j^n)_{0\leq j<\nu_n})_{n\geq 1}$ be a sequence of partition of $[0,1)$ such that:

(C1) $\lim_{n\to\infty} \max_{0\leq j<\nu_n} |I_j^n| = 0$,

(C2) $\forall n, k, \quad I_k^n$ is an interval, semi-open to the right.

Define:

$$\begin{aligned}
\mathcal{H}^s_{\mu,\delta}(E) &:= \inf\{\sum_{i=1}^{+\infty} \mu(E_i)^s \ / E \subset \bigcup_i E_i, \quad \mu(E_i) \leq \delta, E_i \in \mathcal{P} \quad \forall i\} \\
\mathcal{H}^s_{\mu}(E) &:= \lim_{\delta\to 0} \mathcal{H}^s_{\mu,\delta}(E) \\
\dim_\mu(E) &:= \inf\{s \ / \ \mathcal{H}^s_\mu(E) = 0\} = \sup\{s \ / \ \mathcal{H}^s_\mu(E) = +\infty\}
\end{aligned}$$

Under additional conditions on $\mathcal{P}$, $\dim_\mu(E)$ coincides with the Hausdorff dimension of E w.r.t. to μ.

Let $I^n(x)$ denote the interval I_j^n which contains x. The Hölder coarse grain exponent is defined as:

$$\alpha_n(x) := \frac{\log c_n(I^n(x))}{\log \mu(I^n(x))}$$

and the Hölder pointwise exponent as:

$$\alpha(x) = \lim_{n\to\infty} \alpha_n(x)$$

when the limit exists. Let

$$E_\alpha := \{x \in [0,1) \ / \ \alpha(x) = \alpha\}$$

The Hausdorff singularity spectrum of c is defined as:

$$f_h(\alpha) := \dim_\mu E_\alpha$$

Let $n \in \mathbb{N}$ and $\varepsilon > 0$. Note:

$$K^n_\varepsilon(\alpha) := \left\{k \in \{0,\ldots,\nu_n - 1\} \ / \ \alpha_k^n := \frac{\log c_n(I_k^n)}{\log \mu(I_k^n)} \in (\alpha-\varepsilon, \alpha+\varepsilon)\right\}$$

and

$$N_\varepsilon^n(\alpha) := \#K_\varepsilon^n(\alpha)$$

Let, for $\beta > 0$,

$$\begin{aligned} S_\varepsilon^n(\alpha,\beta) &:= \sum_{k\in K_\varepsilon^n(\alpha)} \mu(I_k^n)^\beta \\ S_\varepsilon(\alpha,\beta) &:= \limsup_{n\to+\infty} S_\varepsilon^n(\alpha,\beta) \end{aligned}$$

There exists a real $f_g^\varepsilon(\alpha)$ such that:

$$\begin{aligned} \beta < f_g^\varepsilon(\alpha) &\implies S_\varepsilon(\alpha,\beta) = +\infty \\ \beta > f_g^\varepsilon(\alpha) &\implies S_\varepsilon(\alpha,\beta) = 0 \end{aligned}$$

The large deviation spectrum is:

$$f_g(\alpha) := \lim_{\varepsilon\to 0} f_g^\varepsilon(\alpha)$$

When all intervals have the same size and μ is the Lebesgue measure $\mathcal{L}$,

$$f_g(\alpha) = \lim_{\varepsilon\to 0} \limsup_{n\to+\infty} \frac{\log N_\varepsilon^n(\alpha)}{\log \nu_n}$$

Finally, let:

$$X_n(x,y) := \sum_{j<\nu_n}{}' c_n(I_j^n)^x \mu(I_j^n)^{-y}$$

and

$$X(x,y) := \limsup_{n\to\infty} n^{-1} \log X_n(x,y)$$

where $\sum'$ means that the summation runs through the indices j such that $c_n(I_j^n)\mu(I_j^n) \neq 0$. X is convex function, non decreasing in y and non increasing in x. Let: $\Omega := \{(x,y)/X(x,y) < 0\}$. There exists a concave function τ such that:

$$\mathring{\Omega} = \{(x,y) \in \mathbb{R}^2 \ / \ y < \tau(x-0)\}$$

The Legendre multifractal spectrum is defined as τ^*, the Legendre transform of τ:

$$f_l(\alpha) := \tau^*(\alpha) := \inf_q [q\alpha - \tau(q)]$$

When all intervals have the same size and $\mu = \mathcal{L}$,

$$\tau(q) = \limsup_{n \to +\infty} \frac{\log \sum_k{}' \mu(I_k^n)^q}{\log \nu_n^{-1}}$$

As mentioned above, a central concern of the multifractal theory is to relate f_l, f_g and f_h. This is the so called multifractal formalism. It has important applications in our case of image analysis:

- f_g measures, loosely speaking, the probability of finding a given value of α_n at resolution n. It gives a probabilistic description of the distribution of the singularities.
- f_h measures the Hausdorff dimension of the set of points having a given α. It gives a geometrical description of the distribution of the singularities.

However, both f_g and f_h are difficult to compute on real discrete data. On the other hand, f_l is much easier to compute, because it involves only the evaluation of average quantities. As a counterpart, f_l generally contains less information, because it is always a concave function. The following results hold.

Theorem: *Under conditions (C1) and (C2),*

$$f_h \leq f_g \leq f_l$$

Theorem: *Under conditions (C1) and (C2),*

$$f_l = f_g^{**}$$

3.4 An example: the binomial measure

Define a sequence of measures μ_n on $[0,1)$ as follows: let $m_0 \in (0,1), m_1 = 1 - m_0$. For $n \in \mathbb{N}^*$ and for $k \in \{0, \ldots 2^{n-1}\}$ let:

$$\mu_n[k\ 2^{-n}, (k+1)2^{-n}) = m_0^{n\phi_0^n} m_1^{n\phi_1^n}$$

where ϕ_0^n is the proportion of 0's in the base-2 expansion of $x \in [k\ 2^{-n}, (k+1)2^{-n})$ up to rank n, and ϕ_1^n the proportion of 1's:

$$\begin{array}{lcllll} x & = & \sum_{i=1}^{\infty} x_i 2^{-i} & x_i & \in & \{0,1\} \\ \phi_0^n(x) & = & \frac{\sum_{i=1}^{n}(1-x_i)}{n} & \phi_1^n(x) & = & \frac{\sum_{i=1}^{n} x_i}{n} \end{array}$$

The sequence $(\mu_n)_{n\geq 1}$ has a weak* limit μ, which is called a binomial measure. It is easy to see that, if $\phi_0(x) = \lim_{n\to\infty} \phi_0^n(x)$ exists, then $\alpha(x) = -\phi_0 \log_2 m_0 - \phi_1 \log_2 m_1$.

As for the multifractal spectra, it can be proved that:

$$\left\{ \begin{array}{lcl} \alpha(\phi_0) & = & -\phi_0 \log_2 m_0 - \phi_1 \log_2 m_1 \\ f_h(\phi_0) & = & f_g(\phi_0) = f_l(\phi_0) = -\phi_0 \log \phi_0 - \phi_1 \log \phi_1 \end{array} \right.$$

and the Hausdorff dimension of the set of points "without" a ϕ_0 is 1. Figure 6 displays a 2D version of the above construction (a "quadrinomial measure"). Figures 5 and 6 illustrate a striking difference between classical and multifractal methods for image segmentation: as said above, the latter will detect only one edge in figure 5 but several ones in image 6. The former will in general detect several edges in both images. Edges detected by IMA and multifractal analysis on figure 6 do however coincide with an appropriate choice of the parameters.

A typical spectrum of a binomial measure is shown on figure 7. Figure 8 displays a 1D random version of a multinomial measure.

The above are simple cases where the multifractal formalism holds. Let us now briefly describe three examples where it fails.

1. Consider the following measure on $[0,1)$:

$$\nu(A) = \frac{1}{2}(\nu_1(A) + \nu_2(A))$$

where:

- A is a Borel set in $[0,1)$.
- ν_1 is an m_0-binomial measure on $\left[0, \frac{1}{2}\right)$.

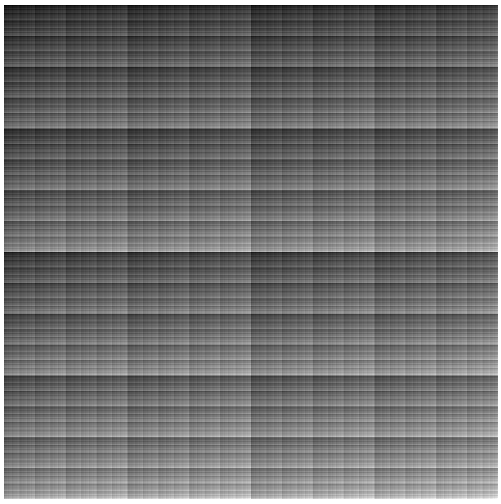

FIG. 6 - *A 2D quadrinomial measure.*

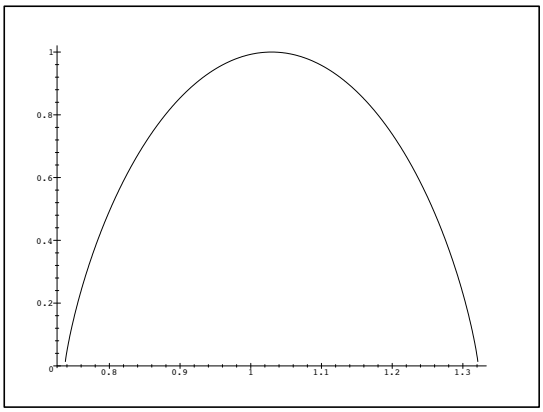

FIG. 7 - *Spectrum of a binomial measure.*

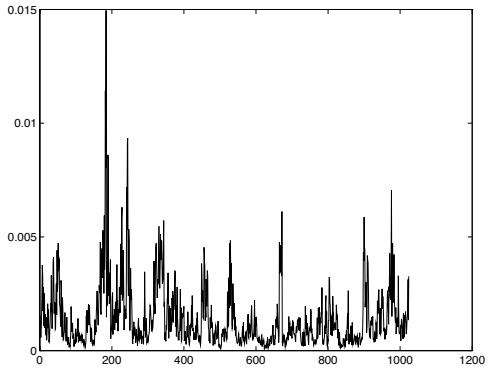

FIG. 8 - *A random multinomial measure.*

- ν_2 is an m_0'-binomial measure on $\left[\frac{1}{2}, 1\right)$ and $m_0 \neq m_0'$.

Then it is easy to show that (figure 9):

$$\begin{array}{lcl} f_h^\nu = f_g^\nu & = & \sup(f_h^{\nu_1}, f_h^{\nu_2}) \\ f_l^\nu & = & (f_g^\nu)^{**} \end{array}$$

- Let now:

$$\nu(A) = \frac{1}{2}[\nu^1(A) + \nu^2(A)]$$

where:

- ν^1 is an m_0-binomial measure on $[0, 1)$,
- ν^2 is an m_0'-binomial measure on $[0, 1)$ and $m_0 \neq m_0'$.

Let $[\alpha_{min}^i, \alpha_{max}^i]$ be the support of $f_h^{\nu^i}$ (i.e. $f_h^{\nu^i} = 0$ outside $[\alpha_{min}^i, \alpha_{max}^i]$). Assume w.l.o.g. that $\alpha_{min}^1 < \alpha_{min}^2 < \alpha_{max}^2 < \alpha_{max}^1$. Then (figure 10):

$$\begin{array}{lcll} f_h^\nu = f_g^\nu & = & f_h^{\nu^1} & on \quad [\alpha_{min}^1, \alpha_{min}^2] \\ f_h^\nu = f_g^\nu & = & f_h^{\nu^2} & on \quad [\alpha_{min}^2, \alpha_{max}^2] \\ f_h^\nu = f_g^\nu & = & -\infty & otherwise \\ f_l^\nu & = & (f_g^\nu)^{**} & \end{array}$$

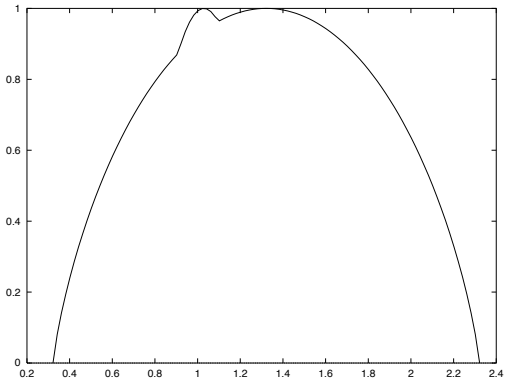

FIG. 9 - *Spectrum of a lumping of two binomial measures.*

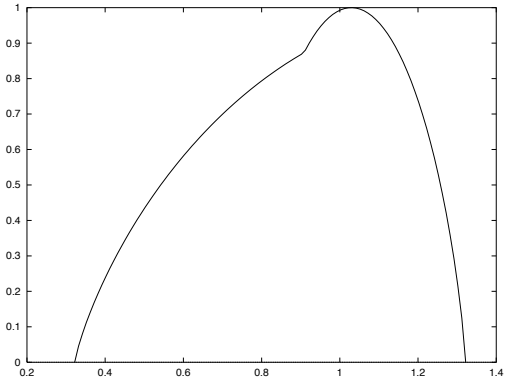

FIG. 10 - *Spectrum of a sum of two binomial measures.*

- The last example deals with a case where $f_h(\alpha) < f_g(\alpha) = f_l(\alpha)$ on an interval. Let

$$c_n(A) = \max\{2^{-nk2^{-n}}\ ;\ k\,2^{-n} \in A\}$$

With a dyadic partition of $[0,1)$, we get:

$$f_h(\alpha) = \begin{cases} 0 & \text{if } \alpha \in [0,1] \\ -\infty & \text{otherwise} \end{cases}$$

and

$$f_g(\alpha) = f_l(\alpha) = \begin{cases} 1 & \text{if } \alpha \in [0,1] \\ -\infty & \text{otherwise} \end{cases}$$

Let us finally mention that, while f_l is always a concave function, it is possible to construct a sequence of capacities having a prescribed f_h spectrum under very weak assumptions on f_h. Figure 11 displays an example.

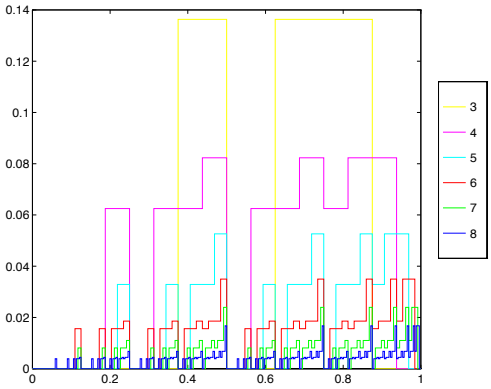

FIG. 11 - *The first elements of a sequence of capacities whose spectrum is x^2 on an interval.*

We end this section with a few words on the role of the reference measure. Let:

- ν_1 be a trinomial measure (0.1, 0.8, 0.1) on $\left[0, \frac{1}{3}\right]$

- ν_2 be a binomial measure $(p, 1-p)$ on a Cantor triadic set defined on $\left[\frac{2}{3}, 1\right]$
- $\nu = (\nu_1 + \nu_2)/2$

Let the reference measure, μ, equal $\mathcal{L}$ (the Lebesgue measure). Then it is possible to choose p such that:

$$f_h^{\nu} = \sup(f_h^{\nu_1}, f_h^{\nu_2}) = f_h^{\nu_1}$$

so that the singularities coming from ν_2 are "hidden" by those coming from ν_1. However, if we take $\mu = \nu_1$, the part of the spectrum f_h^{ν} due to ν_1 will reduce to the point (1,1) and the singularities coming from ν_2 will be "seen" on the spectrum.

3.5 Multifractal correlations

It is an obvious fact that the multifractal spectra do not uniquely characterize a measure. To obtain a finer analysis, one can compute "multifractal correlations". These correlations allow to describe the local interplay of singularities. This topic has been investigated for instance in [10, 15, 36].

Let:

$$\begin{aligned} E_{\alpha,\alpha',r} \quad = \quad & \{x \in [0\,;1[\ / \ \lim_{n\to+\infty} \alpha_n(x) = \alpha \ , \\ & \limsup_{n\to\infty} \alpha_n(x + r_n) = \alpha'\} \end{aligned}$$

where $r := (r_n)_n$ is a sequence in $]0, 1[$ such that $\lim_n r_n = 0$. Define:

$$f_h(\alpha, \alpha', r) := \dim_H E_{\alpha,\alpha',r}$$

The generalized definitions of f_g, τ and f_l, although conceptually simple, are a little involved and we don't write them here. The interested reader is referred to [36]. The following results hold:

Theorem: *Under conditions (C1) and (C2),*

$$f_h(\alpha, \alpha', r) \leq f_g(\alpha, \alpha', r) \leq f_l(\alpha, \alpha', r)$$

Theorem: *Under conditions (C1) and (C2),*

$$f_l(\alpha, \alpha', r) = f_g^{**}(\alpha, \alpha', r)$$

Higher order multifractal spectra are usually difficult to compute analytically. We give a result for a deterministic multinomial measure with weights $(m_0, m_1, \ldots, m_{B-1})$. Consider a sequence $(p(n))_n$ such that $\beta := \lim p(n)/n$ exists and let:

$$r_n := B^{-p(n)}$$

In such a case, the multifractal formalism does not hold since, in general,

$$f_h(\alpha, \alpha', \beta) < f_g(\alpha, \alpha', \beta) = f_h(\alpha, \alpha', \beta)$$

More precisely (see figures 12 and 13):

$$f_h(\alpha, \alpha', \beta) = \begin{cases} f_h(\alpha) & \text{if } (\alpha' = \alpha) \text{ or} \\ & (\alpha' > \alpha \text{ and } \alpha = -\log_B m_{B-1}) \\ -\infty & \text{if } (\alpha' < \alpha) \text{ or} \\ & (\alpha' > \alpha \text{ and } \alpha \neq -\log_B m_{B-1}) \end{cases}$$

$$\begin{aligned} \tau(q, q', r) &= -\lim_{n \to +\infty} \frac{1}{n} \log_B \sum_{k=0}^{B^n - 1} \mu(I_k^n)^q \, \mu(I_k^n + B^{-p(n)})^{q'} \\ &= \tau(q + q') - \beta \min(0, L(q, q')) \\ L(q, q') &:= \tau(q + q') + q \log_B m_{B-1} + q' \log_B m_0 \\ \lambda &:= \lambda(\alpha, \alpha') := \frac{\alpha' - \alpha}{\log_B \frac{m_{B-1}}{m_0}} \\ A &:= A(\alpha, \alpha') := \frac{\alpha + \lambda \log_B m_{B-1}}{1 - \lambda} \end{aligned}$$

- If $m_0 \neq m_{B-1}$,

$$f_l(\alpha, \alpha', \beta) = \begin{cases} (1 - \lambda) f_l(A) & \text{if } A \in]\alpha_{\min}, \alpha_{\max}[\\ & \text{and } \lambda \neq 1, \lambda \geq 0 \\ 0 & \text{if } \lambda = 1 \\ & \text{and } \alpha = -\log_B m_{B-1} \\ -\infty & \text{else} \end{cases}$$

- If $m_0 = m_{B-1}$,

$$f_l(\alpha, \alpha', \beta) = \begin{cases} f_l(\alpha) & \text{if } \alpha = \alpha' \\ -\infty & \text{if } \alpha \neq \alpha' \end{cases}$$

Finally:

$$f_g = f_l$$

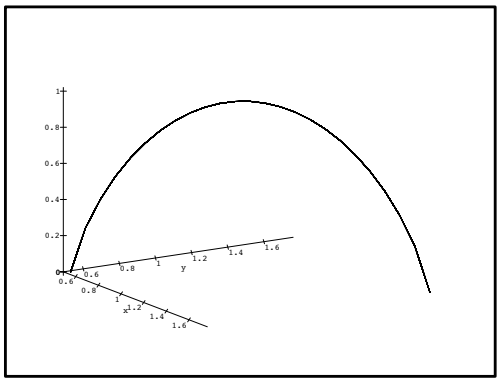

FIG. 12 - f_h *second order spectrum of a multinomial measure.*

4 Application to image analysis

4.1 Introduction

Throughout the remaining of the paper, we make the following assumption:

$$f := f_h = f_g$$

The multifractal analysis of images consists in defining a sequence of capacities on the image, computing its multifractal spectrum, and classifying each point according to the corresponding value of $(\alpha, f(\alpha))$, both in a geometric and a probabilistic fashion. The value of α gives a *local* information

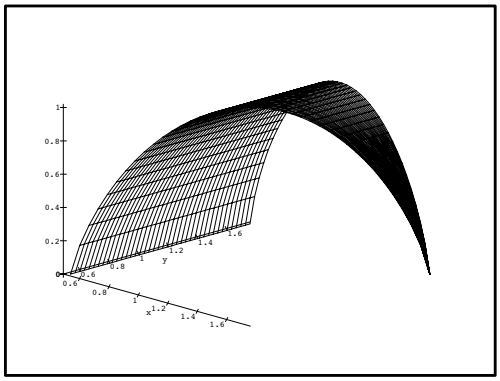

FIG. 13 - *$f_l = f_g$ second order spectrum of a multinomial measure.*

about the pointwise regularity: for a fixed capacity, an ideal step edge point in an image without noise is characterized by a given value. The value of $f(\alpha)$ yields a *global* information: a point on a smooth contour belongs to a set E_α whose dimension is 1, a point contained in a homogeneous region has $f(\alpha) = 2$, etc ... The probabilistic interpretation of $f(\alpha)$ corresponds to the fact that a point in a homogeneous region is a frequent event, an edge-point is a rare event, and, for instance, a corner an even rarer event (see figures 14 and 15). Indeed, if too many "edge points" are detected, it is in general more appropriate to describe these points as belonging to a homogeneous (textured) zone.

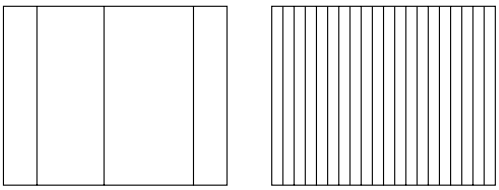

FIG. 14 - *Three edges, a texture.*

In other words, the assumption that $f_g = f_h$ allows to link the geometrical and probabilistic interpretations of the spectrum. Points on a smooth contour

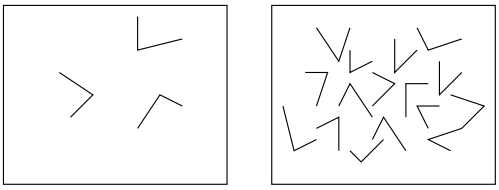

FIG. 15 - *Three corners, a texture.*

have an α such that:

- $f_h(\alpha) = 1$ because a smooth contour fills the space as a line.
- $f_g(\alpha) = 1$ because a smooth contour has a given probability to appear.

In fact, we may define the type of a point (i.e. edge, corner, smooth region ...) through its associated $f(\alpha)$ value:

Definition: *Let c be a sequence of capacities defined on the image. x is called a c-edge point if $f_h^c(\alpha(x)) = 1$. More generally, for $t \in [0, 2]$, x is called a point of type $c - t$ if $f_h^c(\alpha(x)) = t$. The following terminology is used:*

- *points of type $c - 2$ are called smooth points,*
- *points of type $c - 0$ are called rare points.*

A benefit of the multifractal approach is thus that it allows to define not only edge points, but a continuum of various types of points (see figure 27). Note that, in this setting, the type of a point has no absolute meaning. It depends on the sequence of capacities used to analyze the image.

Recall that in most "classical" approaches, an edge point is a point where some sort of "gradient" is maximal in its direction, or more generally, a point where the image exhibits a specific local behavior (even if the size of the neighborhood may evolve, as in IMA). In the multifractal approach, an edge point is defined through both a local ($\alpha(x)$) and a global property ($f(\alpha)$) of the image: one can not decide whether a point is on an edge only by looking at a neighborhood of this point (indeed, all edge points in an image are not characterized by the same value of α: the actual value depends on the contrast of the image and on the noise). Such a use of a global information is meaningful only if one assumes that the image is "homogeneous" in the

following sense: the sequence c of capacities used to analyze the data is such that, for any subregions Ω_1 and Ω_2 in the image,

$$f_h^{\Omega_1}(\alpha) = f_h^{\Omega_2}(\alpha) \qquad \forall \alpha \in support(f_h^{\Omega_1}) \cap support(f_h^{\Omega_2})$$

where $f_h^{\Omega_i}$ is the spectrum of c computed on the restriction of the image to Ω_i. Note that we allow the possibility of finding a given value of α in only one of the two regions (i.e. if $f_h^{\Omega_1}(\alpha) = -\infty$, $f_h^{\Omega_2}(\alpha)$ may assume any value), which is essential for applications. This property can be verified by local computations of the spectrum. When it is not fulfilled, one needs to perform separate analysis for each homogeneous region of the image. Of course, from a practical point of view, this only makes sense if the whole image can be divided in a few sufficiently large homogeneous zones.

Another condition for the above definition to be relevant is that there exists a sequence of capacities that assigns different Hölder exponents to the set of points one wants to extract (the "foreground", for instance the edges) and to the "background". Remark that it is not necessary that all "foreground" points have the same exponent. Infinitely many different exponents may correspond to the same type t, as long as the corresponding subsets of points in the image all have the same dimension (see figures 16 and 17. The *sum* measure and *iso* capacity are defined below, and the procedure to extract the edges is explained in section 4.5).

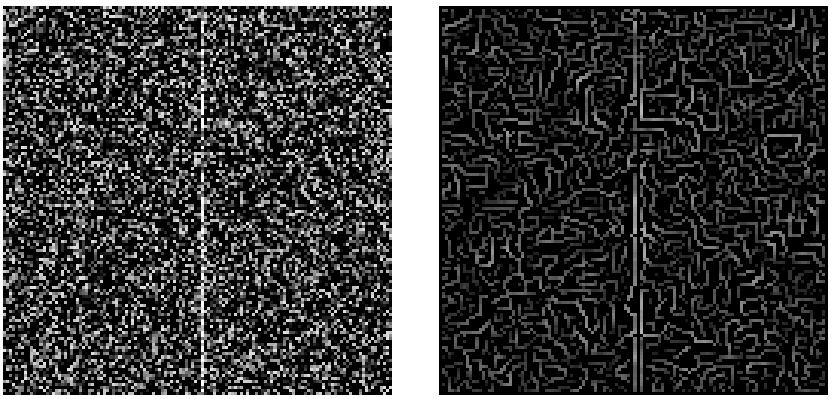

FIG. 16 - *Original image (left) and edges obtained with Canny's detector.*

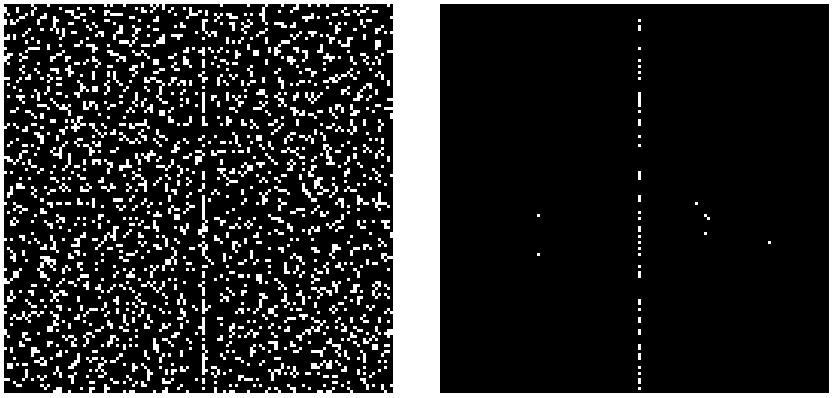

FIG. 17 - *Edges obtained with the sum measure (left) and with the iso capacity(right).*

As is probably clear to the reader by now, the core of the multifractal analysis of images lies in the choice of a relevant sequence of capacities for describing the scene. The following fact is easy to prove:

Proposition: *Let E_n be the set of edge points at time n derived from a multiscale analysis. Then there exist a sequence of capacities $c = (c_n)$ such that, for all n, $E_n = \{(x,y)/\alpha(x,y) = \alpha_0\}$.*

This result is however of little practical interest, and the problem of finding an optimal c in a general setting is still unsolved. Preliminary investigations are described in [9]. In practice, one uses myopic capacities because they allow to take into account the resolution in a simple way. The most basic functions are the following ones:

Definitions: *Assume the image is defined on $[0,1] \times [0,1]$, Let $\mathcal{P} := ((I_k^n)_{0 \leq k < \nu_n})_{1 \leq n \leq N}$ be a sequence of partitions of $[0,1] \times [0,1]$ and (x_k^n, y_k^n) be any point in I_k^n. Each I_k^n is made of an integer number of pixels. Let $L(I_k^n)$ denote the sum of the grey levels in I_k^n. Let (x,y) denote a generic pixel in the image and $L(x,y)$ denote the grey level at (x,y). Let $(p_n)_{1 \leq n \leq N}$ be a fixed sequence of positive integers and Ω be a region in the image.*

sum **measure:**

$$c^s(\Omega) = \sum_{(x,y)\in\Omega} L(x,y)$$

max **capacities:**

$$c_n^m(\Omega) = \max_{I_k^{n+p_n}/(x_k^{n+p_n},y_k^{n+p_n})\in\Omega} L(I_k^{n+p_n})$$

$$c^M(\Omega) = \max_{(x,y)\in\Omega} L(x,y)$$

iso **capacities:**

$$c_n^i(\Omega) = \max_l \#\left\{k \;/\; L(I_k^{n+p_n}) = l, (x_k^{n+p_n}, y_k^{n+p_n}) \in \Omega\right\}$$

$$c^I(\Omega) = \max_l \#\left\{(x,y) \;/\; L(x,y) = l, (x,y) \in \Omega\right\}$$

Note that, since in practice resolution is finite, all sequences above are also finite (they are in fact finite versions of the myopic capacities defined in 3.2). It is easy to show that:

- $c^s(\Omega)$ depends on both the grey level values and their distribution in Ω,
- $c_n^m(\Omega)$ and $c^M(\Omega)$ only depend on the grey level values,
- $c_n^i(\Omega)$ and $c^I(\Omega)$ only depend on the grey level distribution.

Thus, (c_n^m, c^M) and (c_n^i, c^I) give in some loose sense "orthogonal" information about the image. Furthermore, it can be shown that they are more robust to noise than c^s.

4.2 Numerical estimation of f_g

Before we apply the above ideas to specific problems in image analysis, let us say a few words about the numerical estimation of f_g (estimation of f_l is considered in [13] and estimation of f_h in [26]). The simplest method for estimating f_g involve the following steps ([13]):

- for each n, compute all α_k^n;

- compute $\alpha^n_{\min} = \min_k \alpha^n_k, \alpha^n_{\max} = \max_k \alpha^n_k$, and divide $[\alpha^n_{\min}, \alpha^n_{\max}]$ into N boxes
- compute the number N^n_i of intervals I^n_k whose α^n_k falls in the i-th box, $i = 1, \ldots N$;
- find $f_g(\alpha)$ by a linear regression on $(\log n, \log N^n_i)$.

Although, this "histogram method" may yield satisfactory results on strictly multiplicative cascades, it fails to estimate a good approximation in more complex situations, in particular when f_g is not a concave function. For instance, when dealing with a non stationary process, like the lumping of two binomial measures, or a compound process, like the sum of two binomial measures, the histogram method will tend to hide the non concavity of f_g (see figures 9 and 10). There are two types of problems:

- The choice of the number of "boxes" on the α axis: this choice is ad-hoc, although it influences a lot the result. In particular, it does not take into account the ε limit in the definition of f_g.
- The choice of the averaging procedure: when dealing with, say, a trinomial measure, averaging two adjacent boxes at each step produces strong oscillations with induce errors in the regression estimation step, and thus both in the exponent and dimension computations.

As said above, this method in fact introduces an implicit dependence between n and ε through step 2. Although it is desirable, in numerical methods, to pass from two limits on n and on ε to just one limit, it is clear that the dependency between n and ε should be carefully designed.

Indeed, more precise results may be obtained using a classical tool in density estimation: the kernel method, or more precisely the double kernel method ([11]). Kernel methods have been extensively studied and powerful theorems are known that assess the quality of the results. The difficulty here is that f_g is not the density corresponding to the α's, but rather a double logarithmic normalization of this density. Applying a kernel procedure to estimate f_g has the following advantages:

- The choice of the number of boxes is no longer ad-hoc but driven by the data. The dependence on ε is made explicit.
- No averaging needs to be done, since the spectrum is evaluated at only one resolution. Of course, in the general case, one needs to verify

that the spectra estimated at each resolution are consistent. Such a verification should be performed carefully. Note that other authors have considered one resolution estimation methods ([17]).

- The use of smooth kernels allows to obtain more regular estimates.

For simplicity, we only consider the 1D case with $\mathcal{P}$, the sequence of partitions $\mathcal{P}_n$ of $[0,1[$, being made of dyadic intervals, i.e.:

$$\mathcal{P}_n = \{I_k^n\}_{0 \leq k < 2^n}, I_k^n = [k2^{-n}, (k+1)2^{-n}[$$

(recall section 3.3 for the notations.)

The starting point of the method is to write $N_\varepsilon^n(\alpha)$ as:

$$N_\varepsilon^n(\alpha) := \#\{\alpha_k^n / \alpha_k^n \in \mathcal{B}(\alpha, \varepsilon)\} = 2^{n+1}\varepsilon K_\varepsilon * \rho_n(\alpha)$$

where $\mathcal{B}(\alpha, \varepsilon)$ is the open ball centered at α of radius ε, ρ_n is the density of the $(\alpha_k^n)_k$, and K_ε is the rectangular kernel: $K(x) = 1$ for $x \in \left[-\frac{1}{2}, \frac{1}{2}\right]$, 0 outside, and $K_\varepsilon(x) = \frac{1}{\varepsilon} K\left(\frac{x}{\varepsilon}\right)$.

It is easily shown that K may be in fact chosen to be any compactly supported kernel. A classical problem in density estimation is the choice of an ε such that $K_\varepsilon * \rho_n$ is as close as possible to the limiting density ρ. In this setting, ε becomes a function of n. Adapting classical techniques allows to reduce the two limits in f_g to one limit, with an optimal choice of ε. However, one has to check under which conditions it is indeed possible to get rid of the limit on ε.

Let $f_g^\varepsilon(\alpha) = \overline{\lim}_{n \to \infty} \frac{\log N_\varepsilon^n(\alpha)}{n}$. The following results hold ([1]):

Theorem: *Let $\varepsilon_0 > 0$. The following statements are equivalent:*

- $f_g^\varepsilon(\alpha) = \sup_{x \in \mathcal{B}(\alpha, \varepsilon)} f_g(x) \qquad \forall \varepsilon \in (0, \varepsilon_0)$.
- $\varepsilon \mapsto f_g^\varepsilon(\alpha)$ *is left continuous for* $\varepsilon \in (0, \varepsilon_0)$.

Theorem: *Assume f_g is continuous on a neighborhood of α. Then, for sufficiently small ε,*

$$f_g^\varepsilon(\alpha) = \sup_{x \in \mathcal{B}(\alpha, \varepsilon)} f_g(x)$$

The following corollaries are then easy to prove:

Corollary: *Assume f_g is continuous on a neighborhood of α_0. Then $\alpha \mapsto f_g^\varepsilon(\alpha)$ is continuous at α_0.*

Corollary: *Assume f_g is continuous on $[\alpha_m, \alpha_M]$. Then $(f_g^\varepsilon)_\varepsilon$ converges uniformly to f_g when ε goes to 0^+.*

The two following results deal with the problem of passing from two limits to one limit in the computation of f_g.

Assume f_g is greater than or equal to 0 on $[a, b]$ and $-\infty$ outside. Then it is easy to see that $\overline{\bigcup_{n,k} \alpha_k^n} = [a, b]$. Let $(\beta_j^n)_j$ denote the ordered set of $(\alpha_k^n)_k$ at resolution n without multiplicity, i.e.

$$\beta_0^n = \min_k \alpha_k^n, \qquad \beta_1^n = \min_{k/\alpha_k^n \neq \beta_0^n} \alpha_k^n, \qquad \ldots.$$

With a slight abuse of notation, let $N_0^n(\alpha) := \#\{k \ / \ \alpha_k^n = \alpha \ \}$, and define $\widehat{\beta^n(\alpha)}$ as the smallest β_j^n that verifies:

$$|\beta_j^n - \alpha| < \varepsilon \qquad N_0^n(\beta_j^n) \geq N_o^n(\beta_k^n) \qquad \forall \beta_k^n \in \mathcal{B}(\alpha, \varepsilon)$$

Define:

$$\tilde{f}_g(\alpha) := \lim_{\varepsilon \to 0} \overline{\lim_{n \to \infty}} \frac{\log N_0^n(\widehat{\beta^n(\alpha)})}{n}$$

Finally, let $\varepsilon_\alpha(n) = |\alpha - \widehat{\beta^n(\alpha)}|$ and

$$\varepsilon_n := \sup_\alpha \varepsilon_\alpha(n) = 1/2 \ \max_j |\beta_{j+1}^n - \beta_j^n|$$

Define:

$$f_g^n(\alpha) := \frac{\log N_{\varepsilon_n}^n(\alpha)}{n}$$

The following results hold:

Proposition:

$$\forall \alpha, \qquad \tilde{f}_g(\alpha) \leq \overline{\lim_{n \to \infty}} f_g^n(\alpha) \leq f_g(\alpha)$$

Proposition: *Assume that there exist positive reals a, b, ε_0 such that*

$$\forall \varepsilon < \varepsilon_0, \forall n, \forall \alpha, \ N_\varepsilon^n(\alpha) \leq b \, n^a \ N_0^n(\widehat{\beta^n(\alpha)})$$

Then

$$\forall \alpha, \qquad \tilde{f}_g(\alpha) = f_g(\alpha)$$

Corollary: *Under the same hypothesis as the previous proposition,*

$$\forall \alpha, \qquad \tilde{f}_g(\alpha) = \overline{\lim_{n\to\infty}} f_g^n(\alpha) = f_g(\alpha)$$

The last series of propositions are specific to the case of multinomial measures. Generalizing the above definition, let, for an arbitrary sequence $(\varepsilon_n)_n$ of positive reals:

$$f_g^{\varepsilon_n}(\alpha) := \frac{\log N_{\varepsilon_n}^n(\alpha)}{n}$$

Proposition: *Let μ be a binomial measure with weights (m_0, m_1), and $a = \left|\log_2 \frac{m_1}{m_0}\right|$. Let $(\varepsilon_n)_n$ be such that: $\varepsilon_n \geq \frac{a}{n^2}, \varepsilon_n \to 0$. Then, for all α, there exists a subsequence $s(n)$ (depending on α) such that*

$$\lim_n f_g^{\varepsilon_{s(n)}}(\alpha) = f(\alpha)$$

Moreover, if $\varepsilon_n \geq \frac{a}{2n}, \varepsilon_n \to 0$, then

$$\lim_n \sup_\alpha \left| f_g^{\varepsilon_n}(\alpha) - f(\alpha) \right| = 0$$

If we replace the rectangular kernel by a Gaussian kernel, we obtain

Proposition: *For a binomial measure, the optimal ε_n (in the sense of uniform convergence) for a Gaussian kernel is such that there exists $c > 0$ with $\lim_n(\varepsilon_n / c\sqrt{n}) = 1$.*

The double kernel method consists in choosing $(\varepsilon_n, \varepsilon'_n)$ such that the estimates $f_g^{\varepsilon_n}$ and $h_g^{\varepsilon'_n}$ with two different kernels are as close as possible. For instance, let $f_g^{\varepsilon_n}$ correspond to the rectangular kernel and $h_g^{\varepsilon'_n}$ correspond to the triangular kernel ($T(x) = 1 - x$ for $x \in [0,1], T(x) = 0$ for $x > 1, T(-x) = T(x)$). The following propositions are valid for finite sums or lumpings of binomial measures:

Proposition: *Let $(\varepsilon_n, \varepsilon'_n)$ be sequences that minimizes*

$\sup_\alpha(f_g^{\varepsilon_n}(\alpha) - h_g^{\varepsilon'_n}(\alpha))$. *Then* $f_g^{\varepsilon_n} \to f(\alpha)$ *for all* α.

Proposition: *Let* ε_n *be such that there exists* $c > 0$ *with* $\lim_n \frac{\varepsilon_n}{cn} = 1$. *Then* $\lim_n \sup_\alpha(f_g^{\varepsilon_n}(\alpha) - f(\alpha)) = 0$.

It should be noted that the double kernel method does not in general give the optimal sequence ε_n. Some results using a modification of the above ideas in a $Max - plus$ frame (see [2],[5]) are shown below. The theoretical f_g along with its kernel estimate for a lumping (resp. a sum) of two binomial measures is shown on figure 18 (resp. figure 19). A slightly more complex example is shown on figure 20, which is the estimated f_g spectrum of a trinomial measure $\mathcal{T}$ with overlapping: the first map sends $[0,1]$ to $\left[0, \frac{1}{2}\right]$ and has weight $m, 0 < m < \frac{1}{2}$, the second map sends $[0,1]$ to $\left[\frac{1}{4}, \frac{3}{4}\right]$ and has weight $\frac{1}{2}$, and the third map sends $[0,1]$ to $\left[\frac{1}{2}, 1\right]$ and has weight $\frac{1}{2} - m$. The theoretical f_g spectrum is not known in this case. However, it is readily proven, using for instance Fourier transformation, that $\mathcal{T}$ is the convolution of the Lebesgue measure with a binomial measure with weights $(m, 1-m)$. In particular, all points have Hölder exponent greater than one, a fact which is reasonably well verified on figure 20. Further experiments may be found in [25].

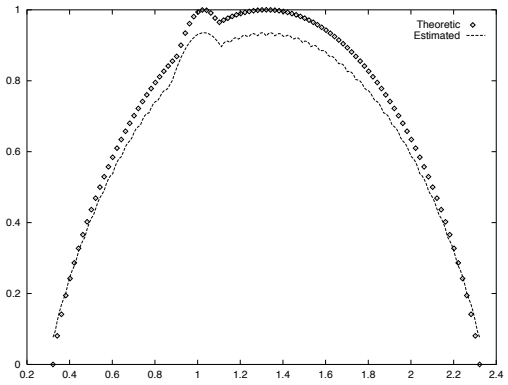

FIG. 18 - *Theoretical and estimated* f_g *spectrum for the lumping of two binomial measures.*

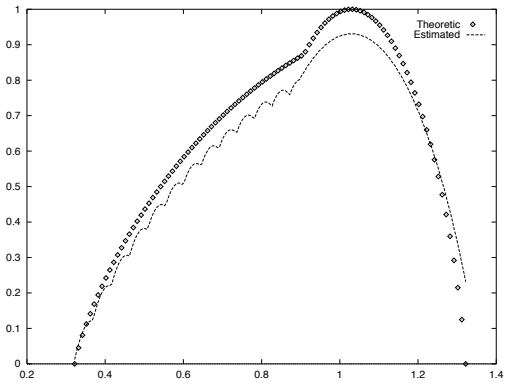

FIG. 19 - *Theoretical and estimated f_g spectrum for the sum of two binomial measures.*

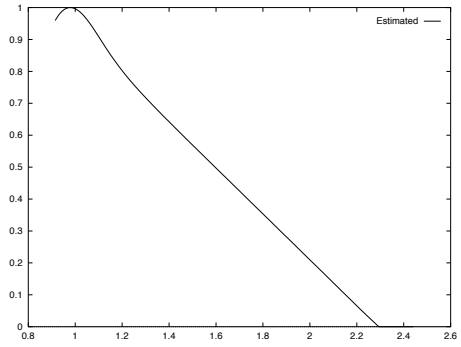

FIG. 20 - *Estimated f_g spectrum for a trinomial measure with overlapping.*

4.3 Image enhancement

The simplest application of multifractal theory in image analysis is *enhancement*. Consider for instance the *sum* measure. It is easily shown that points in smooth zones have $\alpha = 2$. Moreover, when one computes the spectrum of this measure on a "typical" image (for instance a non noisy optical image of an indoor scene), one finds that $f(\alpha = 2) = 2$: "most" points have exponent 2, because there exist large smooth regions in the signal. On the contrary, the spectrum computed on, say, a radar image has $f(\alpha_0) = 2$ for an α_0 strictly lower than 2 and $f(\alpha = 2) < 2$. A natural idea to "de-noise" the radar image is to transform it in such a way that its spectrum is shifted by an amount $\delta = 2 - \alpha_0$ along the α axis. The new image will then have a set of points in "smooth" zones whose dimension is 2. Other distortions of the spectrum may of course be considered. More on this topic can be found in [18]. We simply show here an example on a radar image (figure 21). Figure 22 displays a de-noising by a wavelet method (the wavelet coefficients soft thresholding of Donoho, [12]). Figure 23 shows the result of a spectrum shift. Note that, contrarily to the wavelet method, the multifractal one is not a filtering, since it is invertible. Figures 24, 25 and 26 display the result of Canny's edge detector as applied on the three images, the threshold on the norm of the gradient being set such that no spurious contours appear in the sea region.

4.4 Change detection

An important application of image analysis is to provide means to monitor a scene over a period of time and to detect changes in the content of the scene. In photo-interpretation, change detection consists in finding significant differences – most of the time man-made changes in opposition to natural and/or seasonal changes – between a new image and site models derived from older images. In biomedical imagery, the aim is to control a disease evolution and its cure. In both contexts, most existing methods require a prior knowledge of the objects to be extracted in the new image.

The basic idea of change detection in the multifractal frame is to analyze the new image w.r.t. the older ones. More precisely, one defines a sequence of capacities on the new image, and compute its multifractal spectrum using the *sum* measure on some average of the old ones as the *reference* measure (μ with the notations of section 3.3). Intuitively, the spectrum will then highlight the importance of the changes in the new scene. More details on this method can be found in [8].

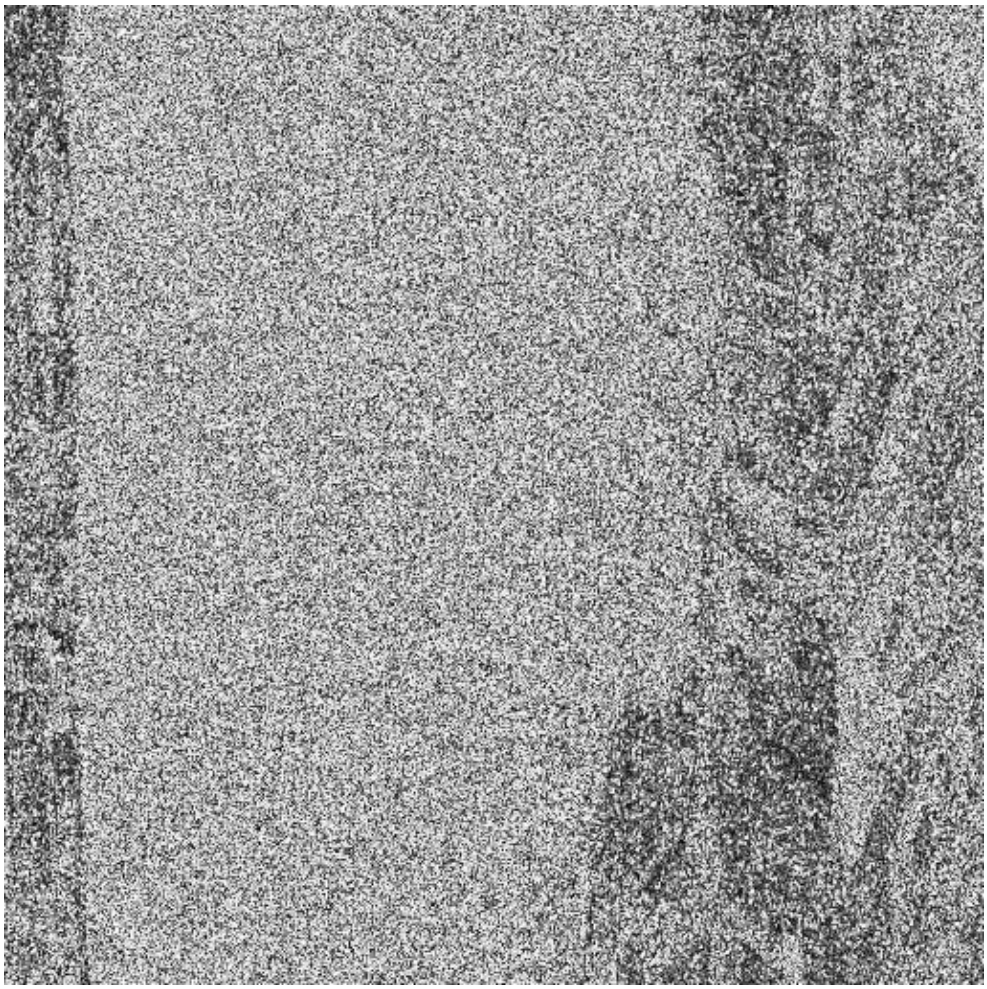

FIG. 21 - *A radar image, displaying sea in its middle part and land on the left and right sides (courtesy Alcatel Espace).*

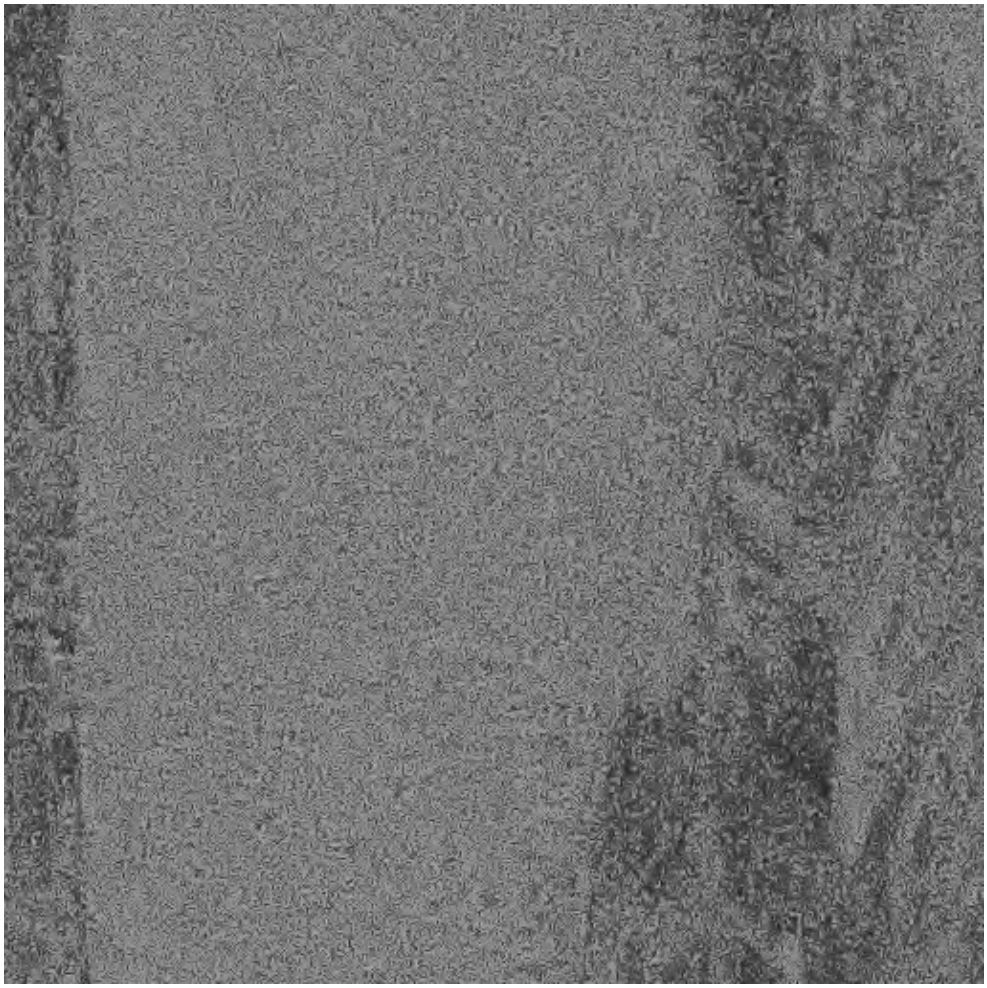

FIG. 22 - *De-noising by soft thresholding of the wavelet coefficient.*

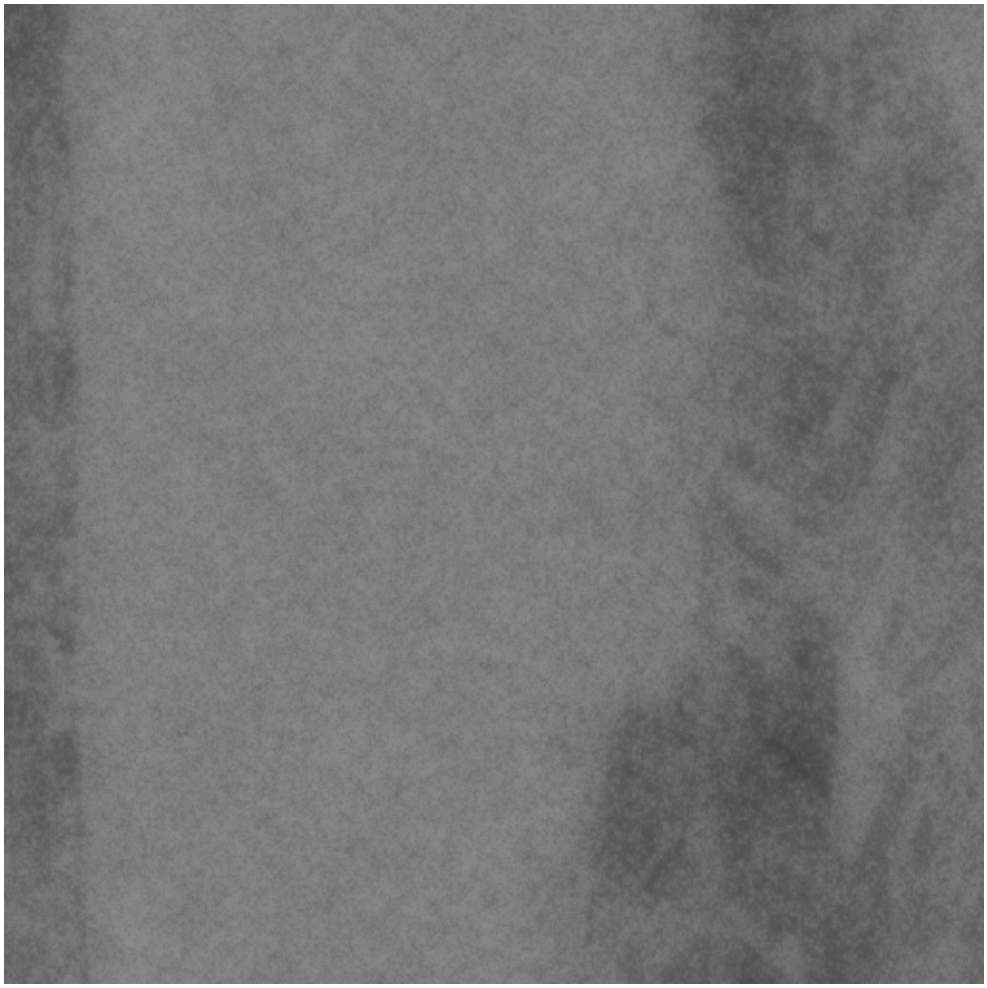

FIG. 23 - *De-noising by multifractal spectrum shifting.*

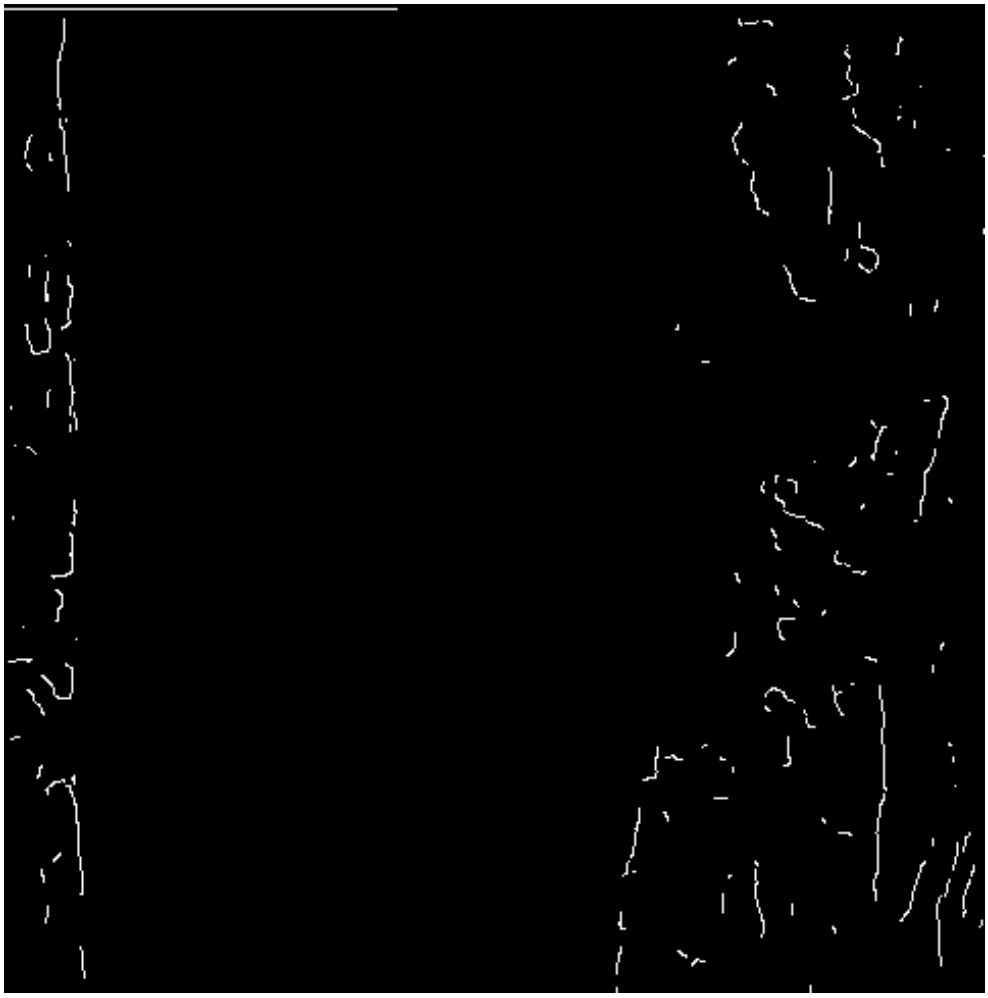

FIG. 24 - *Edges obtained from image 21 with Canny's detector, thresholded such that no edges appear in the sea region.*

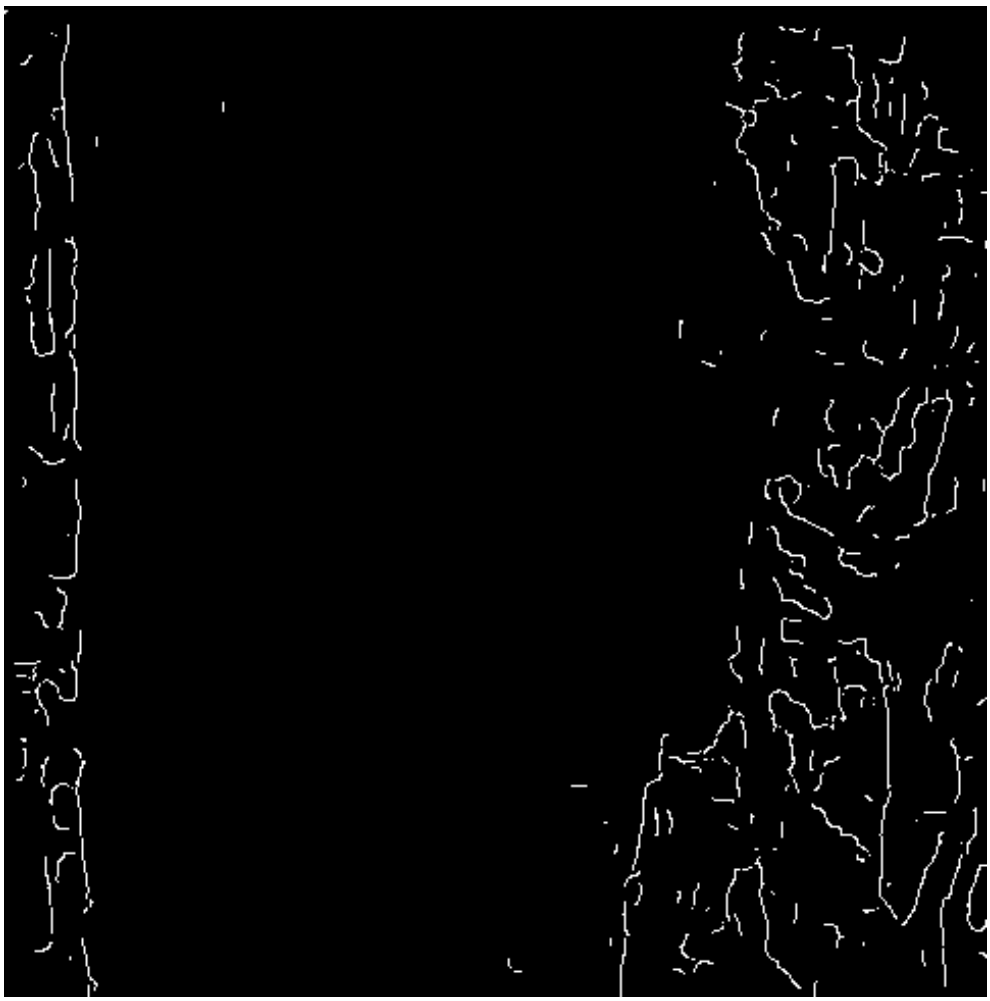

FIG. 25 - *Edges obtained from image 22 with Canny's detector, thresholded such that no edges appear in the sea region.*

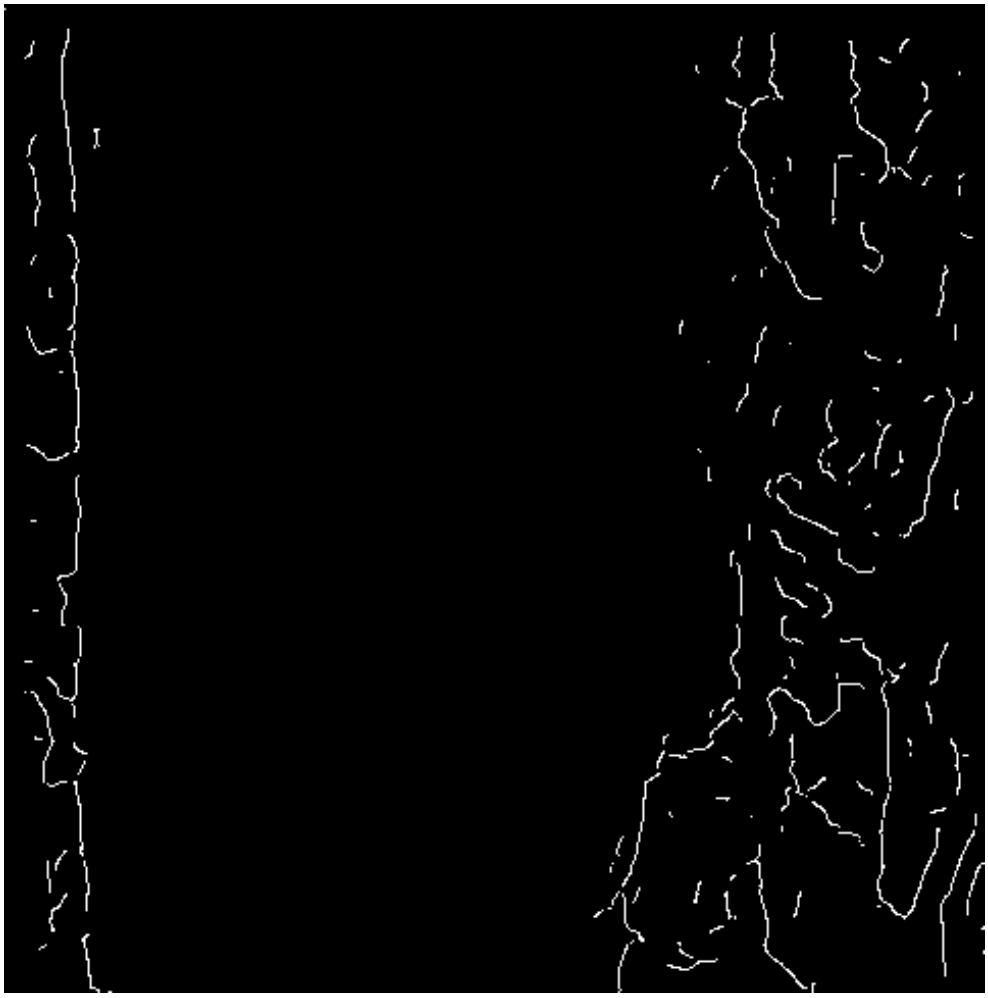

FIG. 26 - *Edges obtained from image 23 with Canny's detector, thresholded such that no edges appear in the sea region.*

4.5 Edge detection

The simplest procedure for extracting edges using multifractal analysis is as follows :

- Choose a sequence c of capacities.
- Compute the Hölder exponent of c at each point of the image.
- Compute the multifractal spectrum of c.
- Declare as "smooth" edge points those belonging to the set(s) E_α whose dimension is one.
- Declare as "irregular" edge points those belonging to the set(s) E_α whose dimension is between two fixed values, typically 1.1 and 1.5.

Results of segmentation using this approach are shown on figures 27, 28, 29 and 30.

FIG. 27 - *Original image.*

The *sum* measure, *max* and *iso* capacities are the three basic kinds of capacities used for analyzing images. There are cases where additional specific capacities have to be designed in order to get robust and precise results.

FIG. 28 - *Original image with histogram equalization.*

FIG. 29 - *Smooth edge points obtained with max capacity.*

FIG. 30 - *Irregular edge points obtained with max capacity.*

This happens for instance in medical imaging. In certain situations, it is not enough to use one capacity (or one sequence of capacities), and one has to define a vector of capacities. To fully exploit the information provided by the different capacities and their spectra, it is convenient to adopt a Bayesian approach: a given point (x, y) in the image is characterized by a couple (t, λ), where t denotes the type of the singularity, and λ its height.

Let A be the vector of Hölder exponents at point (x, y). For instance, $A = (\alpha_{sum}, \alpha_{max}, \alpha_{iso})$. As is usual, one writes:

$$\Pr((t,\lambda)/A) = \frac{\Pr(A/(t,\lambda)) \times \Pr(t,\lambda)}{\Pr(A)}$$

and one looks for the couple (t, λ) that maximizes the left hand side. Since $\Pr(A)$ is constant, this is equivalent to maximizing the product $\Pr(A/(t, \lambda)) \times \Pr(t, \lambda)$. The computation of the conditional probability $\Pr(A/(t, \lambda))$ is quite involved in the general case. Only a few simple situations allow for an analytical computation. In other cases, numerical simulations are performed.

The computation of $\Pr(t, \lambda)$ is based on the fact that, when (x, y) does not lie in a smooth region, t and λ are independent:

$$\Pr(t,\lambda) = \Pr(t) \times \Pr(\lambda)$$

It may be shown that:

$$\begin{cases} \Pr(t \in T) &= \Pr(\alpha_n^{iso} \in I(T)) \\ \Pr(\lambda \in \wedge) &= \Pr(\alpha_n^{max} \in M(\wedge)) \end{cases}$$

with explicit relations between T and $I(T)$, and $\wedge$ and $M(\wedge)$. It is thus possible to evaluate the prior probabilities using the f_g spectra related to the *iso* and *max* capacities.

A different approach that allows to refine the segmentation is to use multifractal correlations. In this case, the vector of Hölder exponents at point (x, y) is $A = (\alpha, \alpha')$ as defined in section 3.5. More complex situations call for the use of correlations of order higher than 2. A complete presentation of this topic is however beyond the scope of this paper.

4.6 Acknowledgments

Parts of the work described in sections 3 and 4 were done in collaboration with my students J.P. Berroir, J. Bestel, C. Canus, B. Guiheneuf, P. Mignot and R. Vojak. I warmly thank M. Akian and R. Riedi for many stimulating discussions. Parts of this research were funded by Alcatel Espace, Alcatel ISR, INRETS and Dassault Aviation.

Références

[1] M. Akian. Personnal communication.

[2] M. Akian. Densities of idempotent measures and large deviations. Technical Report 2534, INRIA, April 1995.

[3] L. Alvarez, F. Guichard, P. L. Lions, and J. M. Morel. Axioms and fundamental equations of image processing. *Archi. for Rat. Mech.*, 3(123):199–257, 1993.

[4] M. Arbeiter and N. Patzschke. Random self-similar multifractals. *Adv. Math.*, to appear.

[5] J. Bestel and J. Lévy Véhel. A $Max - plus$ approach to large deviation multifractal spectrum computation. Technical report, INRIA, September 1996.

[6] G. Brown, G. Michon, and J. Peyrière. On the multifractal analysis of measures. *Journal of Stat. Phys.*, t. 66:775–790, 1992.

[7] J. F. Canny. Finding edges and lines in images. Technical Report T.R. 720, MIT, June 1983.

[8] C. Canus and J Lévy Véhel. Change detection in sequences of images by multifractal analysis. In *ICASSP*, May 1996.

[9] C. Canus and J. Lévy Véhel. Sub-optimal choice a sequence of capacities for multifractal image analysis. Technical report, INRIA, September 1996.

[10] M.E. Cates and J.M. Deutsch. Spatial correlations in multifractals. *Phys. Rev. A.*, 35(11):4907–4910, 1987.

[11] L. Devroye. The double kernel method in density estimatio. *Ann. Inst. Henri Poincaré.*

[12] D.L. Donoho. De-noising by soft thresholding. *IEEE Trans. on Info. Theory*, (41):613–627, 1995.

[13] C.J.G. Evertsz and B. Mandelbrot. *Multifractal Measures*, pages 199–257. Springer Verlag, 1992.

[14] K.J. Falconer. The multifractal spectrum of statistically self-similar measures. *Journal of Theo. Prob.*, 7(3), July 1994.

[15] F. Family. *Phys. Rev. E*, 47:2281, 1993.

[16] U. Frisch and G.Parisi. *Turbulence and predictability in geophysical fluid dynamics and climate dynamics*, page 84. M. Ghil, R. Benzi and G. Parisi, Amsterdam (Holland), 1895.

[17] P. Grassberger, R. Badii, and A. Politi. Scaling Laws for Invariant Measures on Hyperbolic and Nonhyperbolic Atractors. *Journal of Statistical Physics*, 51(1/2):135–178, July 1988.

[18] B. Guiheneuf and J. Lévy Véhel. Image enhancement through multifractal analysis. Technical report, INRIA, July 1996.

[19] H.G.E. Hentschel and I. Procaccia. The infinite number of generalized dimensions of fractals and strange attractors. *Physica 8D*, 1983.

[20] R. Holley and E. Waymire. Multifractal dimensions and scaling exponents for strongly bounded random cascades. *Ann. App. Prob.*, 2(4):819–845, 1992.

[21] S. Jaffard. Multifractal formalism for functions, part 1 and part 2. *SIAM Journal Math. Anal.* to appear.

[22] S. Jaffard. Construction de fonctions multifractales ayant un spectre de singularités prescrit. *C.R. Acad. Sci. Paris*, pages 19–24, 1992. T. 315, Série I.

[23] J.-P. Kahane and J. Peyrière. Sur certaines martingales de B. Mandelbrot. *Adv. Math.*, 22:131–145, 1979.

[24] J. J. Koenderink. The struture of images. *Biol. Cybern*, (50):363–370, 1984.

[25] J Lévy Véhel. Numerical computation of the large deviation multifractal spectrum. In *CFIC*, September 1996.

[26] J. Lévy Véhel and C. Canus. Hausdorff dimension estimation and application to multifractal spectrum computation. Technical report, INRIA, June 1996.

[27] J. Lévy Véhel and R Vojak. Multifractal analysis of choquet capacities: Preliminary results. *Adv. Appl. Math.*, to appear.

[28] B.B. Mandelbrot. A class of multinomial multifractal measures with negative (latent) values for the dimension $f(\alpha)$. In *Fractals (Proceedings of the Erice meeting)*. L.Pietronero, New York, 1989.

[29] B.B. Mandelbrot. Fractal measures (their infinite moment sequences and dimensions) and multiplicative chaos: early works and open problems. Technical report, Physics Department, IBM Research Center, Mathematics Deparment, Harvard University, Cambridge, MA 02138, USA, 1989.

[30] L. Olsen. *Random Geometrically Graph Directed Self-similar Multifractals*. Pitman Research Notes in Mathematic, 1994. Series 307.

[31] L Olsen. A multifractal formalism. *Adv. Math.*, to appear.

[32] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. PAMI*, 7(12):629–639, 1990.

[33] R. Riedi. An improved multifractal formalism and self-similar measure. *Journal Math. Anal. Appl*, (189), 1995.

[34] M. Schmitt and J. Mattioli. *Morphologie mathématique.* Masson, 1994.

[35] J. Serra. Image analysis and mathematical morphology. *Academic press*, 1982.

[36] R. Vojak and J. Lévy Véhel. Higher order multifractal analysis. *submitted to SIAM Journal on Math. An.*