Multifractal Segmentation of Images

Jacques Lévy Véhel, Pascal Mignot

Multifractal Segmentation of Images

Jacques LEVY VEHEL, Pascal MIGNOT
INRIA
Domaine de Voluceau
Rocquencourt - B.P. 105
78153 Le Chesnay Cedex
FRANCE
e-mail: jlv@bora.inria.fr, pascal@bora.inriafr

Abstract

In this work, we propose a multifractal approach to the problem of image analysis. We show that an alternative description of images, based on a multifractal characterization of the signal, can be used instead of the classical approach that involves smoothing of the discrete data in order to compute local extrema. We classify each point of the image according to two parameters, its type of singularity and its relative height, by computing the spectra associated with different kinds of capacities defined from the grey levels. All these informations are then used together through a Bayesian approach.

1 Introduction

The aim of this work is to show the potentialities of a multifractal approach for image analysis. In the fractal community, "image analysis" usually means that we are given an image representing a certain state of a particular process, and that we want to compute some sort of fractal dimension, which is of interest for characterizing the process, using this image

Our concern here is different, since we want to characterize the image itself in terms of fractal features: in other words, the object of study is the image, and the fractal approach is used to describe its structure. Thus our work is a fractal approach to the widely studied image analysis problem.

In section 2, we state the basic problems of image analysis and describe some of the "classical" solutions that have been proposed. In section 3, we recall some definitions and results of the multifractal theory. These results are used in section 4 for a multifractal description of images. We present some results on synthetic and real images, before concluding and proposing some desirable extensions.

2 Classical Approach to Image Segmentation

Image Analysis is an important research field which has a number of applications in robotics, medical imaging, satellite imaging, etc ...

We restrict ourselves here to the problem of image segmentation: segmentation means that we want to extract from the image a compact description in terms of edges and/or regions. Thus, we do not tackle the problem of higher level interpretations such as recognition for instance.

Essentially, image segmentation consists in finding all the characteristic entities of an image: these entities are either described by their contours (edge detection) or by the region where they lie (region extraction). These two approaches are dual, but their algorithms are very different, and, unfortunately, most of the times lead to different segmentation results.

Edge Detection

It is by far the most widely used approach. The core of the classical methods is the assumption that edges usually corresponds to local extrema of the gradient of the grey levels in the image. In this setting, one then has to tackle the problem of computing some kind of "derivative" of a noisy discrete signal.

Let $I(x, y)$ be the image (noisy) signal. An edge is defined by its type: a step edge is a 0^{th}-order discontinuity of I, a roof-edge is a 1^{st}-order discontinuity of I, ... Let $G(I)$ be the gradient of I. The problem reduces to the determination of a filter yielding a good approximation of G. Under some assumptions on the nature of the noise, it may be shown that the problem is equivalent to that of finding an optimal linear filter f such that:

$$G = (I * f)' = I * f'$$

In order words, we start by smoothing the discrete image data I by convolving it with f, and then compute the gradient by differentiating the smoothed signal. Edge points are then defined to be the local maxima of the gradient's norm in the gradient's direction. Using additionnal criteria, one can derive expressions for optimal filters. A frequently used one is:

$$f(x) = -ce^{-\alpha|x|}\sin(wx) \quad \text{or} \quad f(x) = -cxe^{-\alpha|x|} \qquad \text{(Canny-Deriche filter)}$$

It is also possible to refine the method using a multiresolution scheme: the original image undergoes a series of successive smoothings, and, at each step, some characteristic points (maxima of the transform) are computed. These points are then used in collaboration through a propagation method, to describe more robustly and accurately the edges ([12]).

Region Extraction

The idea here is to separate the image into regions that verify a given uniformity criterion. If we are dealing with very simple images, the criterion might just be that all points belonging to a certain region must have the same grey level. However, in general, images include textured zones, and one has to solve the much harder problem of texture discrimination. For more complete discussion, see citeMonga87,lvpmjpb92.

3 Basics of the Multifractal Theory

We define here our notations and briefly recall some basic facts about the multifractal theory ([1, 2, 3, 4, 14, 13]).

Let μ be a Borel probability measure on $[0, 1] \times [0, 1]$. Let ν_n be an increasing sequence of positive integers, and define:

$$I_{i,j,n} = \left[\frac{i}{\nu_n}, \frac{i+1}{\nu_n}\right] \times \left[\frac{j}{\nu_n}, \frac{j+1}{\nu_n}\right]$$

We consider the following quantities:

$$\tau_n(q) = \frac{1}{\log \nu_n} \log \sum_i{}^* \sum_j{}^* \mu(I_{i,j,n})^q$$

where $\sum^*$ means that the summation runs through those indices (i,j) such that $\mu(I_{i,j,n}) \neq 0$. When the limit exists, we set:

$$\lim_{n\to\infty} \tau_n(q) = \tau(q)$$

We then define $f_l(\alpha)$ as the following Legendre transform of $\tau(q)$:

$$f_l(\alpha) = \inf_{q\in\mathbb{R}} (\alpha q - \tau(q))$$

On the other hand, we consider the sets:

$$E_\alpha = \left\{ (x,y) \in [0,1[\times[0,1[/ \lim_{n\to\infty} \frac{\log \mu(I_n(x,y))}{\log \nu_n} = \alpha \right\}$$

with

$$I_n(x,y) = \{ I_{i,j,n} / (x,y) \in I_{i,j,n} \}$$

α is the local Hölder exponent at point (x,y), and we define $f_h(\alpha)$ as the Haussdorff dimension of E_α.

Finally, we consider the following double limit:

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

where:

$$N_n^\varepsilon(\alpha) = \mathrm{card}\{ I_{i,j,n} / \alpha_n(I_{i,j,n}) \in [\alpha - \varepsilon, \alpha + \varepsilon[\}$$

and α_n is the coarse grained Hölder exponent of μ at $I_{i,j,n}$:

$$\alpha_n(I_{i,j,n}) = \frac{\log \mu(I_{i,j,n})}{\log \nu_n}$$

A central concern of the multifractal theory is to compare the three descriptions of the singularities of the measure, namely the "spectra" $(\alpha, f_l(\alpha))$,$(\alpha, f_g(\alpha))$ and $(\alpha, f_h(\alpha))$. This has important applications. Indeed, $\tau(q)$, and thus $f_l(\alpha))$, is usually much easier to compute on experimental data than the other spectra: $\tau(q)$ is obtained by averaging over many "boxes" and then taking the limit. $f_g(\alpha)$ is more difficult to evaluate, both theoretically and practically, specially on real noisy data, since pointwise computations are necessary. As for $f_h(\alpha)$, it is even much more complex, since the computation of a Hausdorff dimension is typically very involved.

Under very general assumptions, it has been proven that ([1]):

$$f_h(\alpha) \leq f_l(\alpha)$$

It is also possible to prove that in general ([8]):

$$f_g(\alpha) \leq f_l(\alpha)$$

For certain special classes of measures, including multinomial measures, we have an equality:

$$f_n(\alpha) = f_g(\alpha) = f_l(\alpha)$$

when all quantities are the same, we simply note them $f(\alpha)$.

In the case of multinomial measures, $f(\alpha)$ is a bell-shaped curve. This shape is also observed for a number of natural phenomena. However, this is by no way a general property, as one can prove that any ruled function can be the spectrum of a multifractal function (see [5]), or capacity (see [8]).

Other "special" features of f may appear depending on the construction of the measure, as for instance negative values (see [13]). In general, it is easy to construct a measure for which $f_l(\alpha)$ is strictly greater than $f_g(\alpha)$ and $f_h(\alpha)$. We shall call the assumption that $f_g = f_h = f_l$ the "strong assumption", and the assumption that $f_g = f_h$ the "weak assumption".

4 Application to Image Analysis

4.1 Definition of the Measures

Though fractal geometry has been introduced a long time ago in image analysis, it is not yet used extensively [16, 10].

Some authors have used the fractal dimension to perform texture classification and image segmentation, other have used higher order dimensions or measures, as correlation or lacunarity [6, 9], to refine the results and have obtained some interesting results. Very few papers have been devoted to the use of multifractals in image analysis[11], although we believe that approaches based on the computation of the fractal dimension are largely unfounded. This approach assumes that the 2D grey level image can be seen as a 3D surface, or, equivalently, that the grey levels can be assimilated to a spatial coordinate on the z-axis. This assumption has no theoretical basis, since the scaling properties of the grey levels are generally different from those of the space coordinates. Instead, we should look at the grey levels as a measure, laid upon a generally compact set, totally inhomogeneous to space coordinates. This leads to a multifractal analysis.

A natural choice is to define the measure μ as the sum of intensities of pixels in the measured region. This measure will be useful, but it will not be sufficient for a fine description of the image. One possibility is then to define other types of functions of the grey levels, and to apply the multifractal analysis to them. Since the notion of resolution is of great importance in image analysis, we find it more appropriate to work with set functions than with point functions. However, it occurs that those functions that are relevant in our field are not in general measures, but rather capacities. Lack of space prevents us from presenting the extension of the multifractal analysis to capacities, thus we refer the interested reader to [8] and just define the capacities that we will need.

We introduce "max", "min" and "iso" capacities of a region Ω. If $\Omega^\star$ is the subset of Ω where intensity is non-zero, and $p(i)$ is the intensity of the point i, we define:

$$\mu_{\max}(\Omega) = \max_{i\in\Omega} p(i) \quad \text{and} \quad \mu_{\min}(\Omega) = \min_{i\in\Omega^\star} p(i) \tag{1}$$

If $G(\Omega)$ is the geometrical center of Ω, we define:

$$\mu_{iso}(\Omega) = \text{Card}\{i \in \Omega / p(i) = p(G(\Omega))\} \tag{2}$$

The exponents computed with those capacities give different informations on the singularities encountered: $\alpha_{\max}$ and $\alpha_{\min}$ only depends on the height of the singularity, α_{iso} only depends on the kind of singularity, and α_{sum} depends on both height and kind of the singularity.

4.2 Edge Detection Using Multifractal Characterizations

4.2.1 Introduction

The approach here is, in some sense, inverse to the classical one explained in section 2: instead of smoothing the discrete data in order to be able to compute some derivatives, we stay with our initial discrete values and quantify the singularity around each point; we then characterize an edge point as a point having a given value of singularity.

This procedure is based on the idea that, in some cases, it might be impossible to recover an underlying continuous process from the discrete data (if such a process exists ...). Thus it seems more natural to directly model the sampled signal. The advantage is that we do not loose or introduce any information by smoothing. The drawback is that we may well be much more sensitive to noise. This is why we have to define several capacities. Using jointly the local information provided by α and the global one contained in $f(\alpha)$, we are able to construct a operator on the image whose main features are the following ones: it is idempotent (it detects its own result), it reacts differently to different types of singularities (provided that the noise is not too important), and no tuning parameters are needed, as soon as the type and the amplitude of the noise are known. As a drawback, since a more complex analysis of the image is made, the computations are not as fast as with gradient-based edge detectors. A few minutes are needed to analyse a 512x512 image.

4.2.2 Computation of the singularity exponents

We study the behaviour of the singularity exponents for sum, max,min and iso capacities on simplified models of step-edge, corner, line and plane. Figure 1 describes those singularities. There are only two values for the gray levels: p_1, level of the point of interest, and p_2, level of the background. We will denote $V(i)$ the $i \times i$ squared neighborhood centered on the pixel of interest, and $V^\star(i)$ as $V(i)$ minus the pixel of interest.

We stress a very important point here: it is obvious that the objects we consider are far from being "fractals", or even "multifractals". However, our approach does **not** at any point make such an assumption. All we do is suppose that the defined capacity does have a Hölder at each point (a rather weak assumption). This makes it possible to compute a multifractal spectrum, whether the capacity is multifractal or not. Of course, strictly speaking, the

p1 p2

x = 6p1 + 3p2x = 3p1 + 6p2x = 3p1 + 6p2x = 4p1 + 5p2

Figure 1: Step-edge, line , corner line and corner models. Notice that the line and the line corner model give the same capacity, they won't be distinguished further.

spectra corresponding to all of our models reduce to the point (2,2), thus we do not have a multifractal behaviour. The fundamental idea here is that we are not interested in getting the "real" spectrum, but rather to verify that, if we use a certain procedure of estimation, the "spectrum" associated with a typical image will allow a description of the local singularities. In other terms, we are not interested in **absolute** quantitites (the "true" spectrum), but rather in evidencing **differences** between estimated spectrum associated with different sets of images.

α is computed as the estimation of the slope of $\log \mu(V(i))$ versus $\log i$, with $i = 2n + 1$, $n = 0 \ldots$. The maximal size of neighborhoods is related to localization of computation. If we use little neighborhoods, for instance $i \leq 3$, α will react to localized singularities, if we use larger neighborhoods, α will react to more widespread singularities. If we consider a computing neighborhood $V(3)$, it is possible to derive explicitely the probability law of α in some cases, for instance when the noise is additive gaussian or uniform. The calculations are simple but rather tedious, and the formulas are quite long, thus we only give here one law for information:

Law of α (sum measure), for a gaussian additive noise of variance σ: the singularity is characterised by n (n equals 9 for a smooth region, 6 for a step, 4 for a corner, 3 for a line), and $s_1 = \frac{p_1}{\sigma}$ et $s_2 = \frac{p_2}{\sigma}$:

$$
\begin{cases}
f_1(\alpha) &= \frac{1}{9}(3^{2\alpha} - 2\ 3^{\alpha} + 9) \\
S(n) &= n\ s_1 + (9 - n s_2) \\
f_2(\alpha) &= -\frac{2}{9} S(n) 3^{\alpha} + \frac{2}{9}(S(n) + 3^{\alpha} s_1) - 2 s_1 \\
f_\alpha(\alpha) &= \frac{6\sqrt{2}\log 3\ 3^{\alpha}}{\pi f_1(\alpha)}\ \exp\left(-\frac{1}{16}S^2 - \frac{9}{16}s_1^2 + \frac{1}{8}S(n)s_1\right) \times \\
& \qquad \left(-\frac{8}{9} - \frac{\sqrt{\pi} f_2(\alpha)}{3\sqrt{f_1(\alpha)}}\ \exp\left(\frac{9 f_2(\alpha)^2}{64 f_1(\alpha)}\right)\ \mathrm{erf}\left(\frac{3 f_2(\alpha)}{8\sqrt{f_1(\alpha)}}\right)\right)
\end{cases}
$$

Observation of plots of different laws show that there is nearly no chance of confusion between the different types of singularity when the amount of noise is not too large, and that the max and min capacities are more robust to uniform noise.

4.2.3 Results

We first present the detection of a step singularity blurred with rayleigh noise. Result is shown on figure 2. We can see that edge detected by a Canny-Deriche filtering is irregular, and edges detected by the multifractal exponent are far less sensitive to noise. On figure 3, we can see the detection of a line blurred with uniform noise. The line is detected with a fair accuracy by multifractal exponent, and not at all by Canny's filter: we should here have used a specific filter for lines. However, the same multifractal exponent is able to detect both step-edge and line.

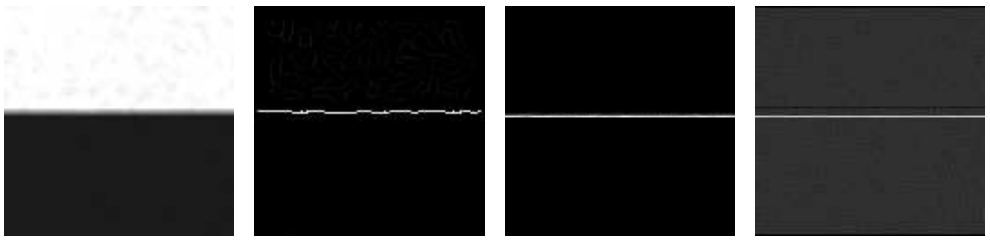

Figure 2: from left to right: blurred step, Canny's edge, edge detected with max capacity on $V(3)$, with sum measure on $V(3)$.

We then present a compared result on a natural scene (figure 4). We can see on that figure that the multifractal exponent is able to detect small details accurately. The most remarkable is the accuracy of the detection of the corners of the door and of the limits of the bush, when Canny's edges only gives good results in presence of a step.

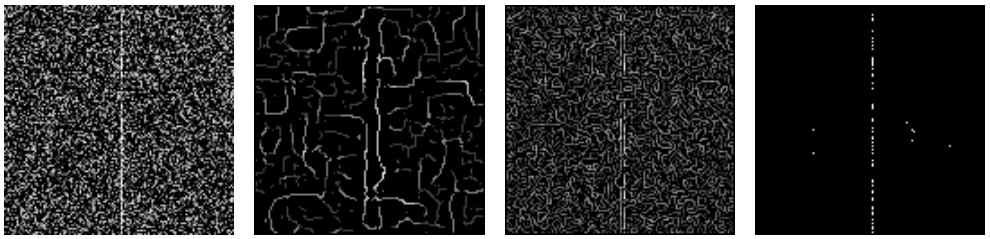

Figure 3: from left to right: blurred line, Canny's edge with large bandwidth, with small bandwidth and edge detected with iso capacity, 9 gray levels on $V(3)$.

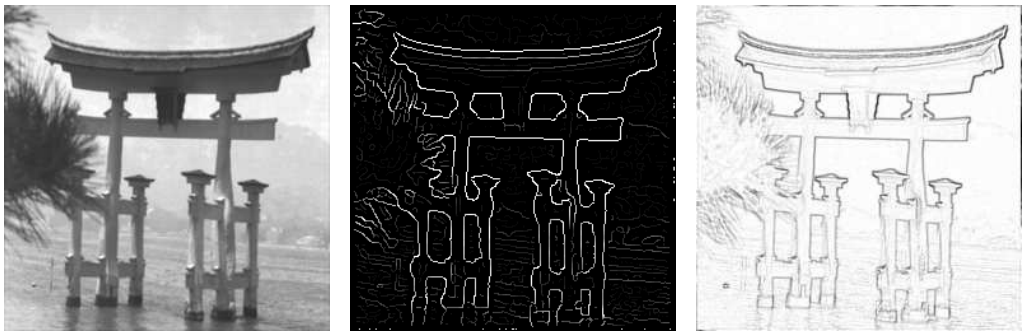

Figure 4: Left: original image, middle: Canny's edges, right: exponent computed with min capacity on $V(3)$.

4.2.4 Use of $f(\alpha)$

In the images presented so far, the use of α computed with well chosen capacities has proven to be sufficient. However, this will not always be the case. In this section, we indicate how the use of $f(\alpha)$ can help us refine our image description. Let us consider figure 5.

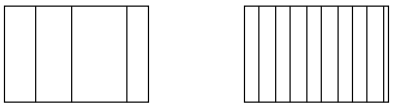

Figure 5: Left: Some edges. Right: A texture

On the left, anyone would see three edges, that are easily detected by any edge detector. On the right, we have done nothing else than triple the number of lines in the image. Of course, it is still possible to interpret this image as being composed of nine edges, but most people would prefer to talk of a binary texture. However our local computation of exponent α would be the same in both situations.

Here appears another characteristic feature of an edge: an edge does correspond to a certain type of singularity in the images, or to an extremum of the gradient (local characterization), but also to a "rare" event, in some sense that has to be defined. In other words, if too many "edges" are detected in a portion of an image, then the human visual system will have a tendency to talk of a textured zone, rather than of a concentration of edges.

This is where we can use the $f_g(\alpha)$ characterization. Remember that $f_g(\alpha)$ measures,

loosely speaking, how rare or frequent an event of singularity α is.

Now if we assume that $f_g(\alpha)$ and $f_h(\alpha)$ are equal (weak assumption), we may assess how "rare" a smooth edge is, because a smooth edge point will belong to a set E_α whose dimension is one. We simply use here the connection between geometry and probability provided by the assumption made on the two spectra. The general line of reasoning is the following one: from a geometrical point of view, a point with prescribed singularity belongs to a set of given $f_h(\alpha)$. If the weak assumption holds, then $f_g(\alpha)$ is also given, and we know the probability of finding such a point in the image **at a given resolution** (this means that all the quantities are computed at these resolution).

In this sense, we may precisely say how an edge, for instance, is characterized both by a given singularity value (local condition) and by the fact that it is a rare event (global condition). To illustrate this, we show in figure 6: the points of figure 4 (original image) belonging to the sets E_α (there might be several such E_α sets) such that $f(\alpha) \approx 2$ (we keep here all the points lying inside regions), and the points where $f(\alpha) \approx 1$ (one can verify that we get most edge points of the original image),

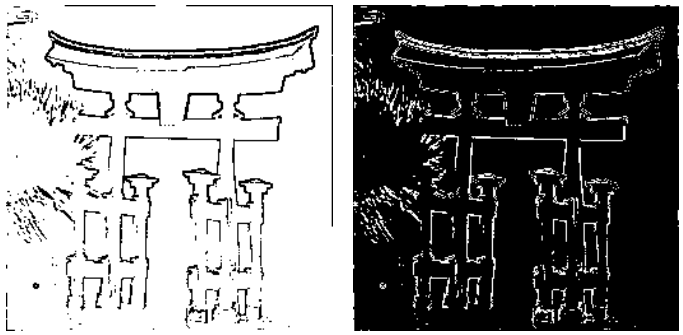

Figure 6: Left: image of points (in white) whose $f(\alpha) = 1.93$. Right: image of points (in white) whose $f(\alpha) = 1.1$.

These ideas can be used more rigorously in a probabilistic setting. The general framework is that of Bayesian optimisation. We restate the problem as follows: at a given point (x, y) in the image, we look for the most probable couple (t, λ), where t is the type of singularity and λ the relative height of singularity at (x, y). Let us denote by A the vector of computed local Hölder exponents at point (x, y), with different measures or capacities. Typically, $A = (\alpha_{\min}, \alpha_{\max}, \alpha_{sum}, \alpha_{iso})$. As is usual in image analysis, we use Bayes rule to write:

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

and we look for the couple (t, λ) that maximizes the left hand side of the above equality. This equivalent to maximize the product $Pr(A/(t, \lambda)) \times Pr(t, \lambda)$, since $Pr(A)$ is a constant here. Thus we have to evaluate two quantities: the conditional probability of a vector of Hölder exponents given a singularity, and the prior probability of a given singularity.

Computation of the conditional probability

This probability is difficult to compute theoretically, and only the cases of uniform noise with $A = (\alpha_{\max}, \alpha_{sum})$ or $A = (\alpha_{\min}, \alpha_{sum})$ have been completed (see [7]).

In the general case, one has to perform computer simulations to obtain the conditional laws.

Computation of the prior probability

Two cases have to be considered: when the point does not lie in a smooth region, it is reasonable to assume that t and λ are independant. Thus:

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

On the other hand, we know that the iso capacity reacts only to the type of the singularity, and that the max capacity reacts only to the relative height of the singularity. In our case, we even have an equivalence between (t, λ) and the coarse grained Hölder exponents, which allows us to write:

$$\begin{aligned} Pr(t \in T) &= Pr(\alpha^n_{iso} \in A_i) \\ Pr(\lambda \in \wedge) &= Pr(\alpha^n_{\max} \in A_m) \end{aligned}$$

where we have used an superscript n to indicate that the coarse grained exponent are computed at resolution n. The sets T, A_i and $\wedge, A_m$ are related by expressions that can be derived explicitely.

To evaluate $Pr(t, \lambda)$, we thus need only to evaluate the two spectra $f_g(\alpha_{iso})$ and $f_g(\alpha_{\max})$.

This can be done directly on the data, using an approach described in citelevy-mignot-berroir.

Finally, when (x, y) lies in a smooth region, another approach using only $(\alpha_{\min}, \alpha_{\max})$ have to be used.

Results obtained with this approach are presented on an aerial photograph (figure 7). A blurring with gaussian noise of variance 0.1 has first been preformed, then the method explained above has been applied at three succesive resolutions. Finally, a propagation algorithm have been used.

5 Conclusion

In this work, we have demonstrated that the use of a multifractal characterization of image points can help to solve the problem of segmentation. Our experiments show that, in several cases, this approach gives at least as good results as the classical ones. Much more work is needed in this direction, but these preliminary results show that the $(\alpha, f(\alpha))$ approach might be able to build a bridge between the two so far unconnected methods of edge detection and region extraction.

References

[1] G. Brown, G. Michon, and J. Peyrière. On the multifractal analysis of measures.

[2] 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.

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

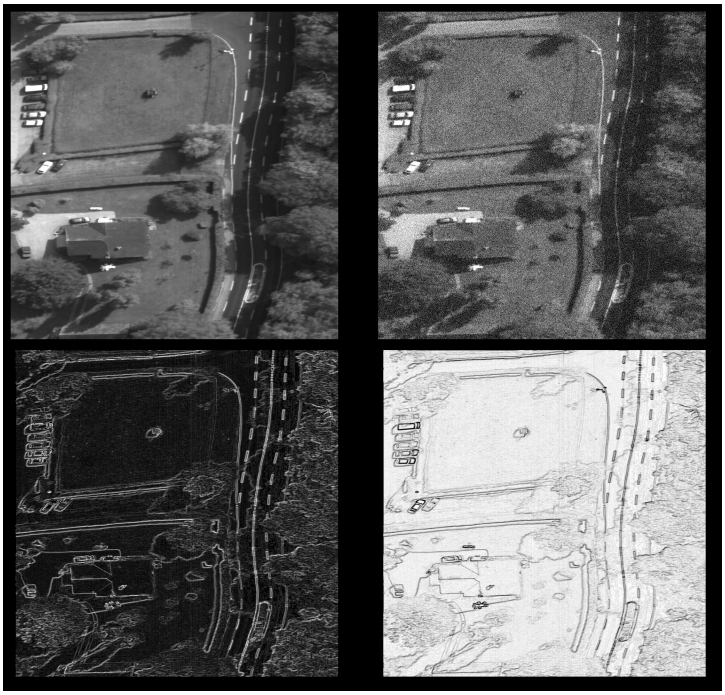

Figure 7: Left: Top left: original image. Top right: blurred

[4] I.Procaccia. The characterization of fractal measures as interwoven sets of singularities : Global universality at the transition to chaos. Technical report, Department of Chemical Physics, The Weizmann Institute of Science, Rehovot 76100, Israel, 1986.

[5] 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.

[6] James M. Keller, Susan Chen, and Richard M. Crownover. Texture description and segmentation through fractal geometry. *Computer Graphics and Image Processing*, 45:150–166, 1989. Edited by Academic Press.

[7] J. Lévy Véhel, , P. Mignot, and J.-P. Berroir. Multifractal image analysis. preprint.

[8] J. Lévy Véhel, , P. Mignot, and R. Vojak. Analyse multifractale de capacités. preprint.

[9] J. Lévy Véhel. About lacunarity, some links between fractal and integral geometry, and an application to texture segmentation. In *ICCV*, 1990.

[10] J. Lévy Véhel. Fractal probability functions : an application to image analysis. In *CVPR*, 1991.

[11] J. Lévy Véhel, P. Mignot, and J.P. Berroir. Multifractals, texture, and image analysis. In *CVPR*, 1992.

[12] S. Mallat and W.L. Hwang. Singularity detection and processing with wavelets. *IEEE Trans. on Information Theory*, 38(2), March 1992.

[13] 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.

[14] 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.

[15] O. Monga. An optimal region growing algorithm for image segmentation. *International Journal of Pattern Recognition and Artificial Intelligence*, 1(3), December 1987. Paris, France.

[16] S. Peleg, J. Naor, R. Hartley, and D. Avnir. Multiple resolution texture analysis and classification. *IEEE*, PAMI-6(4), July 1984.