The multi-fractal structure of contrast changes in natural images: from sharp edges to textures.

ANTONIO TURIEL*and NÉSTOR PARGA†
Departamento de Física Teórica
Universidad Autónoma de Madrid
Canto Blanco, 28049 Madrid, Spain

January 19, 1999

Abstract

We present a formalism that leads very naturally to a hierarchical description of the different contrast structures in images, providing precise definitions of sharp edges and other texture components. Within this formalism, we achieve a decomposition of pixels of the image in sets, the fractal components of the image, such that each set only contains points characterized by a fixed stregth of the singularity of the contrast gradient in its neighborhood. A crucial role in this description of images is played by the behavior of contrast differences under changes in scale. Contrary to naive scaling ideas where the image is thought to have uniform transformation properties [Field 1987], each of these fractal components has its own transformation law and scaling exponents. A conjecture on their biological relevance is also given.

*E-mail: amturiel@delta.ft.uam.es
†To whom correspondence should be addressed. E-mail: parga@delta.ft.uam.es

1 Introduction

It has been frequently remarked that natural images contain redundant information. Two nearby points in a given image are very likely to have very similar values of light intensity, except when an edge lies between them. In this case, the sharp change in luminosity also represents a loss of predictability, because probably the two points belong to two different objects, with almost independent luminosities. Since they cannot be predicted easily, it is frequently said that edges are the most informative structures in the image and that they represent their independent features [Bell and Sejnowski 1997]. Anyone´s subjective experience is consistent with this: most people would start drawing a scene by tracing the lines of the obvious contours in the scene. Even more, the picture can be understood by anyone observing that sketch. Only afterwards, one starts adding texture to the light-flat areas in the drawing. These textures represent the less informative data about the position and intensity of the light sources. It seems that different structures can be distinguished in every image, following a hierarchy defined according to their information content, from the sharpest edges to the softest textures.

Image processing techniques also emphasize the fact that edges are the most prominent structures in a scene. A large variety of methods for edge detection have been proposed, e.g. high pass filtering, zero-crossing, wavelet projections, ... [Gonzalez and Woods 1992]. But once the edges have been extracted from the image, these methods do not address the problem of how to describe and obtain the other structures. In order to detect these other sets and to extract them in the context of an organized hierarchy, it is necessary to provide a more general mathematical framework and, in particular, to define the variables suitable to perform this classification.

In this paper we present a more rigorous approach to these issues, proposing precise definitions of edges and other texture components of images. As we shall see, our formalism leads very naturally to a hierarchical description of the different structures in the images, achieving a decomposition of its pixels in sets, or components, ranging from the most to the less informative structures.

Another motivation to understand the role played by edges in natural images is related to the development of the early visual system. The brain has to solve the problem of how to represent and transmit efficiently the information it receives from the outside world. As it has been frequently pointed out [Barlow 1961, Linsker 1988, Atick 1992, van Hateren 1992], cells located in the first stages of the visual pathway do this by trying to take advantage of statistical regularities in the image ensemble. The idea is that since natural images are highly redundant over space, they have to be represented in a compact and non-redundant way before being transmitted to inner areas in the brain. In order to do this, it is first necessary to know the regularities of the images, which then could be used to build more appropriate and efficient internal representations of the environment.

The statistical properties of natural images have been studied for many years, but the effort was mainly restricted to the second order statistics. The emphasis was put on the two-point correlation first because of the simplicity of the analysis, and secondly because its Fourier transform (i.e. the power spectrum) has a power law behavior that reflects the existence of scale invariant properties in natural images. This implies that the second order statistics does capture some of the regularities. However most of the underlying statistical structure remains unveiled. One can easily convince oneself that most of the correlations in the image are still present after whitening is performed: the contours of the objects in the whitened image are still easily recognizable [Field 1987]. The conclusion is that non-gaussian statistical properties of edges are important to describe natural images, a fact that has also been recognized by [Ruderman and Bialek 1994, Ruderman 1994]. However, no systematic studies of the statistical properties of edges had been done until very recently [Turiel, Mato, Parga and Nadal 1998].

A complete characterization of the regularities of natural images is impossible, and probably useless. Some intuition is necessary to guide the search for regularities. The aim of this work is to exploit the basic idea that changes in contrast are relevant because they occur at the most informative pixels of the scene. Hence we focus our attention in the statistics of changes in luminosity. Since these changes are graded, the task could seem rather difficult and even hopeless, but the solution to this problem turns out to be quite simple and elegant. A single parameter is enough to describe a full hierarchy of differences in contrast and the hierarchy can be explicitly constructed. Moreover, it is intimately related to the scale invariance properties of the image, which are certainly more elaborated than those that have been used until now but are still simple enough and have plenty of structure to be detected by a network adapting to the statistics of the stimuli.

To understand these scale invariances we need first the basic concepts of multiscaling and multifractality [Falconer 1990], together with a simple explanation of their meaning for images. While in naive scaling there is a single transformation law throughout the image (which is then said to be a fractal), in the more general case there is no invariance under a global scale transformation but the space can be decomposed in sets with a different transformation law for each of them (the image is then said to be a multifractal).

To make this concept more precise, let us consider a simple situation where a set of images present a single type of changes in contrast: the luminosity suddenly changes from one (roughly) constant value to another. In this case, a natural description of the properties of changes in contrast would be *to count* the number of such changes that appear along a segment of size r, oriented in a given direction. The statistical distribution of the position and size of these jumps in contrast would give relevant information about an ensemble of such images. However, in real images this counting is not enough to describe well the properties of contrast changes. This is because these changes do not follow such a simple pattern: some of them can be sharp (the notion of sharp changes has to be given), but there will also be all types of softer textures that would be lost by just counting the most noticeable changes. It is then necessary to consider all of them as a whole. A natural way to deal with contrast changes is by defining a quantity that accumulates all of them, whatever their strength, contained inside a scale r, that is:

$$\epsilon_r(\vec{x}) \;=\; \frac{1}{r^2}\int_{x_1-\frac{r}{2}}^{x_1+\frac{r}{2}} dx_1' \int_{x_2-\frac{r}{2}}^{x_2+\frac{r}{2}} dx_2' \quad \mid \nabla C(\vec{x}') \mid \quad . \tag{1}$$

Hereafter, the contrast $C(\vec{x})$ is taken as $C(\vec{x}) = I(\vec{x}) - \langle I \rangle$, where $I(\vec{x})$ is the field of luminosities and $\langle I \rangle$ its average value across the ensemble. The bidimensional integral [1] on the right hand side, defined on the set of pixels contained in a square of linear size r, is a *measure* of that square[2] . It is divided by the factor r^2, which is the usual Lebesgue measure (which we will denote λ) of a square of linear size r. The quantity $\epsilon_r(\vec{x})$ can then be regarded as the ratio between these two measures. More generally, we can define the measure μ of a subset A of image pixels as

$$\mu(A) \;=\; \int_A d\vec{x}' \;\mid \nabla C(\vec{x}') \mid \;, \tag{2}$$

[1] The variables x_1', x_2' are the components of the vector $\vec{x}'$ and $||$ denote the modulus of a vector.

[2] Any measure of a set has the property of being *additive*, that is: if one splits the set in pieces, the measure of the set is equivalent to the sum of the measures of all its pieces.

($\int_A d\vec{x}'$ means bidimensional integration over the set A) what we will call the *Edge Measure* (EM) of the subset A. It is also possible to generalize the definition of ϵ_r for any subset A:

$$\epsilon_A \;=\; \frac{\mu(A)}{\lambda(A)} \qquad . \tag{3}$$

ϵ_A is a density, a quantity that compares how much the distribution of $|\nabla C|$ deviates from being homogeneously distributed on A. We shall call it the *Edge Content* (EC) of the set A. Denoting $B_r(\vec{x})$ the "ball"[3] of radius r centered around $\vec{x}$, it is clear that $\epsilon_r(\vec{x}) \equiv \epsilon_{B_r(\vec{x})}$, so this definition generalizes the previous one. We shall refer to $\epsilon_r(\vec{x})$ as the Edge Content of $\vec{x}$ at the scale r. Notice that $\epsilon_{dr}(\vec{x})$ is the μ-density at $\vec{x}$, $|\nabla C|(\vec{x})$. Thus, eq. (2) can be symbolically expressed as

$$d\mu(\vec{x}) = |\nabla C|(\vec{x}) d\vec{x} \;\; . \tag{4}$$

The main point is that contrast changes are distributed over the image in such a way that the Edge Content has rather large contributions even from pixels that are very close together. As a consequence, the measure presents a rather irregular behavior. But it is precisely in its irregularity where the information about the contrast lies. Let us assume for instance that, as r becomes very small, the measure behaves at every point $\vec{x}$ as

$$\mu\left(B_r(\vec{x})\right) \;\equiv\; \mu_r(\vec{x}) \;=\; \alpha(\vec{x})\; r^{h(\vec{x})+d} \quad , \tag{5}$$

where $d = 2$ is the dimension of the images. For convience a shift in the definition of the exponent of $\mu_r(\vec{x})$ has been introduced. The Edge Content then verifies :

$$\epsilon_r(\vec{x}) \;=\; \alpha(\vec{x})\; r^{h(\vec{x})} \quad , \tag{6}$$

and this choice of $h(\vec{x})$ removes trivial dependencies on r. The exponent $h(\vec{x})$ is just the singularity exponent of $|\nabla C|(\vec{x})$. That is, $h < 0$ indicates a divergence of $|\nabla C|(\vec{x})$ to infinity, while $h > 0$ indicates finiteness and continuous behavior. The greater the exponent the smoother is the density around that point. In general, we will talk about "singularity" even for the positive h cases.

[3] "Ball" in a mathematical sense: it does not need to be true circles, but could also be squares or diamonds. In fact, along this paper the "balls" will be taken as squares.

Thus, the meaning of eq. (5) is that all the points are singular (in this wide sense), and that the singularity exponent is not uniform. This irregular behavior would allow us to classify the pixels in a given image: the set of pixels with a singularity exponent contained in the interval $[h-\Delta_h, h+\Delta_h]$ (where Δ_h is a small positive number) define a class F_h. These classes are the *fractal components* of the image. The smaller the exponent the more singular is the class, and the most singular component is the one with the smallest value of h. A measure μ verifying eq. (5) in every point $\vec{x}$ is called a *multifractal measure.*

We will refer to these components of the images as fractal sets because it will be seen that they are of a very irregular nature. One way to characterize the odd arrangement of the points in fractal sets is by just counting the number of points contained inside a given ball of radius r; we will denote by $N_r(h,\Delta_h)$ this number for the set F_h. As $r \to 0$ it is verified that

$$N_r(h,\Delta_h) \;\sim\; r^{D(h)}. \tag{7}$$

This new exponent, $D(h)$, quantifies the size of the set of pixels with singularity h as the image is covered with small balls of radius r. It is the fractal dimension of the associated fractal component F_h, and the function $D(h)$ is called the *dimension spectrum* (or the *singularity spectrum*) of the multifractal. It is also worth estimating the probability for a ball of radius r to contain a pixel belonging to a given fractal component. Its behavior for small scales r is:

$$Prob(h,\Delta h) \;\sim\; r^{d-D(h)} \quad , \tag{8}$$

where $d=2$ as before. Notice that, since r is small, this quantity increases as $D(h)$ approaches d.[4]

The multifractal behavior also implies a *multiscaling* effect: as every fractal component possesses its own fractal dimension, each of them changes differently under changes in the scale. Thus, although every component has no definite scale and is invariant under the scale transformations, scale invariance is broken for a given image: the points in the image transform with different exponents, according to the fractal component to which they belong. This leads to the presence of intermittency effects: if the image were enlarged, its statistical properties would change. Self-similarity is lost.

[4]It can be proved that any set contained in a d-dimensional real space has fractal dimension D smaller or equal to d.

However, scale invariance is restored in the averages over an ensemble of such images.

At this point one may wonder if a multifractal behavior is actually realized in natural images. One way to check its validity is to verify that eq. (5) is indeed correct. This in turn leads us to the technical problem of how to obtain in practice the exponent $h(\vec{x})$ at a given pixel $\vec{x}$ of the image. This is a local analysis where the singularity is found by looking at the neighbourhood of the point $\vec{x}$. It has the advantage that it not only provides us with a tool to check the multifractal character of the measure but it also gives a way to decompose each image in its fractal components. These issues are addressed in the next Section. The multifractality of the measure can also be studied by means of statistical methods. In this case one looks for the effect of the singulaties on the moments of marginal distributions of, e.g., the Edge Content defined in eq. (1). This is done in Section 3, where the important notions of self-similarity, extended self-similarity and multiplicative processes are explained. Once the relevant mathematical tools are given, these are used in Section 4 to analyze natural images. First in subsection 4.2 we check that the measure introduced in eq. (2) is indeed multifractal. Examples of fractal components of natural images are also presented in the same section. After this we study the consistency between the singularity analysis and the statistical analysis of the measure (subsection 4.3). The results of the numerical study of self-similarity properties of the contrast $C(\vec{x})$ and of the measure density $|\nabla C|(\vec{x})$ are presented in subsection 4.4. The discussion of the results and perspectives for future work are given in the last section. Some technical aspects are included in the Appendix.

2 Singularity analysis: The wavelet transform

In this section we present the appropriate techniques to deal with multifractal measures, and especially with those adapted to natural images. To check if μ defines a multifractal measure and to obtain the local exponent $h(\vec{x})$, we need to analyze the dependence of $\mu(B_r(\vec{x}))$ on the scale r. Unfortunately, direct logarithmic regression performed on eq. (5) yields rather coarse results when it is applied on discretized images, allowing only the detection of very sharp, well isolated features. It is then necessary to use more sophisticated methods. A convenient technique for this purpose [Mallat and Huang 1992, Arneodo et al. 1995, Arneodo et al. 1996] is the *wavelet transform* (see e.g. [Daubechies 1992]). Roughly speaking, the wavelet transform provides a way to interpolate the behavior of the measure μ over balls of radius r when r is a non-integer number of pixels. Besides, if the image is a discretization of a continuous signal, the wavelet analysis is the appropriate technique to retrieve the correct exponents, as it is explained later on.

The wavelet transform of a measure μ is a function that is not localized, neither in space nor in frequency. Its arguments are then the site $\vec{x}$ and the scale r. It also involves an appropriate, smooth interpolating function $\Psi(\vec{x})$. More precisely, the wavelet transform of μ is:

$$T_\Psi^r d\mu(\vec{x}) \equiv \int d\mu(\vec{x}') \, \Psi_r(\vec{x}-\vec{x}') = \frac{d\mu}{d\vec{x}} \otimes \Psi_r(\vec{x}) \ , \tag{9}$$

where $\Psi_r(\vec{x}) \equiv \frac{1}{r^d}\Psi(\frac{\vec{x}}{r})$. The important point here is that for the case of a positive measure it can be proved that:

$$T_\Psi^r d\mu(\vec{x}_0) \sim r^{h(\vec{x}_0)} \tag{10}$$

if and only if μ verifies eq.(5) (h being exactly the same) and Ψ decreases fast enough (see [Arneodo et al. 1996] and references therein). In that way, this property allows us to extract the singularities directly from $T_\Psi^r d\mu$. It is very important to remark that Ψ can be a positive function, what makes an essential difference with respect to the analysis of multiaffine functions (see section 3.4).[5]

In general the ideal, continuous measure $d\mu(\vec{x})$ is unknown. In these cases the data is given by a discretized sampling, which in our case is a collection of pixels. One can reasonably argue that the pixel is the result of the convolution of an ideal signal with a compact support function $\chi_a(\vec{x})$ (describing, e.g., the resolution $\sim a$ of the optical device). In this case, under convenient hypothesis, the wavelet transform yields the correct projection of the ideal signal. Let us call $\vec{x}_{(n_1,n_2)} \equiv \vec{x}_{\vec{n}}$ the positions of the pixels in the discretized sample and $C_{\vec{n}}$ the values of the discretized version of the contrast; then

$$C_{\vec{n}} = T_\chi^a C(\vec{x}_n) \equiv C \otimes \chi_a(\vec{x}_{\vec{n}}) \ , \tag{11}$$

Taking the sample $\{\vec{x}_{\vec{n}}\}$ in such a way that the centers of adjacent pixels are at a distance larger than a, it is verified that:

[5]This means that the exponents can be obtained with a *non-admissible* wavelet Ψ, i.e. a function with a non-zero mean

$$\nabla C_{\vec{n}} \equiv \Big(C_{(n_1+1,n_2)} - C_{(n_1,n2)} \ , \ C_{(n_1,n_2+1)} - C_{(n_1,n2)} \Big) \sim \nabla C \otimes \chi_a(\vec{x}_{\vec{n}}) \tag{12}$$

and so the discretized version $\mu^{(dis)}$ of the ideal measure μ, can be considered as a wavelet coefficient of the continuous μ with the wavelet χ at the scale a, namely

$$d\mu_{\vec{n}}^{(dis)} \sim d\vec{x} \ \ |\nabla C| \otimes \chi_a(\vec{x}_{\vec{n}}) \ = \ d\vec{x} \ \ T_\chi^a d\mu(\vec{x}_{\vec{n}}) \ . \tag{13}$$

Taking into account that the convolution is associative, the discretized wavelet analysis of $d\mu_{\vec{n}}^{(dis)}$ with a wavelet Ψ can be simply expressed as:

$$T_\Psi^r d\mu^{(dis)}(\vec{x}_{\vec{n}}) \sim T_{\Psi \otimes \chi_{\frac{a}{r}}}^r d\mu(\vec{x}_{\vec{n}}) \tag{14}$$

that is, analyzing the discretized measure $\mu^{(dis)}$ with a wavelet Ψ is equivalent to analyze the continuous measure μ with a wavelet $\Psi \otimes \chi_{\frac{a}{r}}$. But if the scale r is large enough compared to the photoreceptor extent a, $\chi_{\frac{a}{r}}(\vec{x}) \sim \delta(\vec{x})$, which does not depend on r. Thus, if the internal size a is small enough to allow fine detection, we can recover the correct exponent at the point $\vec{x}_{\vec{n}}$ by just analyzing the discretized measure. Conversely, the extent a imposes a lower cut-off in the details that can be observed in any image constructed by pixels: as it is just the size of the pixel, the estimation of the singularity exponent at a given point requires to perform the analysis over scales containing several pixels.

This analysis is similar to that performed in [Mallat and Zhong 1991, Mallat and Huang 1992]: any image is analyzed by means of its scaling properties under wavelet projection. There are however important differences between the study performed by these authors and the present work. The first difference concerns the basic scalar field to be analyzed: Mallat and Zhong considered $C(\vec{x})$ instead of $|\nabla C|(\vec{x})$, which we prefer because, as it will be shown later on, it is the μ-density of a multifractal measure μ (see Section 4). Another difference is methodological: in this paper we intend to classify the points according its singularity exponent to form every fractal component, while the previous papers were devoted to obtain only the wavelet transform modulus maxima, a concept related but different from what we will call the most singular component, that is, only one (although the most important) of the fractal components. The third difference concerns motivation: we want to classify every point with respect to a hierarchical scheme, while these authors search a wavelet-based optimal codification algorithm, in the sense of providing an (almost) perfect reconstruction algorithm from the wavelet maxima.

3 Statistical self-similarity of natural images

In this section we introduce the basic concepts of the statistical approach, defining at the same time new variables which are closer to the wavelet analysis.

3.1 SS and ESS

From the statistical point of view, μ_r is a random variable, in the sense that given a random point $\vec{x}$ in an arbitrary image $\mathcal{I}$ the value of $\mu_r^{\mathcal{I}}(\vec{x})$ cannot be deterministically predicted.

By definition, the multifractal structure is a multiscaling effect. This multiscaling should be somehow apparent from the dependence of the probability density of μ_r on the scale parameter r. To determine the whole probability density function of μ_r requires a large dataset, especially for the rare events; besides its dependence on r could mix multiscaling effects in a complicated way. On the contrary, the p-moments of μ_r (that is, the expectation values of μ_r raised to the p-th power) suffice to characterize completely the probability density,[6] and given that they are averages of powers of μ_r they are likely to depend on r in a simple way, somehow related to the behavior of the measure in eq. (5).

It is more convenient to deal with the p-moments of normalized versions of μ_r such as the Edge Content $\epsilon_r \equiv \frac{1}{r^2}\mu_r$ or the wavelet projections $|T_\Psi^r d\mu|$. They are normalized in such a way that the first order moments, $\langle\epsilon_r\rangle$ and $\langle|T_\Psi^r d\mu|\rangle$, do not depend on r: the trivial factor r^{pd} is hence removed from the random variable.

As we will see later on, the moments $\langle\epsilon_r^p\rangle$ and $\langle|T_\Psi^r d\mu|^p\rangle$ exhibit the remarkable scaling properties of Self-Similarity (SS):

$$\langle\epsilon_r^p\rangle \;=\; \alpha_p\, r^{\tau_p} \tag{15}$$

(analogously for $|T_\Psi^r d\mu|$ with exponents τ_p^Ψ) and of Extended Self-Similarity (ESS) (referred to the moment of order 2) [Benzi et al 1993b, Benzi et al 1993c, Benzi et al 1995]:

[6] Provided they do not diverge too fast with p.

$$\langle \epsilon_r^p \rangle \ = \ A(p,2) \left[\langle \epsilon_r^2 \rangle \right]^{\rho(p,2)} \quad , \tag{16}$$

(and similarly for $|T_\Psi^r d\mu|$ with exponents $\rho^\Psi(p,2)$). ESS has been referred to the moment of order 2, but eq (16) trivially implies that any moment can be expressed as a power of the moment of order q with an exponent $\rho(p,q) = \frac{\rho(p,2)}{\rho(q,2)}$.

In this way, we have the sets of SS exponents τ_p and ESS exponents $\rho(p,2)$ for ϵ_r and similarly the exponents τ_p^Ψ and $\rho^\Psi(p,2)$ for $T_\Psi^r d\mu$. Notice that if SS holds also does ESS, in which case there is a simple relation between τ_p and $\rho(p,2)$:

$$\tau_p \ = \ \tau_2 \, \rho(p,2) \ \ . \tag{17}$$

The actual dependence of τ_p and $\rho(p,2)$ on p determines the fractal structure of the system. A trivial dependence $\tau_p \propto p$ would reveal a monofractal structure. A different dependence on p indicates the existence of a more complicated (multifractal) geometrical structure.

3.2 The multiplicative process

A random variable subtending an area of linear size r and which possesses SS (eq. (15)) and ESS (eq. (16)) can be described statistically by means of a multiplicative process [Benzi et al 1993a, Novikov 1994]. The statistical formulation is rather simple: given two different scales r and L such that $r < L$, there is a simple stochastic relation between the Edge Contents at these two scales :

$$\epsilon_r \ \doteq \ \alpha_{rL} \, \epsilon_L \tag{18}$$

where α_{rL} is a random variable, independent of the Edge Content, which describes how the change between these two scales takes place. The symbol "$\doteq$" indicates that both sides of the equation are distributed in the same way, but this does not necessarily imply the validity of the relation at every pixel of a given image. The random variables α_{rL} between all possible pairs of scales r and L are said to define a *multiplicative process*.

Let us understand in a more intuitive way the meaning of this process. It is rather clear that given the distribution of ϵ_L at a fixed large scale L, to compute the distribution of the Edge Content ϵ_r it is enough to know the distribution of α_{rL}. But let us now introduce an intermediate scale r'. One could also obtain first $\epsilon_{r'}$ using $\alpha_{r'L}$ and then ϵ_r by means of $\alpha_{rr'}$. Thus, the multiplicative process must verify the cascade relation:

$$\alpha_{rL} \;=\; \alpha_{rr'}\,\alpha_{r'L} \quad , \tag{19}$$

that tells us that the variables α_{rL} are *infinitely divisible*. Infinitely divisible stochastic variables are then completely characterized by their behavior under infinitely small changes in the scale, that is, by $\alpha_{r,r+dr}$. The simplest example of an infinitely divisible random variable α_{rL} is the one having a binomial infinitesimal distribution, namely:

$$\alpha_{r,r+dr} \;=\; \begin{cases} 1-\Delta\frac{dr}{r} & , \quad \text{with probability } 1-[d-D_\infty]\frac{dr}{r} \\ \\ \beta(1-\Delta\frac{dr}{r}) & , \quad \text{with probability } [d-D_\infty]\frac{dr}{r} \end{cases} \tag{20}$$

The form of this process is determined by its compatibility with eq. (19); see for instance [She and Leveque 1994, She and Waymire 1994, Castaing 1996] for a detailed study. Moreover, there are only two free parameters: it is verified that $d-D_\infty = \Delta/(1-\beta)$. They have also to satisfy the following bounds: $0<\beta<1$ and $0<\Delta<1$.

The infinitesimal process (20) can be easily interpreted in terms of the Edge Content and the measure density:

- **Probability** $=1-[d-D_\infty]\frac{dr}{r}$ **:** this is the most likely situation. In this case the Edge Content (and consequently the measure) changes smoothly under small changes in scale: $\alpha_{r,r+dr}$ is close to one and its variation is proportional to $d\ln r$.

- **Probability** $=[d-D_\infty]\frac{dr}{r}$ **:** in this unlikely case, under infinitesimal changes in scale the Edge Content undergoes finite variations, which is reflected in the fact that $\alpha_{r,r+dr}$ deviates from one by a factor β. The μ-density $|\nabla C|$ must then be divergent somewhere along the boundary of the ball of radius r, and the parameter β is a measure of how sharp this divergence is.

The process for non-infinitesimal changes in scale can be easily derived from the infinitesimal process presented in eq (20). In this case the random variable α_{rL} follows a Log-Poisson process [She and Waymire 1994]; its probability distribution $\rho_{\alpha_{rL}}(\alpha_{rL})$ has the form:

$$\rho_{\alpha_{rL}}(\alpha_{rL}) \;=\; \left[\frac{r}{L}\right]^{d-D_\infty} \sum_{n=0}^{\infty} \frac{(d-D_\infty)^n}{n!} \left[\ln\frac{L}{r}\right]^n \delta(\alpha_{rL} - \beta^n \left(\frac{r}{L}\right)^{-\Delta}) \quad (21)$$

If a multiplicative process is realised the Edge Content has the SS and the ESS properties, eqs. (15) and (16). Thus, knowledge of the ESS exponents $\rho(p,2)$ and of τ_2 is enough to compute τ_p using eq. (17). For the Log-Poisson process the ESS exponents are:

$$\rho(p,2) \;=\; \frac{p}{1-\beta} \;-\; \frac{1-\beta^p}{(1-\beta)^2} \quad (22)$$

which depend only on β. This process was proposed in [She and Leveque 1994] for turbulent flows, so we will refer to it either as the She-Leveque (S-L) or as the Log-Poisson model. The SS exponents are:

$$\tau_p \;=\; -\Delta p \;+\; (d-D_\infty)(1-\beta^p) \quad (23)$$

which compared with eq. (17) gives the following set of relations :

$$\left\{ \begin{array}{ccccc} \Delta & = & -\frac{\tau_2}{1-\beta} & & \\ d - D_\infty & = & -\frac{\tau_2}{(1-\beta)^2} & = & \frac{\Delta}{1-\beta} \end{array} \right. \quad (24)$$

This means that the model has only two free parameters, that can be chosen to be τ_2 and β or Δ and D_∞, for instance. The geometrical interpretation of the last two is very interesting [Turiel, Mato, Parga and Nadal 1998], and will be explained in the next subsection. For the time being let us notice that the maximum value of the EC, $\parallel \epsilon_r \parallel_\infty$, also follows a power law with exponent $-\Delta$,

$$\parallel \epsilon_r \parallel_\infty = \; \alpha_\infty \, r^{-\Delta} \quad (25)$$

and thus the parameter Δ characterizes the most divergent behavior present in natural images.

3.3 SS and multifractality

Let us now analyze in more detail the geometrical meaning of the model and its relation with the singularities described in Section 2. For a multifractal measure μ with singularity spectrum $D(h)$ there is an important relation between the SS exponents τ_p and $D(h)$. Let us denote by $\rho_h(h)$ the probability density of the distribution of the singularity exponents h in the image. Then, partitioning the image pixels according to their values of h, eq. (6) can be expressed, for r small enough, as:

$$\langle \epsilon_r^p \rangle \sim \int dh\, \rho_h(h) \langle \alpha^p \rangle_{F_h} r^{hp+d-D(h)} \quad . \tag{26}$$

The factor $r^{d-D(h)}$ comes from the probability of a randomly chosen pixel being in the fractal F_h (see eq. (8)) and guarantees the correct normalization of $\langle \alpha^p \rangle_{F_h}$ (see e.g. [Frisch 1995, Arneodo et al. 1996] for details). For very small r the application of the saddle point method to eq. (26) yields:

$$\langle \epsilon_r^p \rangle = \alpha_p r^{\tau_p} \propto r^{\min_h \{ph+d-D(h)\}} \tag{27}$$

This gives a very interesting relation between τ_p and $D(h)$: τ_p is the Legendre transform of $D(h)$. The important point is that $D(h)$ is easily expressed in terms of τ_p by means of another Legendre transform, namely:

$$D(h) = \min_p \{ph + d - \tau_p\} \quad . \tag{28}$$

If one has a model for the τ_p's (the S-L model, eq. (23)) the dimension spectrum can be predicted:

$$D(h) = D_\infty - \frac{h+\Delta}{\ln\beta}\left[1 - \ln\left(-\frac{h+\Delta}{(d-D_\infty)\ln\beta}\right)\right] \quad . \tag{29}$$

This is represented in Fig. 8. There are natural cut-offs for the dimension spectrum; in particular, there cannot be any exponent h below the minimal value $h_\infty = -\Delta$. Thus, Δ defines the most singular of the fractal components, which has fractal dimension $D(h_\infty) = D_\infty$. We could use these properties of the most singular fractal (its dimension and its associated exponent) to characterize the whole dimension spectrum, which is in agreement with the arguments given in section 3.2.[7]

3.4 Multiaffinity

Related to multifractality there exists the simpler but more unstable property of multiaffinity [Benzi et al 1993a]. This is a characterization of chaotic, irregular scalar functions that somewhat generalizes the concepts of continuity and differentiability.

[7] Remember that β can be expressed in terms of Δ and D_∞: $\beta = 1 - \Delta/(d - D_\infty)$

First, let us give the concept of *Hölder exponent*: A scalar function $F(\vec{x})$ is said to be Hölder of exponent $h_F(\vec{x}_0)$ at a given point $\vec{x}_0$ if for any point $\vec{y}$ close enough to $\vec{x}$ the following inequality holds:

$$|F(\vec{y}) - F(\vec{x}_0)| \; < \; A_0 |\vec{y} - \vec{x}_0|^{h_F(\vec{x}_0)} \quad , \tag{30}$$

A_0 being a constant depending on the point $\vec{x}_0$. We define the Hölder exponent of $F(\vec{x})$ at $\vec{x}_0$ as the maximum of the exponents $h_F(\vec{x}_0)$ verifying eq. (30). Defining the Linear Increment (LI) of $F(\vec{x})$ by a displacement vector $\vec{r}$ as $\delta_{\vec{r}} F(\vec{x}) \equiv |F(\vec{x}+\vec{r}) - F(\vec{x})|$, if the function F has a Hölder exponent $h_F(\vec{x})$ at $\vec{x}$ then:

$$\delta_{\vec{r}} F(\vec{x}) \; \sim \; \alpha_F(\vec{x}) r^{h_F(\vec{x})} \tag{31}$$

which is similar to eq. (6). In the same spirit, we say that a function is *multiaffine* if for every point $\vec{x}$ eq. (31) is verified.

Multiaffinity is a more intuitive concept than multifractality, because the Hölder exponents are good characterizations of Taylor-like local expansions of the function $F(\vec{x})$. For instance, $h_F(\vec{x}) > 0$ means that the function is continuous at $\vec{x}$, $h_F(\vec{x}) > 1$ implies that the function has a continuous first derivative and so on. It also works in the other sense: a function having exponent $h_F = -1$ behaves as "bad" as a δ-function (See the Appendix for a brief tutorial about Hölder exponents of simple functions). Even more: a function F has Hölder exponent h_F at a given point if and only if its first order derivatives have Hölder exponent $h_F - 1$ at the same point.

Besides, multiaffinity implies multifractality in the following sense: given a measure μ, if its density $\frac{d\mu}{d\vec{x}}$ is a multiaffine function then μ is a multifractal measure having exactly the same exponents as $\frac{d\mu}{d\vec{x}}$ at every point.[8] For natural images and taking $F = |\nabla C|$, this fact can be expressed as:

$$\epsilon_r \; \sim \; T_\Psi^r d\mu \; \sim \; \delta_{\vec{r}} |\nabla C| \tag{32}$$

in the sense that the three variables have the same statistical dependence on r. This relation can be also expressed in terms of the SS exponents. Let us denote by τ_p the SS exponents of ϵ_r, by τ_p^Ψ those of $T_\Psi^r d\mu$ and by $\tau_p^{\nabla C}$ those of $\delta_{\vec{r}} |\nabla C|$. Then the statistical relation in eq. (32) implies that:

[8] Notice that there is a shift of size d in the definition of the exponents of μ (eq. (5)) which is not present in the definition of the Hölder exponents of a multiaffine function (eq. (30)).

$$\tau_p \;=\; \tau_p^{\Psi} \;=\; \tau_p^{\nabla C} \tag{33}$$

We are also interested in the relation between τ_p and the SS exponents τ_p^C of the moments of $\delta_{\vec{r}}C$. If C exhibits the multiaffine behavior shown in eq. (31), recalling the meaning in terms of differentiability of the Hölder exponents, we have that $\delta_{\vec{r}}|\nabla C|(\vec{x})$ verifies:

$$\delta_{\vec{r}}|\nabla C|(\vec{x}) \;\sim\; \alpha_{|\nabla C|}(\vec{x})\; r^{h_C(\vec{x})-1} \tag{34}$$

This can be statistically interpreted as:

$$\epsilon_r \;\sim\; \delta_{\vec{r}}|\nabla C| \;\sim\; \frac{1}{r}\,\delta_{\vec{r}}C \tag{35}$$

which reflects the shift by -1 in the exponents of the derivative of C.[9] In terms of the SS exponents, this implies that:

$$\tau_p^C \;=\; \tau_p^{\nabla C} + p \;=\; \tau_p + p \tag{36}$$

Let us remark that the singularity spectra associated to the variables $\delta_{\vec{r}}C$, $\delta_{\vec{r}}|\nabla C|$ and $T_\Psi^r d\mu$ are the Legendre transforms of the SS exponents τ_p^C, $\tau_p^{\nabla C}$ and τ_p^{Ψ}, respectively (a fact that can be derived in the same way as eq. (28)).

There still remains the converse question: if multifractality implies multiaffinity. Unfortunately, this is not true (see [Daubechies 1992]). In fact, the $|F(\vec{y})-F(\vec{x}_0)|$ term in eq. (30) suggests the existence of a pseudo-Taylor expansion, in the sense that:

$$F(\vec{y}) \;\approx\; p_n(\vec{y}-\vec{x}_0) \;+\; A_0\,|\vec{y}-\vec{x}_0|^{h_F(\vec{x}_0)} \tag{37}$$

where $p_n(\vec{y}-\vec{x}_0)$ is a polynomial of degree n in the projection of the vector $\vec{y}-\vec{x}_0$ along its direction. We should thus generalize the concept of Hölder exponent: given a point $\vec{x}_0$, we will say that this point possesses a Hölder exponent $n < h_F(\vec{x}_0) < n+1$ (n integer) if there exists a polynomial $p_n(\vec{x})$ of degree n and a constant A_0 such that

$$|F(\vec{y}) - p_n(\vec{y}-\vec{x}_0)| \;<\; A_0\,|\vec{y}-\vec{x}_0|^{h_F(\vec{x}_0)} \tag{38}$$

and $h_F(\vec{x}_0)$ is the maximum of the exponents verifying such relation.

[9] Let us note that the relation $r\epsilon_r \;\sim\; \delta_{\vec{r}}C$ is the analog of the Kolmogorov hypothesis of local similarity [Arneodo et al. 1996]

Multifractal measures with no multiaffine densities (as defined in eq. (30)) have been studied in [Bacry et al. 1993, Arneodo et al. 1996]. The problem can be explained by the masking presence of the regular, polynomial part which stands for global, non-localized effects in the signal field $F(\vec{y})$. To remove this part, those authors propose a generalization of the Linear Increment. Instead of Linear Increments of the function, wavelet transforms of the appropriate one-dimensional restriction are considered, taking the scalar field as a real-valued measure density. This wavelet transform of the function $F(\vec{x})$ is defined as

$$T_\Psi^{\vec{r}} F(\vec{x}_0) \;\equiv\; \int ds\; F(\vec{x}_0 + s\frac{\vec{r}}{r})\Psi_r(s)\;, \tag{39}$$

where Ψ is a real, one-dimensional analyzing wavelet. Since the integral is one-dimensional the wavelet at the scale r is $\Psi_r(s) \equiv \frac{1}{r}\Psi(\frac{s}{r})$. If the scale r is small enough, the integral is dominated by the first terms in the pseudo-Taylor expansion of $F(\vec{x})$ around $\vec{x}_0$, eq. (37):

$$T_\Psi^{\vec{r}} F(\vec{x}_0) \;\approx\; \int ds\; p_n(s)\Psi_r(s) \;+\; A_0 \int ds\; |s|^{h_F(\vec{x}_0)}\Psi_r(s) \tag{40}$$

If the analyzing wavelet is now required to vanish at least the first n integer moments, the singular part appears as the main contribution to the wavelet transform. (For later reference, let us say now that a wavelet with a vanishing zero-th order moment – i.e., its integral is null – is called an *admisible wavelet*). Hence, changing variables $t = \frac{s}{r}$ and defining $A = A_0 \int dt\; |t|^{h_F(\vec{x}_0)}\Psi(t)$, one has:

$$T_\Psi^{\vec{r}} F(\vec{x}_0) \;\approx\; A\, r^{h_F(\vec{x}_0)} \tag{41}$$

where all the dependence on r stands in the prefactor $r^{h_F(\vec{x}_0)}$. In this way, the Hölder exponent (in the sense of eq. (38)) can be obtained by a wavelet projection, provided the wavelet has zero moments up to a large enough order. Since $h_F(\vec{x}_0)$ is non-integer, this procedure cannot cancel the singular part even if the order n of the largest vanishing moment of Ψ is larger than $h_F(\vec{x}_0)$. On the other hand, the wavelet could not be able to detect singularity exponents $h_F > n+1$. This is because a polynomial of order $n+1$ could still be present and dominate the contribution of the singular term. For this reason, it is important to determine the right order of the required zero moments of the wavelet.

This is then the appropriate scheme to detect the Hölder exponents of signals with a regular part. In cases in which the μ-density $|\nabla C|$ and/or its primitive C are not multiaffine, the concept of multiaffinity (eq. (30)) should be applied to their wavelet projections. It is then necessary to determine how many zero moments the wavelet should have, as it provides information on the global aspects of the signal. This generalization is straightforward, replacing in all the cases $\delta_{\vec{r}}C$ by $T_\Psi^{\vec{r}}C$ and $\delta_{\vec{r}}|\nabla C|$ by $T_\Psi^{\vec{r}}|\nabla C|$. For instance, if the wavelet transforms of C and $|\nabla C|$ verify eq. (41) for certain A and $h_F(\vec{x})$, the same arguments that led us to eqs. (35) and (36) can be used to derive that :

$$\epsilon_r \sim T_\Psi^{\vec{r}}|\nabla C| \sim \frac{1}{r}\, T_\Psi^{\vec{r}}C \qquad (42)$$

which is expressed in terms of the SS exponents $\tau_p^{\Psi_C}$ (of $T_\Psi^{\vec{r}}C$) and $\tau_p^{\Psi_{\nabla C}}$ (of $T_\Psi^{\vec{r}}|\nabla C|$) as:

$$\tau_p^{\Psi_C} = \tau_p^{\Psi_{\nabla C}} + p = \tau_p + p \qquad (43)$$

Let us make a final remark: it is possible that the wavelet projections of a given scalar function F verifies the scaling in eq. (41) without verifying eq. (38). This can happen, for instance, for functions with such an irregualar behavior that the pseudo-Taylor expansion is not possible. In those cases, $T_\Psi^{\vec{r}}F$ can be used as a stronger generalization of the usual Linear Iincrement. As it will be shown in Section 4.4), this is the case for $|\nabla C|$.

4 Numerical analysis of images

4.1 Databases

We have used two different image datasets: one from Daniel Ruderman (see [Ruderman 1994]), consisting in 45 B/W images of 256×256 pixels with up to 13 bits in luminosity depth (from now on this set will be referred to as the first dataset); and another containing 25 images taken from Hans van Hateren's database (see [van Hateren and van der Schaaf 1998]), having 1536×1024 pixels and 16 bits in luminosity depth (we will refer to these images as the second dataset). These 25 images are enumerated in table 1

The first dataset has been used for the statistical analysis of the one-dimensional wavelet transform, eq. (39), while the second dataset was used for the more demanding task (from the statistics point of view) of computing the moments of the two-dimensional variable in eq. (1). No matter the ensemble considered, however, the studied properties seem to behave in essentially the same way.

imk00034.imc	imk00801.imc	imk01164.imc	imk02649.imc	imk03807.imc
imk00211.imc	imk00808.imc	imk01406.imc	imk03134.imc	imk03842.imc
imk00478.imc	imk00881.imc	imk02035.imc	imk03514.imc	imk03940.imc
imk00586.imc	imk01017.imc	imk02603.imc	imk03536.imc	imk04042.imc
imk00605.imc	imk01032.imc	imk02626.imc	imk03789.imc	imk04069.imc

Table 1: The 25 images from van Hateren's database used in this study. They can be observed and downloaded from the URL address http://hlab.phys.rug.nl/archive.html

4.2 Multifractality of the measure and singularity analysis

We first address the issue of the multifractality of the measure μ defined in eq. (2). Our numerical analysis gives an affirmative answer to this question: the wavelet transform $T_\Psi^r d\mu(\vec{x}_0)$ is well fitted by $\sim r^{h(\vec{x}_0)}$ in about 98% of the pixels $\vec{x}_0$ in each image of our two datasets. The analysis was done using several wavelets, although the most convenient family of one parameter functions was $\Psi^\gamma(\vec{x}) = \frac{1}{(1+|\vec{x}|^2)^\gamma}$ $(\gamma \geq 1)$. These are positive (non-admissible) wavelets. The singularity exponent of the measure at a given pixel, $h(\vec{x})$, was obtained from a regression on eq. (10). The value of the scale r was typically taken in the range $(0.5, 2.0)$ in pixel units. Notice that for scales smaller than $r \sim 0.5$ the method would detect an artificial "edge" that actually corresponds to the pixel boundaries, what corresponds to $h = -1$. Similarly, as the scale becomes large the wavelet can only discriminate between very sharp edges and, for r of the order of the size of the image it would yield the value $h = 0$ everywhere.

To obtain the fractal components F_h, the pixels of a given image were classified according to the value of their exponents. The exponents lie in the interval $(-0.5, 1.4)$: only a negligible fraction of pixels have exponents less than -0.5 (about 0.8% of them) and none with a singularity below -1. It is also observed that there are no exponents larger than 1.4. This range of possible exponent values coincides exactly with that predicted in [Turiel, Mato, Parga and Nadal 1998], and is in agreemente with the model explained in section 3.3.

The most singular component is the one with $h = -0.5$, defined within a window of size Δh. The visualization of this set is very instructive (see Fig. 1). Clearly, the points appear to be associated with the contour of the objects present in the image. This is in agreement with our expectation that the behavior of the measure at these edges should be singular, and as much as possible. But there is something rather surprising: for a sharp, theta-like contrast, with independent values at both sides of the discontinuity, one would have expected an exponent -1, which appears as a Dirac δ-function in the density measure ∇C (see example number 4 in the Appendix). Our interpretation of the minimal value of h, $h_\infty = -0.5$, is that it reflects the existence of correlations among different fractal components of the images that smoothen the singularities (in particular those of the sharpest edges).

It is also interesting to observe the fractal components with exponents just above the most singular one. Fig. 2 shows the two next components, defined with $\Delta h = 0.05$. A comparison with Fig. 1 shows that pixels in these components are close to pixels in the most singular one. It is also rather clear that the probability that a random ball of size r contains a pixel from the component F_h increases with h, at least for small values of this exponent. Since this probability scales as shown in eq. (8), this behavior reveals an increase with h of the dimensionality of the fractal components. It is also observed (although not shown in the figures) that the dimensionality starts decreasing beyond $h \approx 0.2$. This is in agreement with the arguments of section 3.3: the singularity spectrum $D(h)$ of the S-L model, which is shown in Fig. 8, reflects this behavior of the fractal components.

The analysis done in this subsection justifies the arguments presented in Section 1. One cannot consider the sharpest edges of an image independently of the other texture structures (i.e. the other fractal components). Luminosity jumps from one roughly constant value to another (statistically independent) constant value are not typical in natural images. The sharpest edges appear correlated with smoother textures, an effect that tends to decrease the strentgh of their singularity exponent, reaching the value $h \sim -0.5$. Standard edge-detection tecniques [Gonzalez and Woods 1992] have not been designed to search for singularities of this strength. Good edge-detection algorithms should be specifically constructed to capture singularities close to h_∞.

4.3 Statistical analysis of the measure

We computed the p-moments of three different variables related to the measure:

- **Bidimensional Edge Content (ϵ_r):** Following the definition given in eq. (1). This analysis was done using the 25 images of the second dataset. As the variable itself is positive, we computed directly the quantities $\langle\epsilon_r^p\rangle$.

- **Two one-dimensional restrictions of the wavelet projection of the measure**[10] **($T_\Psi^r d\mu$):** We considered horizontal (h) and vertical (v) restrictions of the measure $d\mu$, that is $d\mu^h = dx|\frac{\partial\mu}{\partial x}|$ and $d\mu^v = dy|\frac{\partial\mu}{\partial y}|$. The appropriate wavelet transforms are denoted by $T_\Psi^r d\mu^l$ ($l = h, v$). As these coefficients need not to be positive (in general Ψ is not positive), we computed the moments of their absolute values, $\langle|T_\Psi^r d\mu^l|^p\rangle$.

It is remarkable that the bi-dimensional Edge Content and the two one-dimensional wavelet projections exhibit the scaling properties of SS (eq. 15) and of ESS (eq. 16) (the ESS tests for the three variables are presented in Fig. 3 and 4). This was also observed in [Turiel, Mato, Parga and Nadal 1998], although for somewhat different variables: the one-dimensional horizontal and vertical restrictions of the local edge variance. Surprisingly, no matter the particular variable, the exponents τ_p and $\rho(p,2)$ obtained are very similar (see Fig. 5 and those in [Turiel, Mato, Parga and Nadal 1998]). Therefore, they should refer to very essential, robust aspects of images.

The S-L model fits very well the experimental exponents (see Fig. 5). It seems then that the multifractal structure underlying the statistical description could be described with the Log-Poisson process, which means that under infinitesimal changes in the scale there are only two possible different types of transformations.

It is really remarkable that for our datasets $\Delta \approx 0.5$ and $D_\infty \approx 1$. This means that the most singular component is a collection of lines (the sharp edges present in the image) which can be characterized by their common exponent $-\Delta = -0.5$. But this is precisely the smallest value we observed previously in our wavelet analysis. The statistical approach confirms the result obtained by a local singularity analysis.

The non-linear dependence of $\rho(p,2)$ on p again indicates that the Edge Measure is multifractal. This has to be contrasted with the results by [Ruderman and Bialek 1994, Ruderman 1994] who find that the log-contrast distribution follows a simple classical scaling. [11] It also holds for any derived

[10] Since the calculations involving $T_\Psi^r d\mu$ are highly computer-time consuming we did not consider the bi-dimensional wavelet transform. FFT cannot be safely used due to aliasing, which heavily changes the tails (rare events) of the distribution.

[11] Other luminosity analysis performed by D. Ruderman provided some evidence of multiscaling behaviour (private communication).

variable as the gradient (see [Ruderman 1994] for details). This particular scaling is related, in that context, with general properties of scale invariance of natural images. The Edge Content and the wavelet projections, being local averages of a gradient, are able to keep the information about the more complex underlying multifractal structure.

The same authors have also designed an iterative procedure that tends to decompose the image into a gaussian-like piece (local fluctuations of the log-contrast) and an image-like piece (local variances). Such a decomposition would suggest the existence of a multiplicative process for the log-contrast itself where the random variable relating the image at two different levels of the hierarchy is given by the local fluctuations. A closer look to the iterative procedure indicates that this is not the case: independence between the two factors of the process is not guaranteed and also the process is not infinitely divisible.

4.4 Multiaffinity properties of C and $|\nabla C|$

In this section we study the scaling properties for the Linear Increments of the contrast $C(\vec{x})$ and of the measure density $|\nabla C|(\vec{x})$. Our aim is then to check if C and $|\nabla C|$ satisfy the scaling defined in eq. (31). Recalling the connection (the analog of eq. (26) for these two quantities) between local scaling (described by the singularity exponents) and statistical scaling (characterized by the SS exponents), one concludes that the validity of SS implies that multiaffinity holds. We have then done a statistical analysis of the Linear Increments of the contrast and the measure density.

We first computed the p-moments of the Linear Iincrements of both variables, looking for SS and ESS scalings. Let us notice that there is no *a priori* argument to expect that $\delta_{\vec{r}} C(\vec{x})$ shows SS. On the other hand, since the measure μ is multifractal (as it was shown in section 4.2) it would be reasonable to expect that $|\nabla C|(\vec{x})$ has a multiaffine behavior. However, this has to be verified explicitly because the multifractality of the measure does not guarantee the multiaffinity of its density, as it was discussed in section 3.4. In fact, our numerical analysis showed that neither $C(\vec{x})$ nor $|\nabla C|(\vec{x})$ have SS. The same is true for ESS, as it is shown in Figs. 6a and 7, where the moments exhibit a rather erratic behavior.

It seems that the failure of SS for $C(\vec{x})$ and $|\nabla C|(\vec{x})$ has to be explained in different terms. In the case of $|\nabla C|(\vec{x})$, this quantity is the measure density and, as it was stated in eq. (10), its convolution with a non-admissible (i.e. positive) wavelet has the same singularities as the measure, eq. (2).

This implies that the lack of SS cannot be explained in terms of a polynomial part as described in eq. (37). In fact, a positive wavelet cannot alter the contribution from this polynomial while one observes that the data for $\delta_{\vec{r}}|\nabla C|(\vec{x})$ (Fig. 7) are different from those of $T_\Psi^r d\mu^h$ and $T_\Psi^r d\mu^v$ (Fig. 4).

On the other hand it is plausible that the failure of SS of $\delta_{\vec{r}} C(\vec{x})$ can be attributed to the presence of a polynomial part in $C(\vec{x})$. Differently to what happened with $|\nabla C|(\vec{x})$, when the contrast is convolved with a non-admissible wavelet, SS and ESS are still not present (that is, one observes a similar erratic behavior as in Fig. 6a). The same behavior is observed when $C(\vec{x})$ is convolved with an admissible wavelet with a zero mean. The situation changes drastically when a wavelet with vanishing zero- and first-order moments is used. The result obtained using the second derivative of a gaussian is presented in Fig. 6b, which clearly shows ESS (it also exhibits SS). This can be understood recalling that the possible values of the Hölder exponent of $C(\vec{x})$ are contained in the interval $(0.5, 2.4)$ (see eq. (34) and section 4.2). A non-admissible wavelet truncates the detected singularities at the value $h = 0$, an admissible one truncates them at $h = 1$. The third wavelet considered, with two vanishing moments, truncates the singularities only from $h = 2$ and then it covers almost the whole interval (In theory a wavelet with another vanishing moment would be more appropriate, however in practice it is numerically more unstable because the resolution of the wavelet is smaller).

Finally, we also checked that the experimental values of τ_p, $\tau_p^{\Psi_C}$ and $\tau_p^{\Psi_{\nabla C}}$ satisfy the relation given in eq. (43).

5 Discussion

We have shown that in natural scenes pixels are arranged in non-overlapping classes, the fractal components of the image, characterized by the behavior of the contrast gradient under changes in the scale. In each of these classes this quantity exhibits a distinct power law behavior under scale transformations, what indicates that there is no a well-defined scale for the isolated components but that at the same time the scene is not scale invariant globally.

This result implies that a given image is composed of many different spatial structures of the contrast associated to the fractal components, which can be arranged in a hierarchical way. In fact, how sharp or soft a change in contrast is at a given point can be quantified in terms of the value of the scaling exponent at that site. The smaller the exponent the sharper the local change in contrast is.

More precisely, we have dealt with two basic quantities: the contrast $C(\vec{x})$ and the edge measure μ, which density is the modulus of the contrast gradient $|\nabla C|(\vec{x})$ (see eq. (2)). Closely related to the measure, we can define the edge content of a region A of the image (eq. (3)), which quantifies how much the contrast gradient deviates from being uniformly distributed inside the region A. This is a suitable definition to study the local behavior of the contrast gradient. As the size of A is reduced, some contributions of the contrast gradient to the edge content are left outside this region. Under an infinitesimal change in the scale very often these contributions are small, but sometimes it does happen that a small change in the scale produces a large change in the edge content. This is due to singularities in the contrast gradient. As a consequence of this, the edge content (and the measure) exhibits large spatial fluctuations. The edge content can have different singularities at neighbouring points, characterized by a power-law behavior with exponents $h(\vec{x})$ which depend on the site, eq. (6). This means that the measure is multifractal, as it has been explicitly verified in this work.

The set of pixels with the same value of $h(\vec{x})$ defines a particular spatial structure of contrast: the fractal component F_h. The smallest $h(\vec{x})$ is the sharpest the contrast gradient at the pixel $\vec{x}$ is. Its minimal value was, for our dataset, $h_\infty = -0.5$. Another question refers to the fractal dimension D_∞ of this component. One finds $D_\infty = 1$ [Turiel, Mato, Parga and Nadal 1998], and when this component is extracted from the image one checks that it corresponds to the boundaries of the objects in the scene. The associated contrast structure is made of the sharpest edges of the image. The other, less singular components represent softer contrast structures inside the objects. It is also observed that there are strong spatial correlations between fractal components with similar exponents.

We have also shown how to extract the fractal components from the image. To do this one needs a suitable technique to compute the scale behavior at a given pixel. A wavelet projection [Mallat and Huang 1992, Arneodo et al. 1996] seems to be the correct way to perform this analysis, since it provides a way of interpolating scale values which are not an integer number of pixels. Our approach differs from the one followed in those references in several aspects, but in particular in that we have been here interested in the characterization of the whole set of fractal components.

We remark the important fact that the multifractal can be explained by a model with only two parameters (e.g. β and Δ). This had been noticed before in terms of a variable that integrates the square of the derivative of the contrast along a segment of size r, both taken either in the horizontal or in the vertical direction [Turiel, Mato, Parga and Nadal 1998] . This quantity admits the interpretation of a local linear edge variance. We have seen that the same model is valid for many other quantities, some of them more general and natural.

The basic idea of the model is that the contrast gradient has singularities that under an infinitesimal change in the size of A produce a substantial change in the edge content inside that region. The simplest stochastic process to assign a value to this change is a binomial distribution, eq (20): its more likely value is very small, but with small probability this change is given by the parameter β, which for our dataset is about 0.5. The other parameter of the model, Δ, measures the strength of the singularity. The geometrical locus of these singularities is the most singular fractal component. For a finite scale transformation this process becomes Log-Poisson. The model appeared before in other problems [She and Leveque 1994].

In a sense, the most singular component conveys a lot of information about the whole image. The two model parameters can be obtained by analyzing properties of this particular class of pixels: its dimension (D_∞) and the scale exponent of the edge content ($\Delta = -h_\infty$). In turn, these two parameters completely define the whole dimension spectrum in the Log-Poisson model. It is then very plausible that other components have a great deal of redundancy with respect to the most singular one.

This hierarchical representation of spatial structure can be used to predict specific feature detectors in the visual pathways. We conjecture that the most singular component contains highly informative pixels of the images that are responsible for the epigenetic development leading to the adaptation of receptive fields. Learning of the statistical regularities [Barlow 1961] present in this portion of visual scenes (using for instance the algorithm in [Bell and Sejnowski 1997]), would give relevant predictions about the structure of receptive fields of cells in the early visual system.

Acknowledgments

We are thankful to Jean-Pierre Nadal, Germán Mato and Dan Ruderman for useful comments and to Angel Nevado for many fruitful discussions that we had during the preparation of this work. We are grateful to Dan Ruderman and Hans van Hateren, some of whose images we used in the present statistical analysis. Antonio Turiel is financially supported by a FPI grant from the Comunidad Autónoma de Madrid, Spain. This work was funded by a Spanish grant PB96-0047.

A Characterization of the singularities of simple functions

It is instructive to compute the singularity exponents defined in eq. (30) for several simple functions.

1. **Continuous functions:** If a function $f(\vec{x})$ is continuous at a given point $\vec{x}$, the value of f at $\vec{y}$ should be rather close to $f(\vec{x})$ provided $\vec{y}$ is close to $\vec{x}$; then $|f(\vec{y})| < |f(\vec{x})| + \epsilon$ so $|f(\vec{x}) - f(\vec{y})| < 2|f(\vec{x})| + \epsilon = A$. That is, f is Hölder of exponent 0 at $\vec{x}$.

2. **Smooth functions with continuous, non-vanishing first derivative:** Using the Taylor expansion, $f(\vec{x}) - f(\vec{y}) = \nabla f(\vec{y}_0) \cdot (\vec{x} - \vec{y})$ with $\vec{y}_0$ a point between $\vec{x}$ and $\vec{y}$. Let us call $A = \max |\nabla f(\vec{y}_0)| > 0$, so $|f(\vec{x}) - f(\vec{y})| < A|\vec{x} - \vec{y}|$. That is, f is Hölder of exponent 1 at $\vec{x}$.

   Analogously, any function f having $n-1$ vanishing derivatives at a point $\vec{x}$ and a non-vanishing continuous n-th derivative, is Hölder of exponent n at $\vec{x}$.

3. **θ-function:** The $\theta(x)$ function is given by

$$\theta(x) = \begin{cases} 0 & x < 0 \\ 1 & x > 0 \\ \text{undefined} & x = 0 \end{cases}$$

   Since for $x \neq 0$ the increments are zero (for small enough displacement) it is Hölder of any order. For $x = 0$ we make use of the property that f is Hölder of exponent h at a given point if and only if its primitive F is Hölder of exponent $h+1$ at the same point.

   One possible primitive $F(x)$ of $\theta(x)$ is the following:

$$F(x) = \begin{cases} 0 & x < 0 \\ x & x \geq 0 \end{cases}$$

   which is a continuous function. The increments $|F(0) - F(r)| = |F(r)| = F(r)$ are thus immediate, and obviously $F(r) < |r|$. Moreover, it is also clear that the maximal h such that $|F(0) - F(r)| < A|r|^h$ is $h = 1$. Thus, F has Hölder exponent 1 at $x = 0$ and therefore $\theta(x)$ has Hölder exponent 0 at $x = 0$.

4. **δ-function:** The Dirac's δ-function is in fact a distribution. Anyway, it is possible to define the Hölder exponent of distributions attending to the fact that any distribution can be expressed as a finite order derivative of a bounded function. So, if the n-th order primitive of a distribution is a Hölder function of exponent h at a given point, the distribution is Hölder of exponent $h-n$ at the same point.

   Since the $\delta(x)$ distribution vanishes in $x \neq 0$ the only non-trivial point is $x=0$. Since the δ is the derivative of $\theta(x)$ and this has Hölder exponent 0 at $x=0$, one concludes that the $\underline{\delta(x) \text{ has Hölder exponent } -1 \text{ at } x=0}$

References

[Arneodo et al. 1995] Arneodo et al. *"Ondelettes, multifractales et turbulence"*. Diderot Editeur. Paris, 1995.

[Arneodo et al. 1996] Arneodo et al. *"Wavelet analysis of fractals: from the mathematical concepts to experimental reality"*. in *Wavelets. Theory and applications* (eds. Erlebacher G., Yousuff Hussaini M. and Jameson L.M.) pp. 349. (Oxford University Press. ICASE/LaRC Series in Computational Science and Engineering. 1996).

[Atick 1992] Atick J. J. *Network* **3** 213-251, 1992.

[Bacry et al. 1993] Bacry E., Muzy J.F. and Arneodo A. *J. of Stat. Phys.*, **70**, 635-673 (1993).

[Barlow 1961] Barlow H. B., in *Sensory Communication* (ed. Rosenblith W.) pp. 217. (M.I.T. Press, Cambridge MA, 1961).

[Bell and Sejnowski 1997] Bell and Sejnowski, *Vision Research* **37**, 3327-3338 (1997).

[Benzi et al 1993a] Benzi R., Biferale L., Crisanti A., Paladin G., Vergassola M. and Vulpiani A. , *Physica D* **65**, 352-358 (1993).

[Benzi et al 1993b] Benzi R., Ciliberto S., Baudet C., Ruiz Chavarria G. and Tripiccione C., *Europhys. Lett.* **24** 275-279 (1993)

[Benzi et al 1993c] Benzi, Ciliberto, Tripiccione, Baudet, Massaioli, and Succi, *Phys. Rev. E* **48**, R29 (1993)

[Benzi et al 1995] Benzi, Ciliberto, Baudet and Chavarria *Physica D* **80** 385-398 (1995)

[Castaing 1996] Castaing, *J. Physique II, France* **6**, 105-114 (1996)

[Daubechies 1992] Daubechies I. *"Ten Lectures on Wavelets"*. CBMS-NSF Series in Ap. Math. *Capital City Press.* Montpelier, Vermont, 1992.

[Falconer 1990] Falconer K. *"Fractal Geometry: Mathematical Foundations and Applications". John Wiley and sons.* Chichester,1990.

[Field 1987] Field D. J., *J. Opt. Soc. Am.* **4** 2379-2394 (1987).

[Frisch 1995] Frisch U., *"Turbulence"*. Cambridge Univ. Press, 1995.

[Gonzalez and Woods 1992] Gonzalez R.C. and Woods R.E., *"Digital image processing"*. Addison-Wesley, 1992.

[van Hateren 1992] van Hateren J.H., "Theoretical predictions of spatiotemporal receptive fields of fly lmcs, and experimental validation". *J. Comp. Physiology A*, **171** 157–170, (1992).

[van Hateren and van der Schaaf 1998] van Hateren J.H. and van der Schaaf A., "Independent component filters of natural images compared with simple cells in primary visual cortex". *Proc.R.Soc.Lond.* **B265**, 359-366 (1998)

[Linsker 1988] Linsker R., " Self-organization in a perceptual network". *Computer*, **21**, 105–17 (1988).

[Mallat and Zhong 1991] Mallat S., Zhong S. *"Wavelet transform Maxima and multiscale edges"*. In *"Wavelets and their applications"*. Jones and Bartlett Publishers, Boston 1991.

[Mallat and Huang 1992] Mallat S., Huang W.L. *IEEE Trans. in Inf. Th.* **38**, 617-643 (1992)

[Novikov 1994] Novikov, *Phys. Rev. E* **50**, R3303 (1994)

[Ruderman and Bialek 1994] Ruderman D. and Bialek W., *Phys. Rev. Lett.* **73**, 814 (1994)

[Ruderman 1994] Ruderman D., *Network* **5**, 517-548 (1994)

[She and Leveque 1994] She and Leveque, *Phys. Rev. Lett.* **72**, 336-339 (1994).

[She and Waymire 1994] She and Waymire, *Phys. Rev. Lett.* **74**, 262-265 (1995).

[Turiel, Mato, Parga and Nadal 1998] Turiel A., Mato G., Parga N. and Nadal J.-P. *Phys. Rev. Lett.* **80**, 1098-1101 (1998)

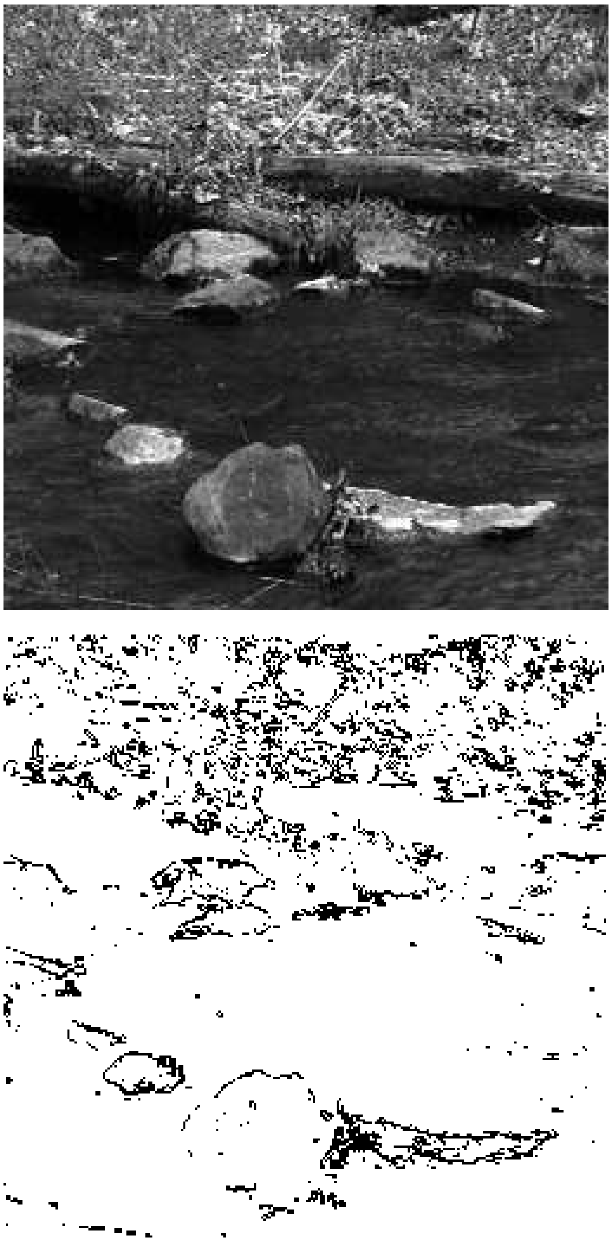

Figure 1: An image from Ruderman's ensemble [Ruderman 1994] and its most singular component (taken as the set of points having a singularity exponent in the interval $h = -0.5 \pm 0.05$).

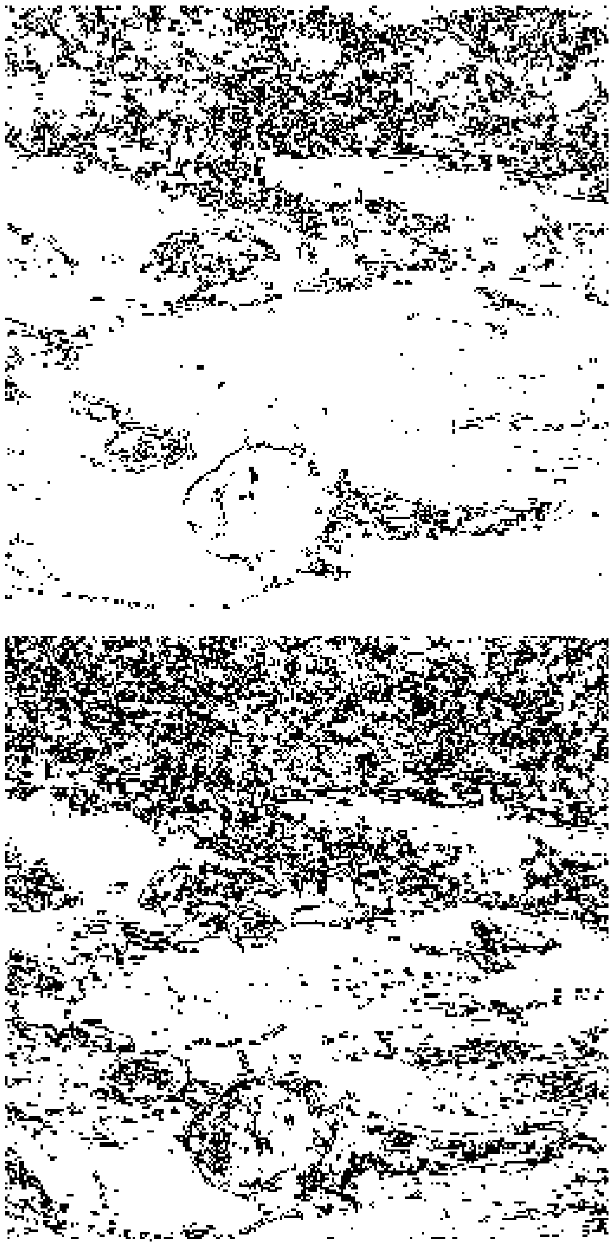

Figure 2: Fractal components with exponents $h = -0.4 \pm 0.05$ (above) and $h = -0.3 \pm 0.05$ (below) for the same image represented in Fig. 1. Notice the spatial correlations between components with similar singularities.

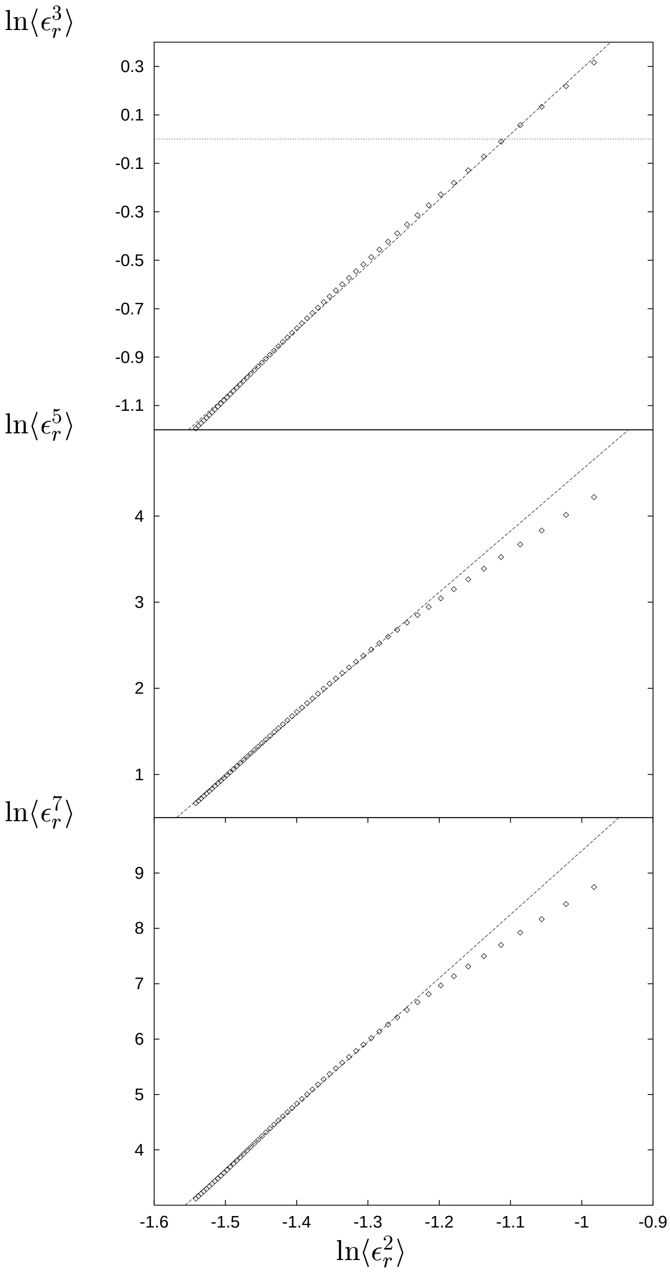

Figure 3: Test of ESS for the bidimensional edge content ϵ_r. The graphs represent the logarithm of its moments of order 3, 5 and 7 versus the logarithm of the second order moment, for distances $r = 4$ to $r = 64$ pixels. The data corresponds to the second image ensemble. According to eq. (16) these graphs should be straight lines; the best linear fits are also represented. Except for lower cut-off effects at very small distances, the scaling law is well verified.

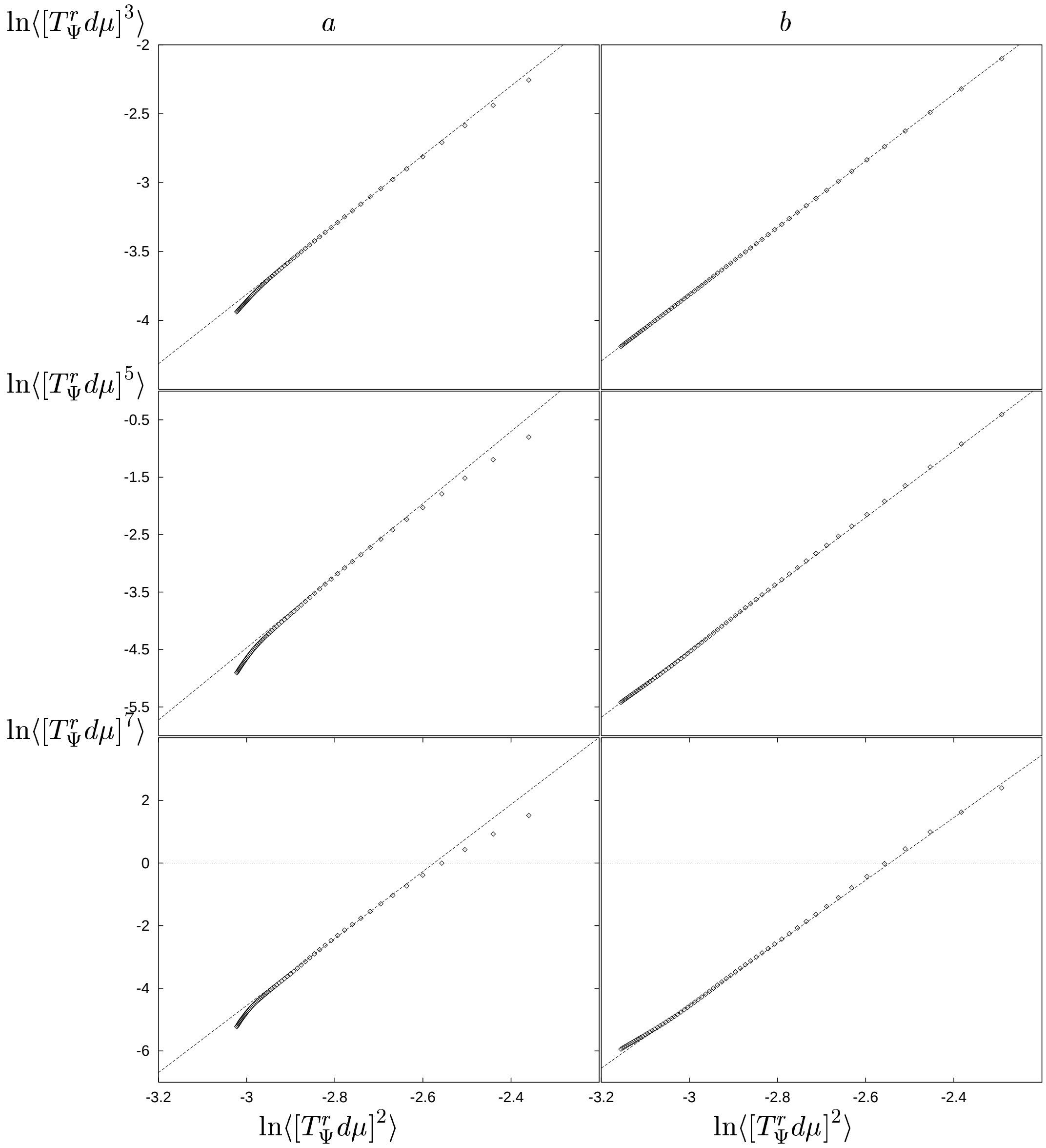

Figure 4: Test of ESS for **a)** $T_\Psi^r d\mu^h$ and **b)** $T_\Psi^r d\mu^v$. The graphs represent the logarithm of the moments of order 3, 5 and 7 for both variables versus the logarithm of the corresponding second order moment. The scales range from $r = 4$ to $r = 64$ pixels. The first dataset was used in this computation. The corresponding best linear fits are also represented. Except for lower and upper cut-off effects at small and large distances, the scaling law verifies well. Although the test is not shown here, SS also holds for these variables.

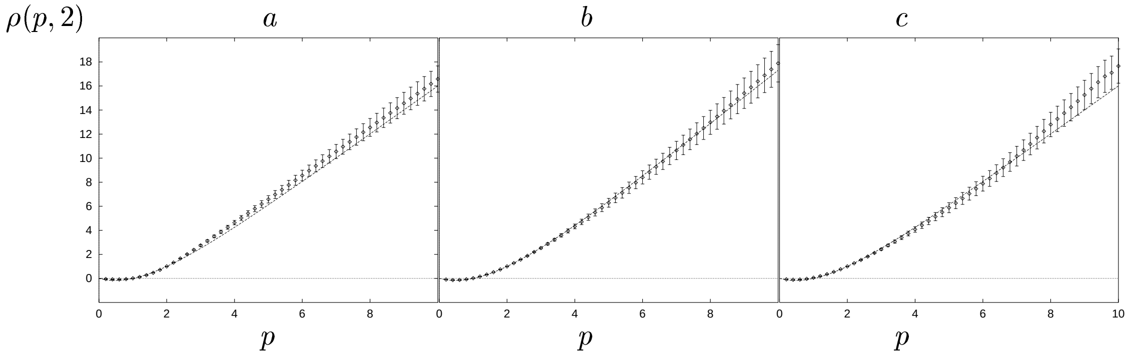

Figure 5: ESS exponents $\rho(p,2)$ for **a)** the bidimensional edge content ϵ_r , **b)** the horizontal wavelet transform $T^r_\Psi d\mu^h$ and **c)** the vertical one, $T^r_\Psi d\mu^h$. The convolution function Ψ was taken to be a gaussian function. Each value of $\rho(p,2)$ was obtained by linear regression of the logarithm of the p-th moment versus the logarithm of the second moment, for r between 8 and 32. The solid line represents the fit with the Log-Poisson process. The best fit is obtained with **a)** $\beta = 0.52 \pm 0.05$ **b)** $\beta = 0.55 \pm 0.06$ and **c)** $\beta = 0.5 \pm 0.07$ The SS parameters τ_2 were also calculated; they turned out to be **a)** $\tau_2 = -0.26 \pm 0.06$, **b)** $\tau_2 = -0.25 \pm 0.06$ and **c)** $\tau_2 = -0.25 \pm 0.08$.

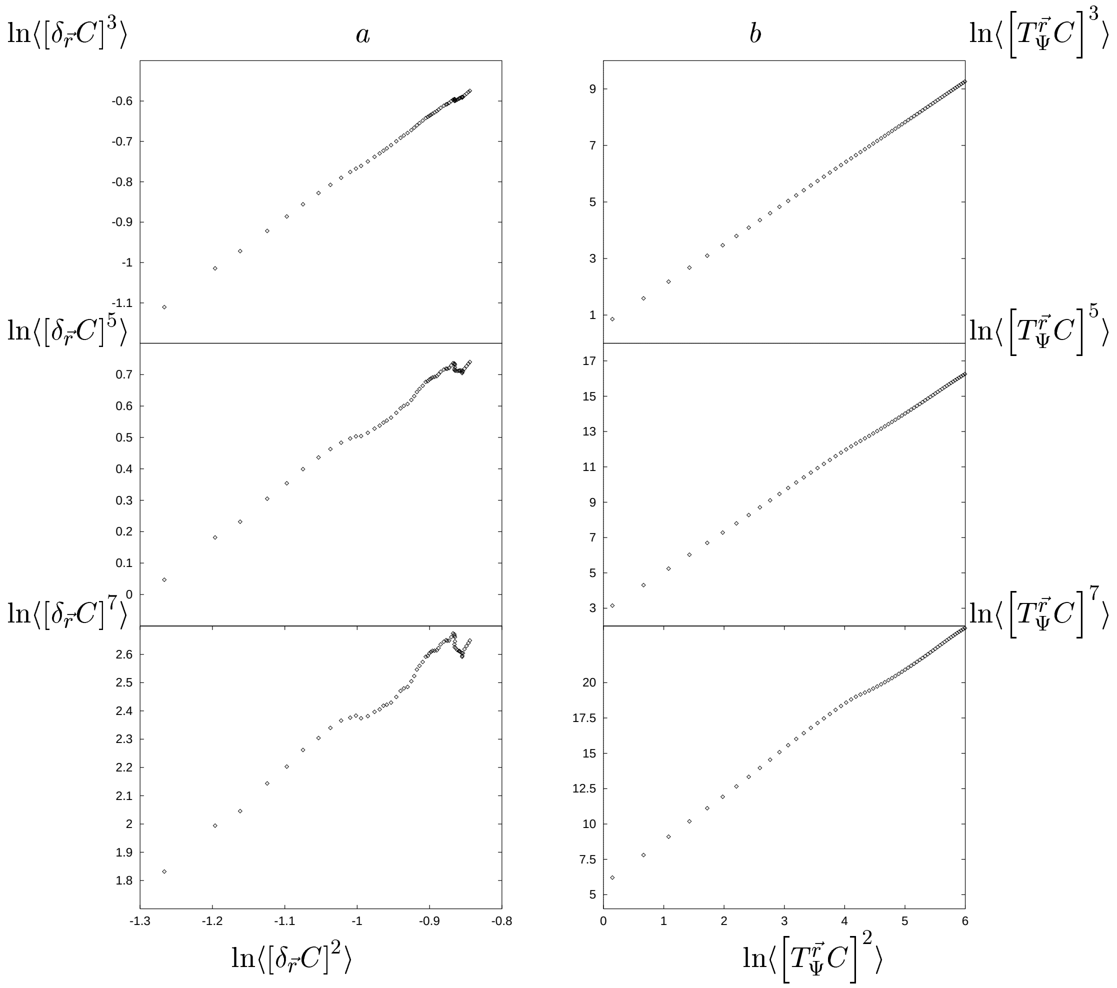

Figure 6: Test of ESS for **a)** the horizontal linear increment of the contrast, $\delta_{r\vec{i}}C$ and **b)** the horizontal wavelet coefficient $|T_\Psi^r C^h|$ (Ψ was taken as the second derivative of a gaussian); the first dataset was used and the range of scales was from $r = 4$ to $r = 64$. ESS is not present for $\delta_{r\vec{i}}C$, but it is clearly observed for the wavelet transform (notice the presence of the cut-offs due to numerical effects). Note that the scale of the axes in **a)** and **b)** is different.

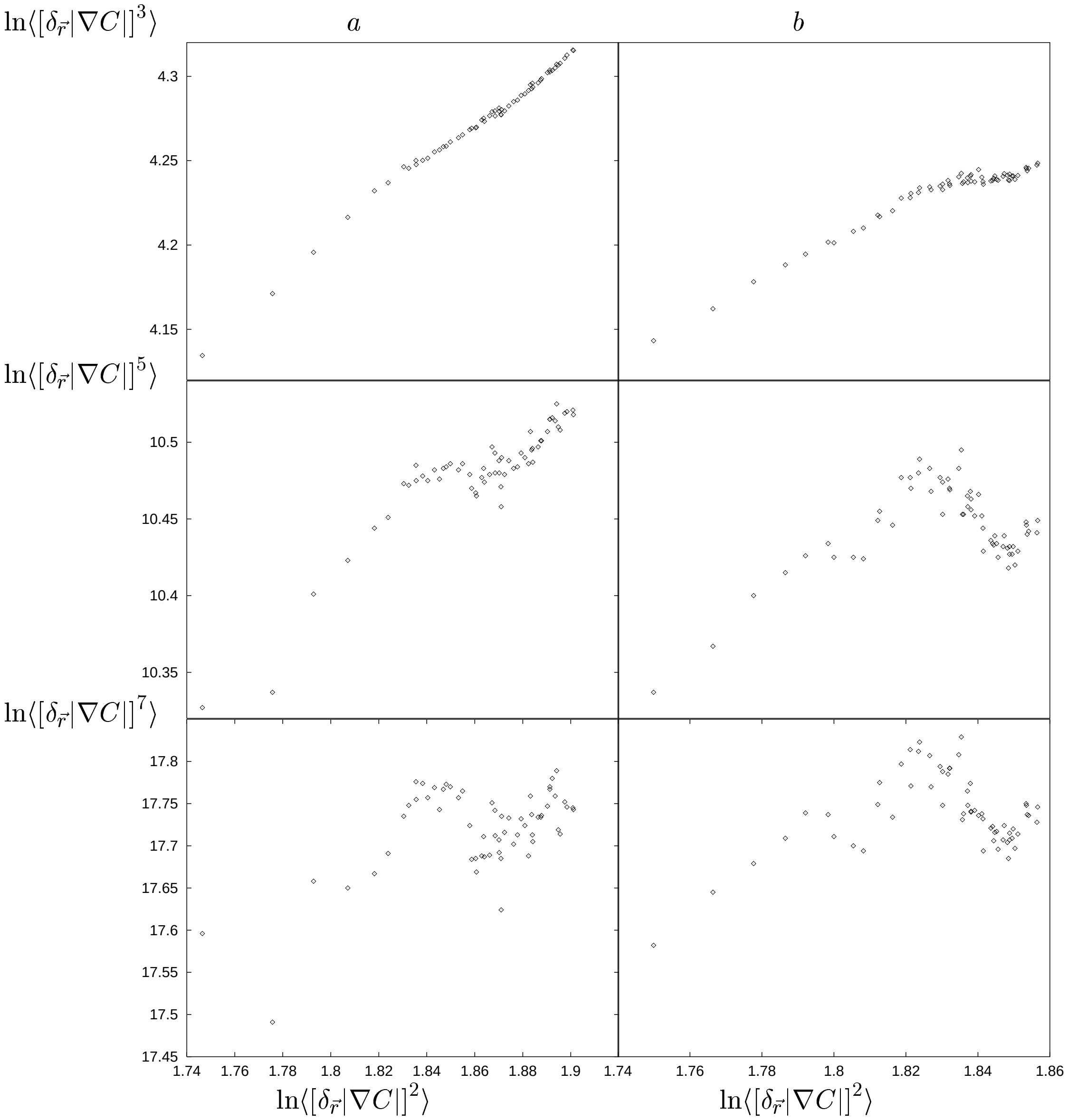

Figure 7: Test of ESS for **a)** horizontal linear increment of $|\nabla C|$, $\delta_{r\vec{i}}|\nabla C|$ and **b)** vertical linear increment of the same function, $\delta_{r\vec{j}}|\nabla C|$. The first dataset was used and the scales were taken ranging from $r = 4$ to $r = 64$ pixels. Again, these variables do not have ESS. The corresponding wavelet projections are shown in Fig. 4 where one can see that ESS holds.

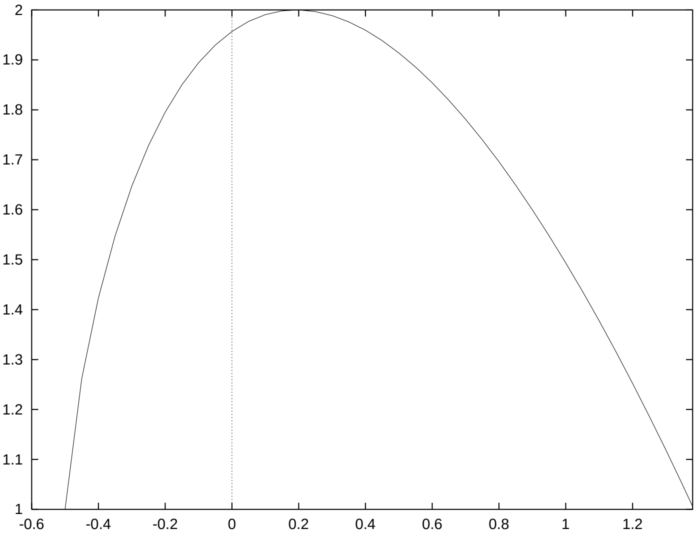

Figure 8: Singularity spectrum of the Log-Poisson model for $\beta = 0.5$ and $\tau_2 = -0.25$. For these values $D_\infty = 1.$ and $\Delta = 0.5$ (see text)