On Generalized Entropies and Scale-Space

Jon Sporring*

(sporring@diku.dk)
Image Group
Dept. of Computer Science
University of Copenhagen
Universitetsparken 1
DK-2100 Copenhagen
DENMARK

Joachim Weickert†

(joachim@cv.ruu.nl)
Image Sciences Institute
Utrecht University Hospital
Heidelberglaan 100
E 01.334
NL-3584 CX Utrecht
THE NETHERLANDS

March 13, 1997

1 Introduction

It has been suggested by Jägersand [12], Oomes *et al.* [17, 16], and Sporring [25] that the entropy measure can be used in image processing as a tool to analyse images in the Scale-Space of Iijima [28], Witkin [30], and Koenderink [14].

The entropy measure was initially proposed independently by Shannon [23] and Wiener [29] as a measure of the information content per symbol emitting from a stochastic information source. Later Rényi [20] extended this to include all generalized means to yield the generalized entropy.

There are several reasons why the generalized entropy function is of interest to image processing. Firstly, Weickert [27] has shown that Lyaponov functionals have monotonically decreasing behavior in Scale-Space and as such may serve as causality measures. In this paper we show that the negative generalized entropies are such functionals. It should be noted that this behavior is not seen for the number of critical points: Although critical points most often disappear when scale is increased, creation of critical points with increasing scale is a generic event [15, 13, 5]. Secondly, generalized entropy is the basis for the theory of Multi-Fractals [10, 18]. We briefly investigate the scaling behaviors of fractal spectra.

*Part of this work was performed during a Joint-Study visit at IBM, Almaden Research Center, 650 Harry Road, San Jose, CA 95120-6099, USA

†Current Address: Dept. of Computer Science, University of Copenhagen, 2100 Copenhagen East, Denmark

Related to this work is Vehel *et al.* [26], where images are studied in the multi-fractal setting, focusing on certain dimensions, and Brink & Pendock [4], and Brink [3] have used the entropy and the closely related Kullback measure to do local thresholding of images.

This article is organized as follows. First, in Section 2 will be given a brief introduction to Linear Scale-Space. Then, in Section 2 will we discuss the generalized entropies, and what their properties are in Scale-Space. Finally, in Section 3 we will discuss an interpretation of images in terms of Multi-Fractals.

2 Information Theory and Scale-Space

It has been argued [2] that the continuum of Generalized Entropies is the only sensible measure of information. As will be shown below, the Shannon entropy is part of this continuum, and further the difference between the normal entropy and the other generalized entropies are intimately linked through the norm operator.

2.1 The Axioms of Information Theory

We will now review the axiomatically derivation of the Generalized Entropies with some brief historical remarks.

The theory of information was born by Hartley's establishment of the additivity axiom [9]:

Axiom 1 (Additivity). *The information contents of two independent events is the sum of the information of the individual events.*

I.e. if one is to distinguish a single element from a set, the set can recursively be broken into two subsets where each subset is followed by a binary decision. One thus needs $\log_2 N$ binary decisions to identify a single element from a set of size N. Further, in product set the number of decisions to identify a specific tuple is in this fashion additive since one first identify the first followed by the second element in a sequential fashion.

This of course assumes that each element is equally probable, and Shannon's contribution was to define the entropy as the linear mean of information [23]:

Axiom 2 (Linear mean). *The entropy of independent events is the mean of the information of the individual events.*

The Shannon entropy $H(P) = -\sum_i p_i \log p_i$ is uniquely determined through the above axiom together with a continuity requirement of H in each p_i, and a requirement that $H(\{1/n\})$ should be monotonical increasing function as n is increased, i.e. when comparing entropies of uniform distributions, but where the number of choices increase.

Finally Rényi [20, 19] relaxed the constraint of the linear mean to generalized mean in the sense of Aczél [1] and others.

Axiom 3 (Generalized mean). *The generalized entropy of independent events is the generalized mean of the information of the individual events.*

The generalized mean $M(P)$ based on simple requirements is found to be either $M(P) = \prod_i p_i^{p_i}$ or $(\sum_k p_i^\alpha)^{1/(\alpha-1)}$ [1, 8, 19] for $\alpha > 0$ and $\alpha \neq 1$. The general entropy of order $\alpha \in \mathbb{R}$ is thus defined as minus the logarithm of $M(P)$, yielding either Shannon's entropy or

$$S_\alpha(p) = \frac{1}{1-\alpha} \log \sum_i p(i)^\alpha \tag{1}$$

for $\alpha \neq 1$[§].

2.2 Some Selected Information Orders

To briefly review, let us investigate some particular information orders. It is easily seen by the Max-norm that

$$\lim_{\alpha \to \infty} S_\alpha(p) = -\log \max_i p(i).$$

In the limit for α going to 1 the Generalized Entropy converges (by l'Hospital's rule) to Shannon's Entropy,

$$\lim_{\alpha \to 1} S_\alpha(p) = S(p) = -\sum_i p(i) \log p(i).$$

Using $\lim_{x \to 0} 0^x = 0$ the convergence towards $\alpha \to 0$ is given by,

$$\lim_{\alpha \to 0} S_\alpha(p) = \log \mathrm{Card}(\{i | p(i) > 0\}),$$

which for non-fractal domains is the dimensionality of the domain. Finally, for negative α's the generalized entropies are undefined for zero valued probabilities and in general unstable for small probability values, but a pragmatic approach is to define the negative information orders on $\tilde{p} = \{p(i) | p(i) > 0\}$, and thus

$$\lim_{\alpha \to -\infty} S_\alpha(p) = -\log \min_i \tilde{p}(i).$$

The reader should note that there are interesting analogies in terms of Thermodynamics, and that the generalized entropy is the basis of the theory of Multi-Fractals.

2.3 The Scale-Space Extension

As in [25] we will now examine the generalized entropies in the Linear Scale-Space perspective [30, 14, 6, 28], to explicitly study the effect of discretization.

[§]Note that for large $|\alpha|$'s this formulation is computationally unsuited. It is much better to calculate $S_\alpha(p) = \frac{1}{1-\alpha}(\log \sum_i (p(i)/\max_i p(i))^\alpha + \alpha \log \max_i p(i))$

The image is extended with the scale or time parameter t by imposing the Heat Diffusion Process, $\partial_t = \partial_{x_i x_i}$, using Einstein's summation convention. It is well known that the Green's function for this process is the Gaussian kernel, and the process can thus be modeled by convolution

$$p_t = G_t * p = \int G_t(x-\alpha)p(\alpha)\, d\alpha = \frac{1}{(\pi t)^{D/2}} \int e^{||x-\alpha||_2^2/t} p(\alpha)\, d\alpha,$$

where $||\cdot||_2$ is the two norm and D is the dimension of the domain. Note that the standard deviation is $\sigma = \sqrt{t/2}$.

To overcome the interpretative difficulties with continuously defined entropies this present work will investigate the following discretization procedure, also known as a histogram,

$$p_t(i) = \int_{ih-\frac{h}{2}}^{ih+\frac{h}{2}} p_t(x)\, dx \simeq \frac{hp_t(ih)}{\sum_j hp_t(jh)} = \frac{p_t(ih)}{\sum_j p_t(jh)}, \tag{2}$$

where i is an integer and h is the width of the bins in the histogram. A purely statistical discussion on the choice of bin size and the effects of smoothing can be found in [24]. In Scale-Space the size of the bins h should be related to the amount of smoothing performed. A linear function of the standard deviation, based on the calculation of the standard deviation of a uniform distribution (a bin) of width n to be $n/\sqrt{12}$. Avoiding details we will in the following just assume

$$h = c\sqrt{t} \tag{3}$$

Combining Equation 2 and 1 yields,

$$S_{\alpha,t}(p) = \frac{1}{1-\alpha} \log \frac{\sum_i (hp_t(ih))^\alpha}{\left(\sum_j hp_t(jh)\right)^\alpha}.$$

For the further analysis we will assume the normalization constant to be independent of t and equal 1. This then allows us to analyse the h dependency,

$$S_{\alpha,t}(p) \simeq \frac{\alpha}{1-\alpha} \log h + \frac{1}{1-\alpha} \log \sum_i p_t(ih)^\alpha.$$

The above formulation depends on the zero-point of the discretization grid, and to avoid this we will take the mean generalized entropy over all grid offsets. By Jensen's inequality it can be seen that,

$$\begin{aligned}
\langle S_{\alpha,t}(p)\rangle(1-\alpha) &= \alpha \log h + \int_0^h \frac{1}{h} \log \sum_i p_t(ih+a)^\alpha\, da \\
&\leq \alpha \log h + \log \int_0^h \frac{1}{h} \sum_i p_t(ih+a)^\alpha\, da \\
&= (\alpha - 1) \log h + \log \int_{x\in\Omega} p_t(x)^\alpha\, dx,
\end{aligned}$$

where Ω is the domain of p, and the integration should be replaced by a sum in the case this domain is discrete. Thus for $\alpha > 1$ we have,

$$\langle S_{\alpha,t}(p)\rangle \geq -\log c - \frac{1}{2}\log t + \frac{1}{(1-\alpha)}\log \int_{x\in\Omega} p_t(x)^\alpha \, dx,$$

and for $\alpha < 1$ we have

$$\langle S_{\alpha,t}(p)\rangle \leq -\log c - \frac{1}{2}\log t + \frac{1}{(1-\alpha)}\log \int_{x\in\Omega} p_t(x)^\alpha \, dx.$$

The equality holds for $\alpha = 1$ [25]. The reader should note that the logarithm of the scale plays a dominating role. But in a study of $\langle S_{\alpha,t}(p)\rangle$ as a function of $\log t$ this term will be less interesting, and we are thus motivated to study the change of $1/(1-\alpha)\log \int_{x\in\Omega} p_t(x)^\alpha \, dx$ by scale. The reader should note that for each fixed scale t, the above equation is only a constant different from the original proposed generalized entropy, and hence all the properties discussed previously are still valid.

In Figure 1 is given an example of the Generalized Entropy surfaces as function of scale. The Scale-Space is a spatial implementation with homogeneous Neumann boundary conditions.

2.4 Some properties of Generalized Entropies in Scale-Space

In this section further details of the mathematical structure of the Generalized entropies will be established.

Proposition 2.1. *The Generalized Entropy is a decreasing function of order.*

Proof. The proof of the monotonically decreasing behavior of the Generalized Entropies is basically a restatement of the proof given in [20, 11]. A generalized Hölder inequality states that for $m_r(a) = \left(\sum_i p_i a_i^r\right)^{1/r}$

$$m_r(a) > m_{r'}(a),$$

for $r > r'$ and $a_i \geq 0$ and not all zero [8, Theorem 16, p. 26]. $m_r(a)$ is also sometimes called the expected norm of a under p. The generalized entropies are defined as minus the logarithm of the expected norm, and since the logarithm is a monotonic function,

$$S_\alpha = -\log m_{\alpha-1}(p) < -\log m_{\alpha'-1}(p) = S_{\alpha'},$$

for $\alpha > \alpha'$. □

Proposition 2.2. *Let $p_i(t) > 0$ for all i and for all $t > 0$. Then the generalized entropies*

$$S_\alpha(P(t)) = \frac{1}{1-\alpha}\ln \sum_{i=1}^{N} p_i^\alpha$$

are increasing in t for $\alpha > 0$, constant for $\alpha = 0$, and decreasing for $\alpha < 0$. For $t \to \infty$, they converge to the zeroth order entropy S_0.

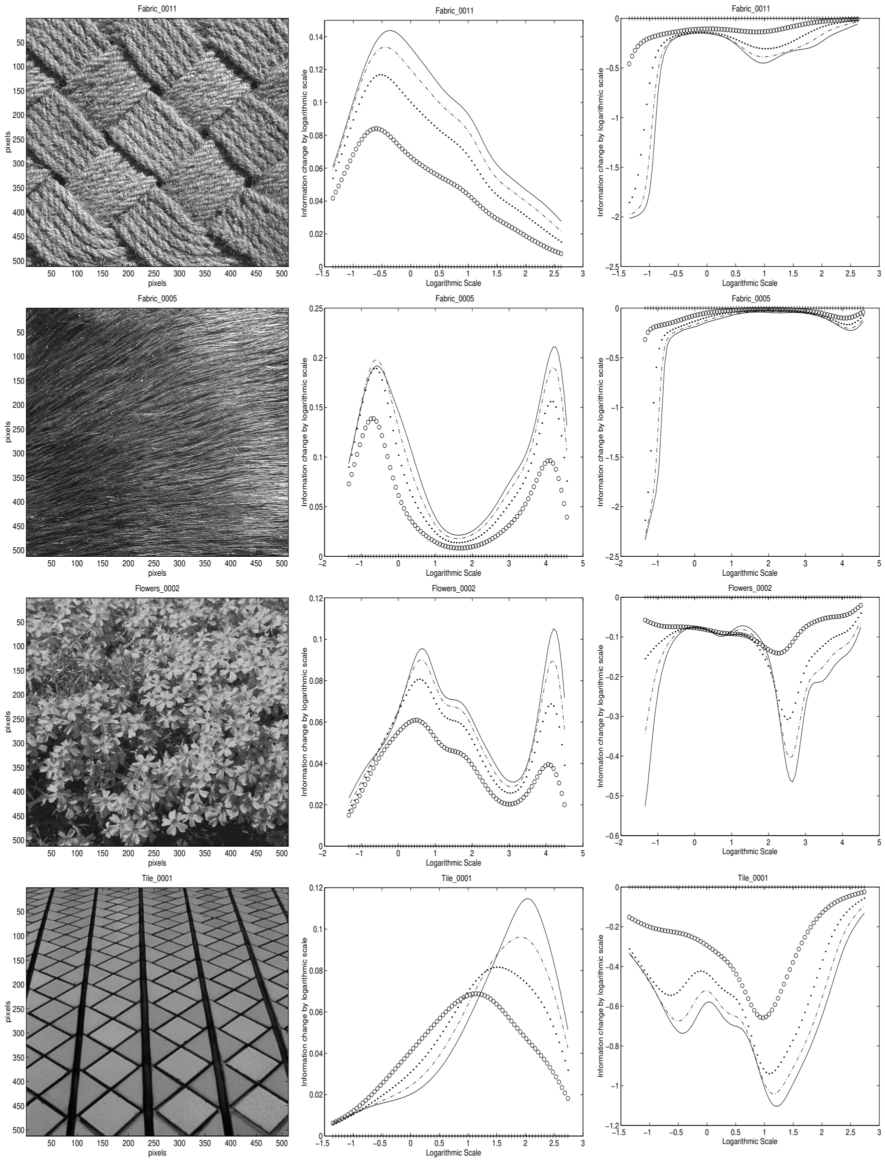

Figure 1: Examples of a generalized entropy functions. Left shows the images, middle and right the corresponding the generalized entropy change by log scale for positive and negative information orders respectively. The curves are slices of fixed order of the order×scale surface. the '+'s is order 0, the 'o's are orders ±5, the '.'s are orders ±10, the '.-'s are orders ±15 and the full lines are orders ±20 respectively.

Proof. The proof is based on results from [27] implying that for an image $P(t)$, which is obtained from a diffusion scale-space, the following holds: The expression

$$V(t) := \sum_{i=1}^{N} r(p_i(t))$$

is decreasing in t for every smooth convex function r. Moreover, $\lim_{t\to\infty} p_i(t) = 1/N$ for all i. See Theorem 5 in [27, p. 68] for full details.

First we prove the monotony of S_α with respect to t.

(a) Let $\alpha > 1$ and $s > 0$. Since $r(s) = s^\alpha$ satisfies

$$r''(s) = \alpha(\alpha - 1)s^{\alpha-2} > 0,$$

it follows that r is convex. Thus,

$$V(t) = \sum_i r(p_i(t)) = \sum_i p_i^\alpha(t)$$

is decreasing in t and

$$S_\alpha(P(t)) := \frac{1}{1-\alpha} \ln V(t)$$

is increasing in t.

(b) Let $\alpha = 1$. By l'Hospital's rule it follows that

$$S_1(P(t)) = -\sum_i p_i(t) \ln p_i(t).$$

This is just the classical entropy. Since $r(s) = s \ln s$ is convex, we conclude that the entropy is increasing in t. This result has already been shown in [27, p. 71].

(c) Let $0 < \alpha < 1$ and $s > 0$. Then $r(s) = s^\alpha$ satisfies $r''(s) < 0$. Thus, $-r$ is convex and

$$V(t) := -\sum_i r(p_i(t)) = -\sum_i p_i^\alpha$$

is decreasing in t. Thus, $\sum\limits_i p_i^\alpha$ and $S_\alpha(P(t))$ are increasing.

(d) Let $\alpha = 0$. Then

$$S_0(P(t)) = \ln \sum_{i=1}^{N} p_i^0(t) = \ln N = \text{const.}$$

for all t.

(e) Let $\alpha < 0$ and $s > 0$. Since $r(s) = s^\alpha$ satisfies $r''(s) > 0$ it follows that r is convex. Thus, $\sum_i p_i^\alpha(t)$ and $S_\alpha(P(t))$ are decreasing in t.

To verify the asymptotic behavior of the generalized entropies we utilize $\lim_{t\to\infty} p_i(t) = 1/N$. For $\alpha \neq 1$ this gives

$$\lim_{t\to\infty} S_\alpha(t) = \frac{1}{1-\alpha} \ln \sum_{i=1}^{N} \frac{1}{N^\alpha} = \ln N = S_0,$$

and $\alpha = 1$ yields

$$\lim_{t\to\infty} S_1(t) = -\sum_{i=1}^{N} \frac{1}{N} \ln \frac{1}{N} = \ln N = S_0.$$

This completes the proof. □

Proposition 2.3. *Let $p_i(t) > 0$ for all i and for all $t > 0$. Then the generalized entropies $S_\alpha(P(t))$ are C^∞ in t and (at least) C^1 in α.*

Proof. C^∞ in t follows directly from the fact that $G_t(x)$ is in C^∞ with respect to t.

In order to prove smoothness with respect to α, we first consider the case $\alpha \neq 1$. Then S_α is the product of the two C^∞ functions $\frac{1}{1-\alpha}$ and $\ln \sum_{i=1}^N p_i^\alpha$, and thus also C^∞ in α.

The smoothness in $\alpha = 1$ is verified by applying l'Hospital's rule. This gives

$$\begin{aligned}
\lim_{\alpha\to 1} \frac{\partial S_\alpha}{\partial \alpha} &= \lim_{\alpha\to 1} \frac{(1-\alpha)\sum_i p_i^\alpha \ln p_i + \sum_i p_i^\alpha \ln \sum_i p_i^\alpha}{(1-\alpha)^2 \sum_i p_i^\alpha} \\
&= \lim_{\alpha\to 1} \frac{(1-\alpha)\sum_i p_i^\alpha (\ln p_i)^2 + \sum_i p_i^\alpha \ln p_i \ln(\sum_i p_i^\alpha)}{2(1-\alpha)\sum_i p_i^\alpha \ln p_i + (1-\alpha)^2 \sum_i p_i^\alpha \ln p_i} \\
&= \frac{-\sum_i p_i (\ln p_i)^2 + (\sum_i p_i \ln p_i)^2}{-2}.
\end{aligned}$$

Thus, $\frac{\partial S_\alpha}{\partial \alpha}$ exists and S_α is in C^1. □

3 A Multi-Fractal Description

The generalized entropies are also the basis of what is known as the theory of Multi-Fractals, and this thus offers a second view on the generalized entropies.

A fractal is a self similar object, i.e. in a zooming sequence the greater 'image' can be found replicated at a smaller scale. It is thus natural to investigate the rate of replication with scale, the so-called Lyaponov exponent. In the theory of multi fractals there is not assumed to be one global scaling exponent but an entire spectrum of exponents each associated with an area in the 'image'.

In the following a summary will be given of multi-fractal theory [10, 21, 22]. Given an density function ρ, a discretization can be performed as,

$$p_{i(l)} = \int_{x \in \Omega_{i(l)}} \rho(x)\, dx,$$

where $\Omega_{i(l)}$ is the i'th box in the grid of boxes of width l covering the domain of x.¶ The point $i(l)$ then has a fractal behavior with scaling exponent $\alpha \in \mathbb{R}$ if it converges like

$$p_{i(l)} \approx l^{\alpha},$$

when $l \to 0$. Remember that ρ is an unobservable quantity. Only p is known in some interval of scales.

We are for the moment interested in the sum of the q powers of p_i,

$$M_q = \sum_i p_{i(l)}^q,$$

and thus we associate an α for each q. Reformulating this using that the density of scaling exponents α can be written as [10],

$$dx = \rho(\alpha) l^{-f(\alpha)}\, d\alpha,$$

where $f : \mathbb{R} \to \mathbb{R}$ is a continuous function to be determined shortly, and measures the Hausdorff dimension of the set of points with exponent α. We can now write

$$M_q \simeq \int l^{\alpha q} \rho(\alpha) l^{-f(\alpha)}\, d\alpha.$$

Because of the smallness of l this integral is dominated by the minimum α' of $\tau(\alpha) = \alpha q - f(\alpha)$ in α, i.e.

$$M_q \approx l^{\tau(\alpha')}.$$

From the fact that τ is minimal in α' we know that $\partial_\alpha \tau = 0$ and $\partial_{\alpha^2} \tau \geq 0$, and this leads to

$$f'(\alpha') = q,$$

and

$$f''(\alpha') \leq 0.$$

Finally define the multi-fractal dimension [11],

$$\begin{aligned} D_q &= -\lim_{l \to 0} \frac{S_q}{\log l} \\ &= -\lim_{l \to 0} \frac{1}{1-q} \frac{\log M_q}{\log l} \\ &\approx -\frac{1}{1-q} \tau. \end{aligned}$$

¶ρ is also called the invariant density since if you reverse the discretization process, this density function is the convergence point invariant to further zooming.

This measure has been shown to unify a great variety of fractal measures used in the past [7], e.g. the Hausdorff dimension can be shown to be D_0, D_1 the information dimension and D_2 the correlation dimension.

Thus given D_q one can calculate α' as,

$$\alpha' = -\frac{d(1-q)D_q}{dq},$$

and hence the multi-fractal spectrum as

$$f(\alpha') = (1-q)D_q + q\alpha'.$$

An example is given in Figure 2. As the image simplifies with scale, the gener-

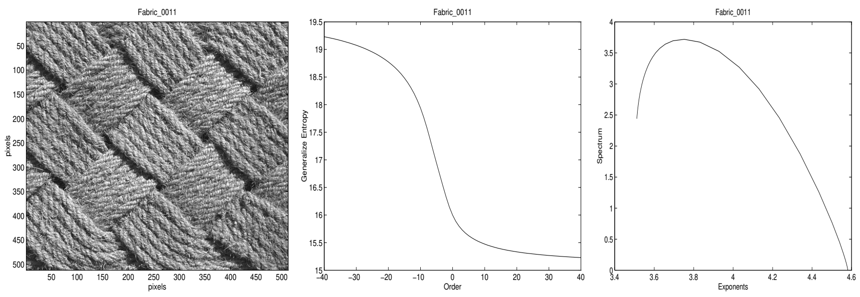

Figure 2: In this figure are the image (left), the generalized entropy at fine scale (middle) and the corresponding spectrum (right) shown.

alized entropy converges to a constant function, and the spectrum converges to a point function.

4 Acknowledgment

We would like to thank Peter Johansen, Mads Nielsen, Luc Florack, Robert Maas and Ole Olsen for the many and enlightening discussions on the subjects related to Scale-Space, Information Theory and the like.

References

[1] J. Aczél. On mean values. *Bulletin of the American Mathematic Society*, 54:392–400, 1948.

[2] J. Aczél and Z. Daróczy. *On information measures and their characterizations.* Academic Press, 1973.

[3] A. D. Brink. Using spatial information as an aid to maximum entropy image threshold selection. *Pattern Recognition Letters*, 17(1):29–36, January 1996.

[4] A. D. Brink and N. E. Pendock. Minimum cross-entropy threshold selection. *Pattern Recognition*, 29(1):179–188, January 1996.

[5] James Damon. Local morse theory for gaussian blurred functions. In Jon Sporring, Mads Nielsen, Luc Florack, and Peter Johansen, editors, *Gaussian Scale-Space Theory.* Kluwer Academic Publishers, The Netherlands, 1997. (in print).

[6] L. M. J. Florack, B. M. ter Haar Romeny, J. J. Koenderink, and M. A. Viergever. Linear scale-space. *Journal of Mathematical Imaging and Vision*, 4(4):325–351, 1994.

[7] P. Grassberger and I. Procaccia. Dimensions and entropies of strange attractors from a fluctuating dynamics approach. *Physica D*, 13:34–54, 1984.

[8] G. H. Hardy, J. E. Littlewood, and G. Pólya. *Inequalities.* Cambridge University Press, 1934.

[9] R. V. Hartley. Transmission of information. *Bell System Technical Journal*, 7:535–563, 1928.

[10] Thomas C. Hasley, Mogens H. Jensen, Leo P. Kadanoff, Itamar Procaccia, and Boris I. Shraiman. Fractal measures and their singularities: The characterization of strange sets. *Physical Review A*, 33:1141–1151, 1986.

[11] H. G. E. Hentschel and Itamar Procaccia. The infinite number of generated dimensions of fractals and strange attractors. *Physica D*, 8:435–444, 1983.

[12] Martin Jägersand. Saliency maps and attention selection in scale and spatial coordinates: An information theoretic approach. In *Fifth International Conference on Computer Vision*, pages 195–202. IEEE Computer Society Press, June 1995.

[13] Peter Johansen. On the classification of toppoints in scale space. *Journal of Mathematical Imaging and Vision*, 4:57–67, 1994.

[14] J. J. Koenderink. The structure of images. *Biological Cybernetics*, 50:363–370, 1984.

[15] L. M. Lifshitz and S. M. Pizer. A multiresolution hierarchical approach to image segmentation based on intensity extrema. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(6):529–541, 1990.

[16] A. Oomes and P. R. Snoeren. Structural information in scale-space. In Peter Johansen, editor, *Proceedings of the Copenhagen Workshop on Gaussian Scale-Space Theory*, pages 48–57, Universitetsparken 1, DK-2100 Copenhagen, Denmark, May10–13 1996. Technical Report DIKU-96/19 ISSN 01078283.

[17] A. H. J. Oomes and E. L. J. Leeuwenberg. A scale-space measure of image complexity. *Perception*, 23:20–21, 1994. supplement.

[18] Edward Ott. *Chaos in dynamical systems.* Cambridge University Press, 1994.

[19] Alfréd Rényi. On the foundations of information theory. In Pál Turán, editor, *Selected Papers of Alfréd Rényi*, volume 2, pages 304–318. Akadémiai Kiadó, Budapest, 1976. (Originally: Rev. Inst. Internat. Stat., 33, 1965, p. 1–14).

[20] Alfréd Rényi. Some fundamental questions of information theory. In Pál Turán, editor, *Selected Papers of Alfréd Rényi*, volume 3, pages 526–552. Akadémiai Kiadó, Budapest, 1976. (Originally: MTA III. Oszt. Közl., 10, 1960, p. 251–282).

[21] Manfred Schroeder. *Fractals, Chaos, Power Laws – Minutes from an infinite paradise.* W. H. Freeman and Company, 1991.

[22] Heinz Georg Schuster. *Deterministic Chaos.* VCH Verlagsgesellschaft mbH, 1989.

[23] C. E. Shannon and W. Weaver. *The Mathematical Theory of Communication.* The University of Illinois Press, Urbana, Chicago, London, 1949.

[24] Jeffrey S. Simonoff. *Smoothing Methods in Statistics.* Springer Series in Statistics. Springer-Verlag, 1996.

[25] Jon Sporring. The entropy of scale-space. In *Proceedings of 13'th International Conference on Pattern Recognition (ICPR'96)*, Vienna, Austria, August 1996.

[26] Jacques Levy Vehel and Pascal Mignot. Multifractal segmentation of images. *Fractals*, 2(3):379–382, June 1994.

[27] J. Weickert. *Anisotropic Diffusion in Image Processing.* PhD thesis, University of Kaiserslautern, Department of Mathematics, Kaiserslautern, Germany, January29 1996.

[28] J. Weickert, S. Ishikawa, and A. Imiya. On the history of gaussian scale-space axiomatics. In Jon Sporring, Mads Nielsen, Luc Florack, and Peter Johansen, editors, *Gaussian Scale-Space Theory.* Kluwer Academic Publishers, The Netherlands, 1997. (in print).

[29] N. Wiener. *Cybernetics.* Wiley, New York, 1948.

[30] A. P. Witkin. Scale–space filtering. In *Proc. 8th Int. Joint Conf. on Artificial Intelligence (IJCAI '83)*, volume 2, pages 1019–1022, Karlsruhe, Germany, August 1983.