On Generalized Entropies and Scale-Space *

Jon Sporring

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

Joachim Weickert

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

Technical Report: DIKU–96/37, ISSN 0107–8283
http://www.diku.dk

March 12, 1997

*Technical Report No. 96/37, ISSN 0107-8283 – Department of Computer Science, University of Copenhagen Universitetsparken 1, DK-2100 Copenhagen, Denmark – http://www.diku.dk

Contents

Chapter 1

Introduction

It has been suggested by Jägersand [13], Oomes and Snoeren [17], and Sporring [27] that the entropy measure can be used in image processing as a tool to analyse images in the Scale-Space of Iijima [31], Witkin [33], and Koenderink [15].

The entropy measure was initially proposed independently by Shannon [25] and Wiener [32] as a measure of the information content per symbol emitting from a stochastic information source. Later Rényi [21] 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 [30] 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 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 [16, 14, 7]. Secondly, generalized entropy is the basis for the theory of Multi-Fractal [11, 18] and it is known that there are very strong algebraic similarities to the fundamental equations of Statistical Mechanics. These are thus well known functions, and while images are not physical systems in classical thermodynamic sense, Linear Scale-Space is governed by the Linear Heat Diffusion Equation, and one could thus without great difficulty extend the view of images to be a classical thermodynamical system for which the Linear Heat Diffusion is valid. Such a system is an ideal gas. These interpretations of images will be discussed in detail in this chapter. Finally, as will be demonstrated the generalized entropies offer practical, mathematical well founded functions to study scaling behaviors of images for scale-selection and texture analysis.

Related to this work is Vehel *et al.* [29], where images are studied in the multi-fractal setting, focusing on certain dimensions, and Brink & Pendock [6], and Brink [5] 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 and linear entropy. Then, in Section 3 will we discuss the generalized entropies, what the difference is to linear entropy, and what their properties are in Scale-Space. Following this, in Section 4 we will discuss a physical interpretation of images both from the view of Multi-Fractals and Thermodynamics. Finally, in Section 5 we will give demonstrations of the applicability of this theory to image processing.

Chapter 2

Information Theory and Scale-Space

The information theoretic entropy of a discrete distribution is defined by Shannon [25] as,

$$S(p) = -\sum_{i=1}^{N} p(i) \log p(i).$$

We will in this paper use the natural logarithm, but it is not of great importance which unit of information is used as long as one is consistent. One interpretation of images is to view them as spatial distributions,

$$p(i) = \frac{I(i)}{\sum_{x \in \Omega} I(\mathbf{x})},$$

where $\Omega \in \mathbb{R}^N$ is the domain of the image and N is the dimensionality. This distribution describes the probability of a photon hitting a certain spatial point. Although there is a relation, this distribution should *not* be confused with the distribution of intensities, sometimes called *the* histogram.

Extending this view with the Linear Scale-Space paradigm [33, 15, 8, 31] enables one to explicitly study the effect of discretization. The image is extended with the scale or time parameter t by imposing a diffusion process according to

$$\partial_t = \partial_{x_i} \partial_{x_i},$$

where Einstein's summation convention is implied, i.e. the right side is a sum over all dimension variables x_i. The Green's function for this process is the Gaussian Kernel, and the process can thus be modeled by convolution

$$p_t = G_t * p = \int G_t(x - \alpha) p(\alpha) \; d\alpha,$$

where

$$G_t(\mathbf{x}) = \frac{1}{(\pi t)^{|x|/2}} e^{\frac{-x_i x_i}{t}},$$

again implying the summation convention. $|x|$ is the number of dimensions. Note that the standard deviation is given as, $\sigma = \sqrt{t/2}$. Even if the distribution prior to the Scale-Space extension was discrete due to the discreteness of the image, it is now continuous. The entropy can also be defined for continuously defined distributions, but as the following simple calculation demonstrates, the continuously defined entropy can yield negative values and is thus difficult to interpret in terms of information content. To see this, write the continuous Entropy of an 1D Gauss function as,

$$
\begin{aligned}
S(G_t) &= -\int_{-\infty}^{\infty} G_t(x) \log G_t(x)\, dx \\
&= -\int_{-\infty}^{\infty} G_t(x) \frac{x^2}{t}\, dx + \frac{1}{2}\log \pi t \\
&= -\int_{-\infty}^{\infty} G_t(x)(H_2(\frac{x}{\sqrt{t/2}}) + 1)\, dx + \frac{1}{2}\log \pi t \\
&= -\int_{-\infty}^{\infty} t\frac{\partial^2 G_t(x)}{\partial x^2} + G_t(x)\, dx + \frac{1}{2}\log \pi t \\
&= -1 + \frac{1}{2}\log \pi t,
\end{aligned}
$$

where H_2 is the second order Hermitian polynomial [28, p. 28], and using the fact that the integral of a Gaussian is one and the integral of any derivative of a Gaussian is zero. Negative entropy is found for $t < e/\pi$.

This present work will therefore investigate the following discretization procedure, also known as a histogram,

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

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 [26]. To simplify we will further approximate the integral as,

$$p_t(i) \simeq \frac{hp_t(ih)}{\sum_j hp_t(jh)} \tag{2.1}$$

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, seems appropriate, 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

$$h = c\sqrt{t} \tag{2.2}$$

Further, averaging the entropy over all grid offsets we find [27],

$$\langle S_t(p)\rangle \simeq -c - \frac{1}{2}\log t - \int_{\mathbf{x}\in\Omega} p_t(\mathbf{x}) \log p_t(\mathbf{x}) dx.$$

This is thus the entropy as a function of scale. The reader should note (at least) two points. Firstly, the introduction of the integral is due to the fact that the averaging operates on continuous distributions but avoids the problems of negative entropies by including the discretization explicitly. Secondly, this defines a transformation of the scale parameter t such that the information loss in terms of the entropy is constant, and it should thus be noted that this refines the natural scale parameter $\log t$ of Koenderink and Florack [15, 8] with an data dependent term.

For the rest of this paper we will study the scale behavior as a function of $\log t$. There is thus little information in the first two terms of the above equation, and we will restrict ourselves to examining the last term: $\langle S_t(p)\rangle + c + \frac{1}{2}\log t$. An example of this function is given in Figure 2.1. As will be discussed later, this has been proven by two independent approaches [27, 30] to be a monotonically growing function of scale t, and thus may well serve as a measurement of Koenderink's causality [15].

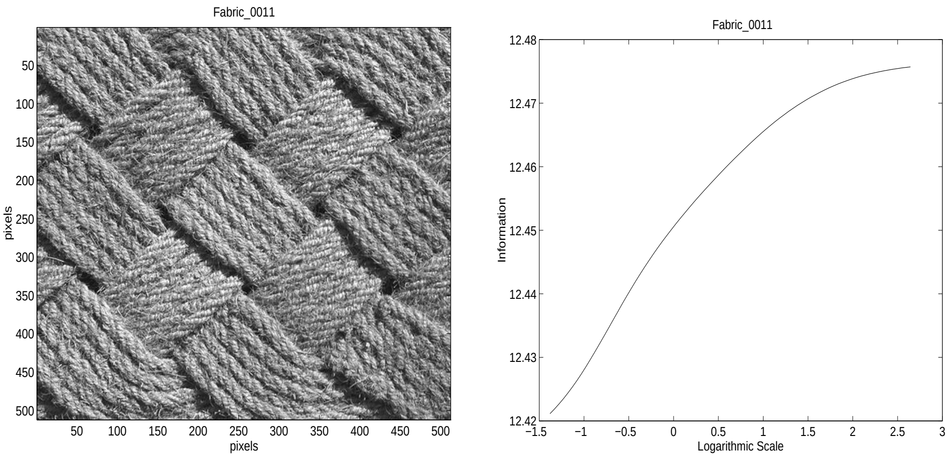

Figure 2.1: An examples of an entropy function: Left is a 512×512 grey-valued image on which Scale-Space has been applied resulting in the Entropy function shown to the right. Actually only $\langle S_t(p) \rangle + c + \frac{1}{2} \log t$ is shown.

Chapter 3

Generalized Entropies

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.

In the following will the axioms of information theory be reviewed, and the concept of Scale-Space in this respect will be discussed in detail.

3.1 The Axioms of Information Theory

To review the axiomatically derivation of the Generalized Entropies we will briefly review the historically developments of the axioms leading first to Shannon's entropy [32, 25] and finally to Rényi's generalized entropies [21, 20].

The theory of information was born by Hartley's establishment of the additivity axiom [10]:

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

It be argued that information is a logarithmic notion. Shannon's contribution was to define the entropy as the linear mean of information:

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

Finally Rényi [21, 20] 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 general entropy of order $\alpha \in \mathbb{R}$ is thus defined as minus the logarithm of the expected $(\alpha - 1)$-norm,

$$S_\alpha(p) = \frac{1}{1-\alpha} \log \sum_i p(i)^\alpha \tag{3.1}$$

for $\alpha \neq 1^*$.

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

3.2 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 general, the generalized entropy of order α is the logarithm of the $\alpha-1$ regular moment of the distribution p under the same distribution divided by $1-\alpha$. The regular moments $m_\alpha = \sum_i p(i)^\alpha$ compare to the central moments as,

$$\begin{aligned} c_\alpha &= \sum_i (p(i)-\overline{p})^\alpha \\ &= m_\alpha + \binom{\alpha}{1}\overline{p}m_{\alpha-1} + \binom{\alpha}{2}\overline{p}^2 m_{\alpha-2} + \cdots + \binom{\alpha}{\alpha}\overline{p}^\alpha, \end{aligned}$$

where $\overline{p} = m_1/N$. For $\alpha = 2$, the relation between moments and centralized moments (the variance) take a particular simple form: $c_2 = m_2 - \overline{p}^2$. The information measure is thus,

$$S_2(p) \simeq -\log c_2 + \mathcal{O}\left(\frac{\overline{p}^2}{c_2}\right).$$

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 the convention of $\lim_{x\to 0} 0^x = 0$ 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, an that the generalized entropy is the basis of the theory of Multi-Fractals. Both these topics will be discussed later.

3.3 The Scale-Space Extension

Returning to the definition of a Scale-Space extended and discretized distribution, we will now examine the behavior of the generalized entropies in Scale-Space. It will be assumed in the following that the discrete distribution is normalized.

Combining Equation 2.1 and 3.1 yields,

$$\begin{aligned} S_{\alpha,t}(p) &= \frac{1}{1-\alpha}\log\sum_i (hp_t(ih))^\alpha \\ &= \frac{\alpha}{1-\alpha}\log h + \frac{1}{1-\alpha}\log\sum_i p_t(ih)^\alpha, \end{aligned}$$

with the choice of h (Equation 2.2) kept in mind. It is easy to see that for $\alpha \to 1$ the Scale-Space extended Generalized Entropy converges to the Scale-Space extended Entropy, thus retaining the properties discussed previously. Again we wish to take the mean generalized entropy over all grid offsets. This time though, it is more complicated. By Jensen's inequality it can be seen that,

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

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 reader should note that the logarithm of the scale again 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 again motivated to study the behavior of $1/(1-\alpha)\log\int_{x\in\Omega} p_t(x)^\alpha \, dx$.

In Figure 3.1 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.

3.4 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 3.4.1. *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 [21, 12]. A generalized Hölder inequality states that for non-negative a and values of a distribution q_i the 'expected norm' $m_r(a) = \left(\sum_i q_i a_i^r\right)^{1/r}$ satisfies

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

for $r > r'$. The generalized entropies are defined as minus the logarithm of the expected norm, and since the logarithm is a monotonic function,

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

for $\alpha > \alpha'$. □

Proposition 3.4.2. *Let $p_i(t) > 0$ for all i and for all $t > 0$. Then the generalized entropies*

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

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 I_0.

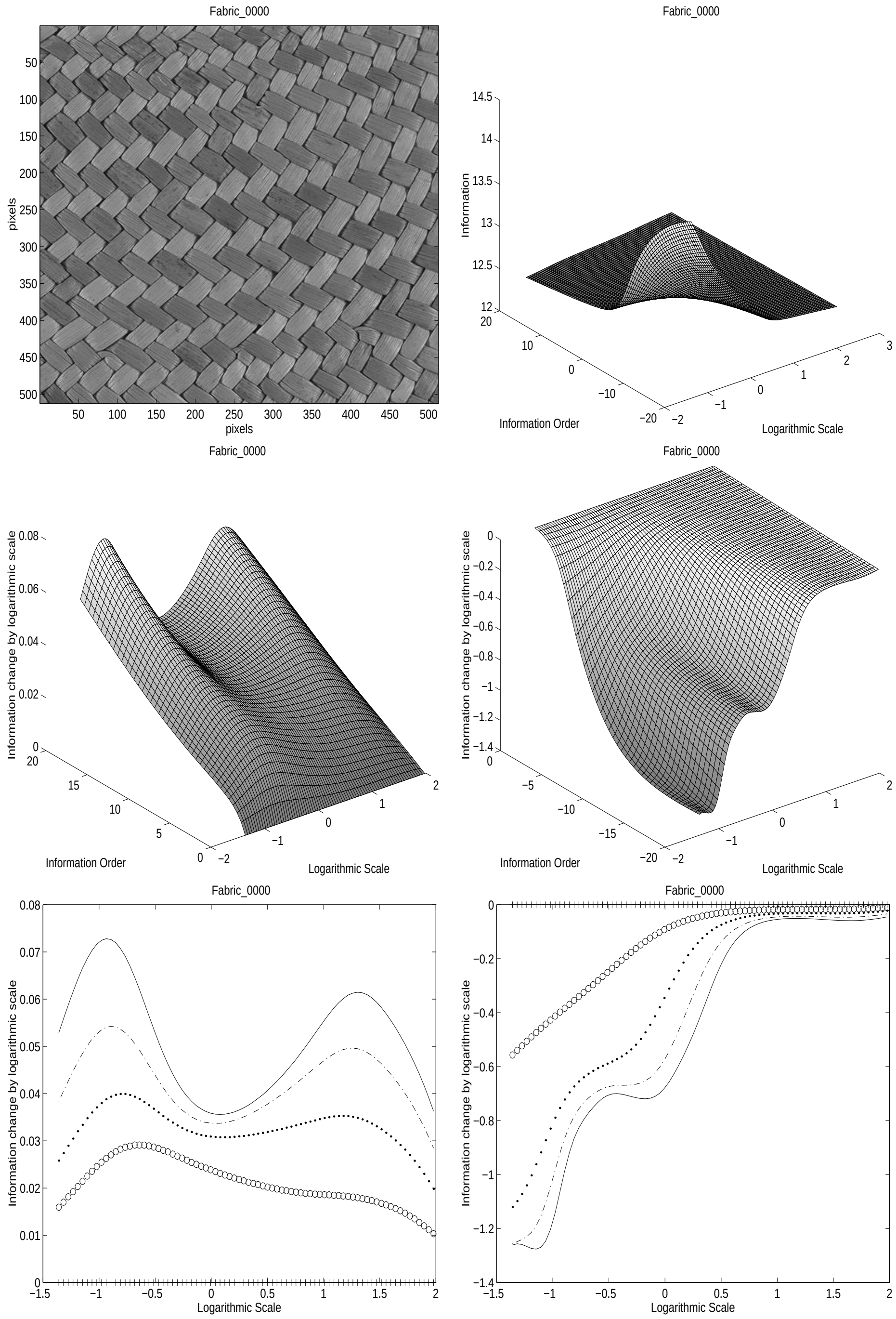

Figure 3.1: An example of a generalized entropy functions. Top left shows the image. Top right shows the generalized entropy as a function of order and logarithmic scale. As described in the text, the generalized entropy function is monotonic in both variable. Hence does the middle row show the derivatives with respect to logarithmic scale. The figure is split into positive and negative information orders. Finally, to emphasize this view, the bottom row shows constant information order slices of these surfaces. The lines of '+'s are order 0, the lines of 'o's are ± 5, the '.'s are order ± 10, the '$\cdot -$'s are order ± 15, and the full lines are order ± 20 respectively.

Proof. The proof is based on results from [30] 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 [30, p. 68] for full details.

First we prove the monotony of I_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

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

is increasing in t.

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

$$I_1(P(t)) = -\sum_i p_i(t) \ln p_i(t).$$

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 [30, 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, $\sum_i p_i^\alpha$ and $I_\alpha(P(t))$ are increasing.

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

$$I_0(P(t)) = \ln \sum_{i=1}^{N} p_i^0(t) = \ln N = \text{const.}$$

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 $I_\alpha(P(t))$ 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

$$\lim_{t\to\infty} I_\alpha(t) = \frac{1}{1-\alpha} \ln \sum_{i=1}^{N} \frac{1}{N^\alpha} = \ln N = I_0,$$

and $\alpha = 1$ yields

$$\lim_{t\to\infty} I_1(t) = -\sum_{i=1}^{N} \frac{1}{N} \ln \frac{1}{N} = \ln N = I_0.$$

This completes the proof. □

Proposition 3.4.3. *Let $p_i(t) > 0$ for all i and for all $t > 0$. Then the generalized entropies $I_\alpha(P(t))$ 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 I_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

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

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

Chapter 4

Information Theory, Heat Diffusion, Thermodynamics, and MultiFractals

Scale-Space is a model on the imaging device, which coincidently can be described as Heat Diffusion. On the other hand, information theory and thermodynamics share common ground through the field of statistical mechanics. Finally, a relative new field of multi-fractals has been developed which provides a fractal interpretation of the generalized entropies. These issues will be discussed in detail in the following.

4.1 The Phase-Space Model

Consider an image to be an ideal gas, such that each spatial position in the image denotes the number of atoms within the aperture. In the ideal case, the gas is so diffuse that no atoms collides, and the interaction can then be neglected. That leaves us $2n$ degrees of freedom per atom, where n is the dimension of the space: The position vector and the impulse moment vector. This is also known as the phase space in classical mechanics [22]. In quantum mechanics every measurement of a system is limited by the Heissenberg's Uncertainty Principle, and in the case of images, a pragmatic extension of this principle is to assume large (infinite) uncertainty of the impulse vectors, and an uncertainty corresponding to the aperture of the position vectors. Conversely, one could also assume that each atom has identical energy. In both cases is it fairly easy to show that the statistical mechanic entropy of the physical system corresponds to the information theoretic entropy of the above mentioned spatial distribution [27]. To resume, given the multiplicity function,

$$W = \frac{N!}{\prod_i n_i!}, \tag{4.1}$$

where n_i is the number of molecules in the phase-space state i, and $\sum_i n_i = N$. Using the Stirling approximation

$$\log N! \simeq N \log N - N,$$

the Boltzmann entropy can be found as,

$$S = k \log W = -kN \sum_i \frac{n_i}{N} \log \frac{n_i}{N} = -kN \sum_i p_i \log p_i.$$

Furthermore, since the Heat-Equation is valid for such a system, it is by the Second Law of Thermodynamics given that the entropy is a monotonically growing function of time [27, 30], and finally, since the Scale-Space by Koenderink and Witkin [33, 15, 31] is such a system, the increase in the entropy of the spatial distribution will be a monotonically growing function of scale.

4.2 A short course on Statistical Thermodynamics

To extend our view to generalized entropies, let us first review the basic machinery of Statistical Mechanics. Assume an assembly of N identical systems, e.g. N gas molecules. Each system can be in one of a given number of states, e.g. somewhere in phase-space. Given a certain population of the states, the multiplicity function W (Equation 4.1) is the the number of realizations of the N systems with such a population. To end the counting, be reminded that for each such realization the number of systems is constant,

$$\sum_i n_i = N, \tag{4.2}$$

where n_i is the number of occupants in the i'th state, and likewise is the total energy,

$$\sum_i n_i \epsilon_i = E \tag{4.3}$$

where ϵ_i is the energy of the i'th state.

For large systems, i.e. large N's, the realizations which maximize W will be totally dominating in the ensemble of all configurations complying with Equations 4.2 and 4.3. This is the method of Most Probable Configuration [22]. This point in $\log W$ is found as the maximum of,

$$\log W - \lambda \sum_i n_i - \beta \sum_i \epsilon_i n_i,$$

yielding

$$n_i = e^{\lambda - \beta \epsilon_i}.$$

The λ term is usually ignored since all thermodynamic functions can be written as functions of N, i.e.

$$n_i = N \frac{e^{-\beta \epsilon_i}}{\sum_i e^{-\beta \epsilon_i}} = N \frac{e^{-\beta \epsilon_i}}{q} = N p_i,$$

where q is called the partition function and p_i is the Boltzmann distribution. Also it is in thermodynamic experiments recognized that

$$\beta = \frac{1}{kT},$$

where k is the Boltzmann constant.

Remarkably, everything there is to know about the thermodynamic state variables of a system from its mechanical description can be found through the analysis of $k \log q$, e.g. the entropy can be written as,

$$S = \frac{U(T) - U(0)}{T} + Nk \log q = -\frac{N}{T} \frac{\partial k \log q}{\partial \beta} + Nk \log q,$$

where U is the internal energy. For further reading, see e.g. [3, 4, 22].

4.3 The Helmholtz Energy

It has been noted by several authors (e.g. [23, p. 206] and [24, p. 130]) that the generalized entropies in the limit of very large 'images' are very similar to the expression of the Helmholtz energy,

$$A(T) = U(T) - TS(T) = -kT \log q = -\log \left(\sum_i \left(e^{-\epsilon_i} \right)^{\beta} \right)^{\frac{1}{\beta}},$$

where U is the internal energy. But, as the reader may well have noted, this is *not* the generalized entropy of the Boltzmann distribution in general, but the generalized entropy of $\exp(-\epsilon_i)/\sum_i \exp(-\epsilon_i)$, i.e. the energy distribution at the specific temperature $T = 1K$ (Kelvin) (for $\beta \gg 1$).

To be a little more specific, the generalized entropy of the Boltzmann distribution is given as,

$$\begin{aligned} S_\alpha &= \frac{1}{1-\alpha}\log\sum_i p_i^\alpha \\ &= \frac{1}{1-\alpha}\log\sum_i\left(\frac{e^{-\beta\epsilon_i}}{\sum_i e^{-\beta\epsilon_i}}\right)^\alpha \\ &= \frac{1}{1-\alpha}\log\sum_i\left(e^{\beta\epsilon_i}\right)^\alpha - \frac{N\alpha\log q}{1-\alpha}. \end{aligned}$$

Given a remapping of the energy levels as $\zeta_i = \beta\epsilon_i$, the information order may now play the role of inverse temperature (for large α's). But note, this is not a thermodynamic temperature as such, since for the view to be consistent, the information theoretical entropy as the limit of the generalized entropy order going to one will be independent on this new temperature, while the thermodynamic entropy is very dependent on the thermodynamic temperature. When $\alpha \gg 1$ the relation is approximately given by,

$$S_\alpha \approx \frac{1}{1-\alpha}\log\sum_i e^{-\alpha\zeta_i} + \text{const.}, = \frac{1}{1-\alpha}A(T/\alpha) + \text{const.},$$

where the constant term is constant with respect to time or scale.

To relate all this to images, we must first consider the concept of temperature. In thermodynamics there is a distinct difference between heat and temperature. Heat is in the absence of mechanical work equivalent to internal energy (U), while temperature is something that is usually measured with a thermometer. The difference can be found through the heat capacity: Some physical objects such as metal require much energy to be raised to a certain temperature, i.e. they feel cold when touched, while others require little. This property is called the heat capacity and relates heat and temperature as, $dQ = CdT$ at constant volume. We are now ready to imagine a system placed in a heat bath such that the temperature can be regulated. We can then imagine an image to be a system where the energy levels corresponds to a physical system at temperature 1K. This then enables us to view the generalized entropies as Helmholtz energy where the information order is regulated through a hypothetical heat capacity and the heat of the surrounding bath.

4.4 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. The theory of multi-fractals will be reviewed below.

A fractal is a self similar object, i.e. in a zooming sequence the greater 'image' can be found replicated at a smaller scale. A behavior that can be found in e.g. the Mandelbrot set. 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 will be given a summary of multi fractal theory [11, 23, 24]. 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 i 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

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

*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.

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 for each q associate an α. Reformulating this in terms of α using that the density of scaling exponents α' can be written as [11],

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

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

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

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 $\partial_{\alpha^2} \tau > 0$, and this leads to

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

and

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

Finally define the multi-fractal dimension [12],

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

This measure has been shown to unify a great variety of fractal measures used in the past [9], e.g. the Hausdorff dimension can be shown to be D_0, D_1 as 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 4.1

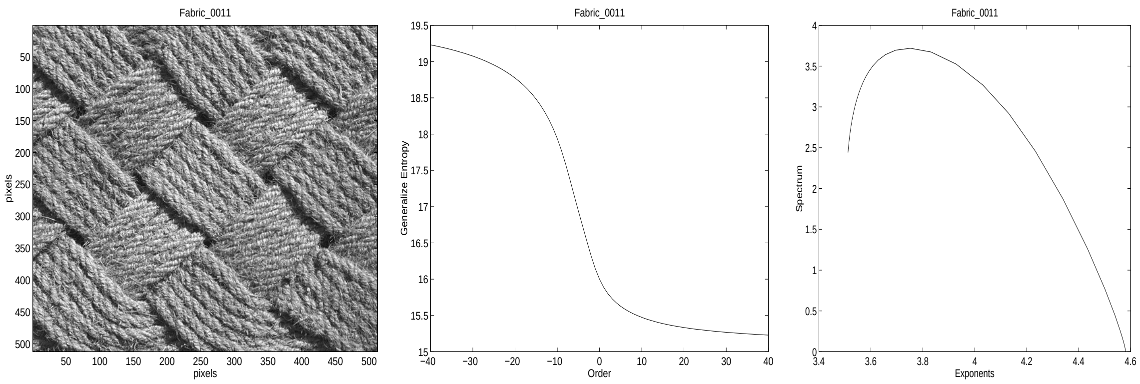

Figure 4.1: In this figure are the generalized entropy (middle) and the spectrum (right) of the left image shown. To calculate the spectrum we used $l = \sqrt{12}/512$ As the image simplifies with scale, the generalized entropy converges to a constant function, and the spectrum converges to a point function. Note also that the derivative of the spectrum is equal to the order, and further that at the point of derivative the spectrum is equal to the D_0.

Chapter 5

Application of the Generalized Entropy in Image Analysis

It has been demonstrated in a previous article [27] that there is a simple relationship between the point of maximum global (Shannon) entropy and the absolute size of simple textures. The generalized entropies gives us a second view on the global scaling behavior of images in Scale-Space. This is best discussed in the terminology of fractal spectra. The isophote of an image is proportional to l^α, as such $f(\alpha)$ is the fractal (Hausdorff) dimension of the $\log(I) = \alpha_0$ manifold. This is a monotonic transformation of the isophotes, and in Scale-Space, the isophotes evolve with different speeds and (fractal) complexities. In general the spectrum will be a smooth, concave and positive function for all α's between $\log \max_x I$ and $\log \min_x I$, and the width of this span will fall monotonically in Scale-Space, such that the spectrum will converge to a point or a Dirac delta function around the exponent of the mean intensity value. In other words, viewing the generalized entropy as the logarithm to the norm, the information order determines which isophotes are to be emphasized. The scaling behavior of the maximum (and minimum) intensities can thus be seen for ($\pm$) infinite order. Likewise, the scaling behavior of the zeroth order is constant as is the case for the scaling behavior of the spectrum at zero derivative.

To demonstrate this, we have given a few examples from the VisTex package [19]. The two top rows showing pieces of fur. We will now interpret the information change for positive orders in terms of image contents. In both graphs the first node is located at almost same scale for both images and all orders, but the the absolute information change of these nodes differ somewhat, i.e. Fabric.0005 shows a more skewed distribution of intensities than Fabric.0004 at low scale. At high scale, the presence of an extra node in Fabric.0004 indicates medio level structure which is not present in the other.

The bottom two rows showing pieces of baskets should also be compared with Figure 3.1. It should be noted that Fabric.0000 has fewer very light and very dark pixels than Fabric.0002 and Fabric.0003, and this in general causes the range of scales to become greater in the latter cases. Also, all graphs show nodes at approximately -1, which is due to the very small scale structure of the threads, while the the node at approximately 1 is due to the prominent junctions between each 'square'. Only Fabric.0002 and Fabric.0003 show very large scale structure which can be accredited the illuminance differences.

Finally, in Figure 5.2 and 5.3 we include some more examples of entropy changes for the reader to enjoy and digest.

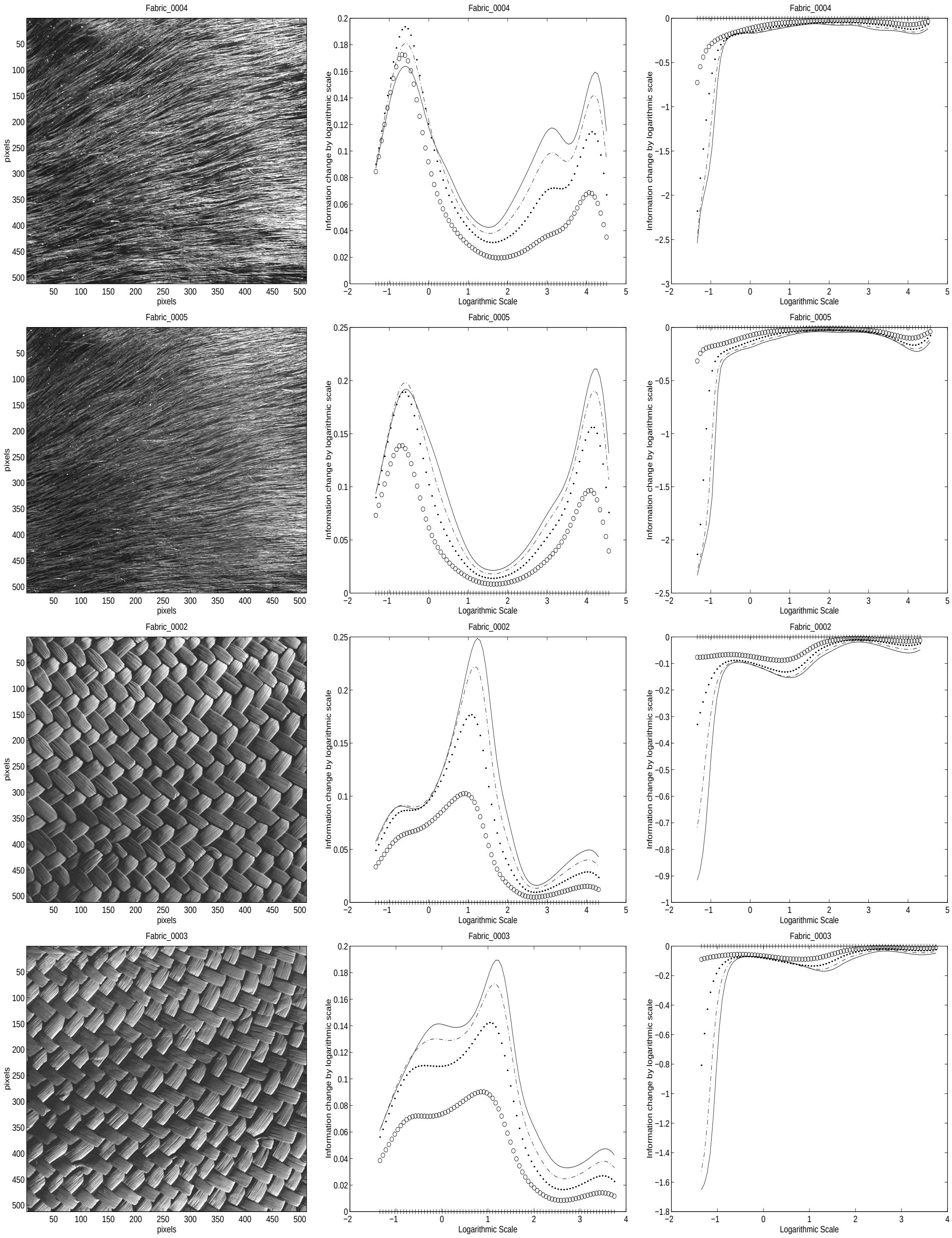

Figure 5.1: A demonstration of the variety of scaling behaviors of generalized entropy function. The left column are the images, the middle and the right are entropy change per log scale of positive and negative orders. The same lines corresponds to the same orders as in Figure 3.1

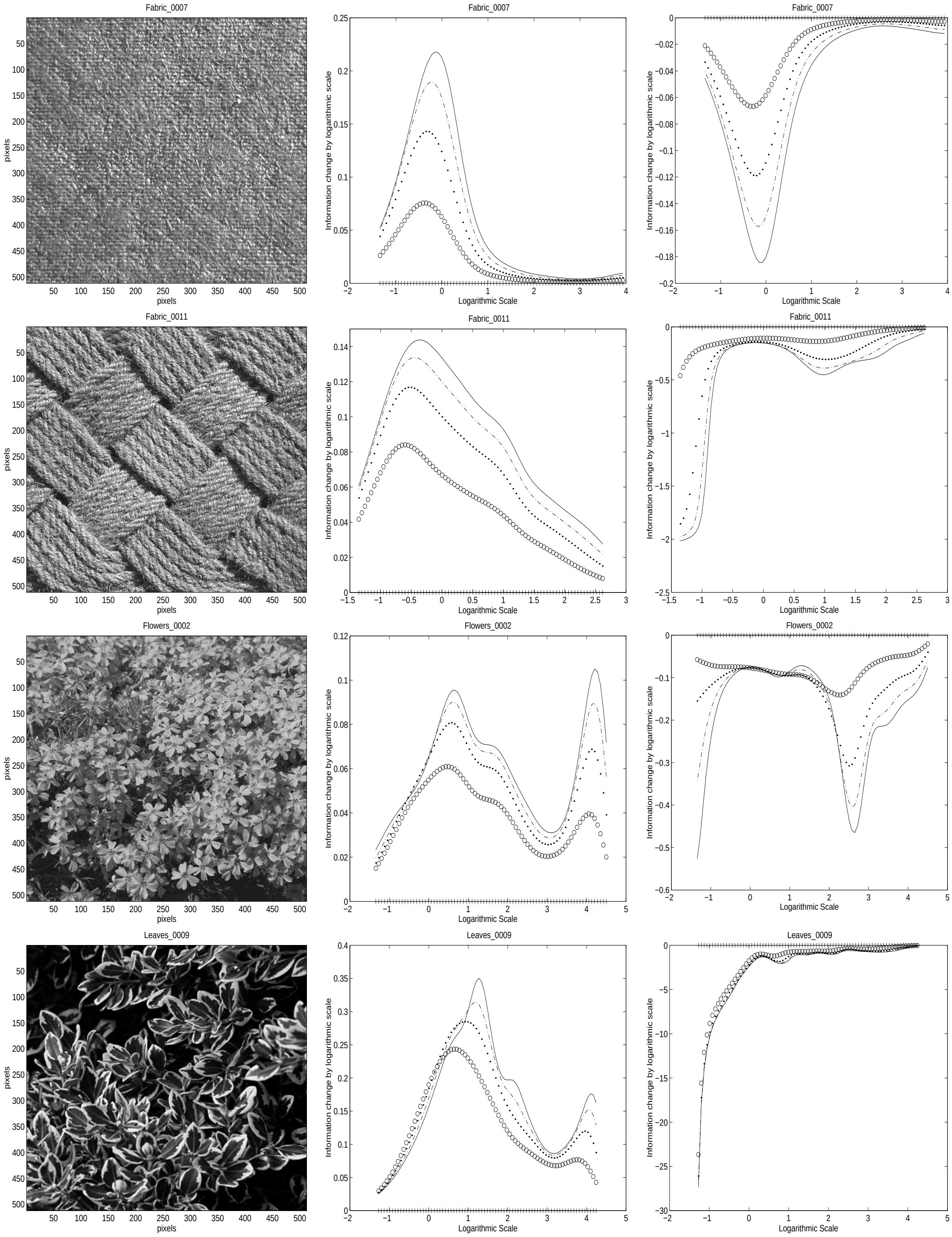

Figure 5.2: More texture images and their entropy changes. See Figure 5.1 for an explanaition of the graphs.

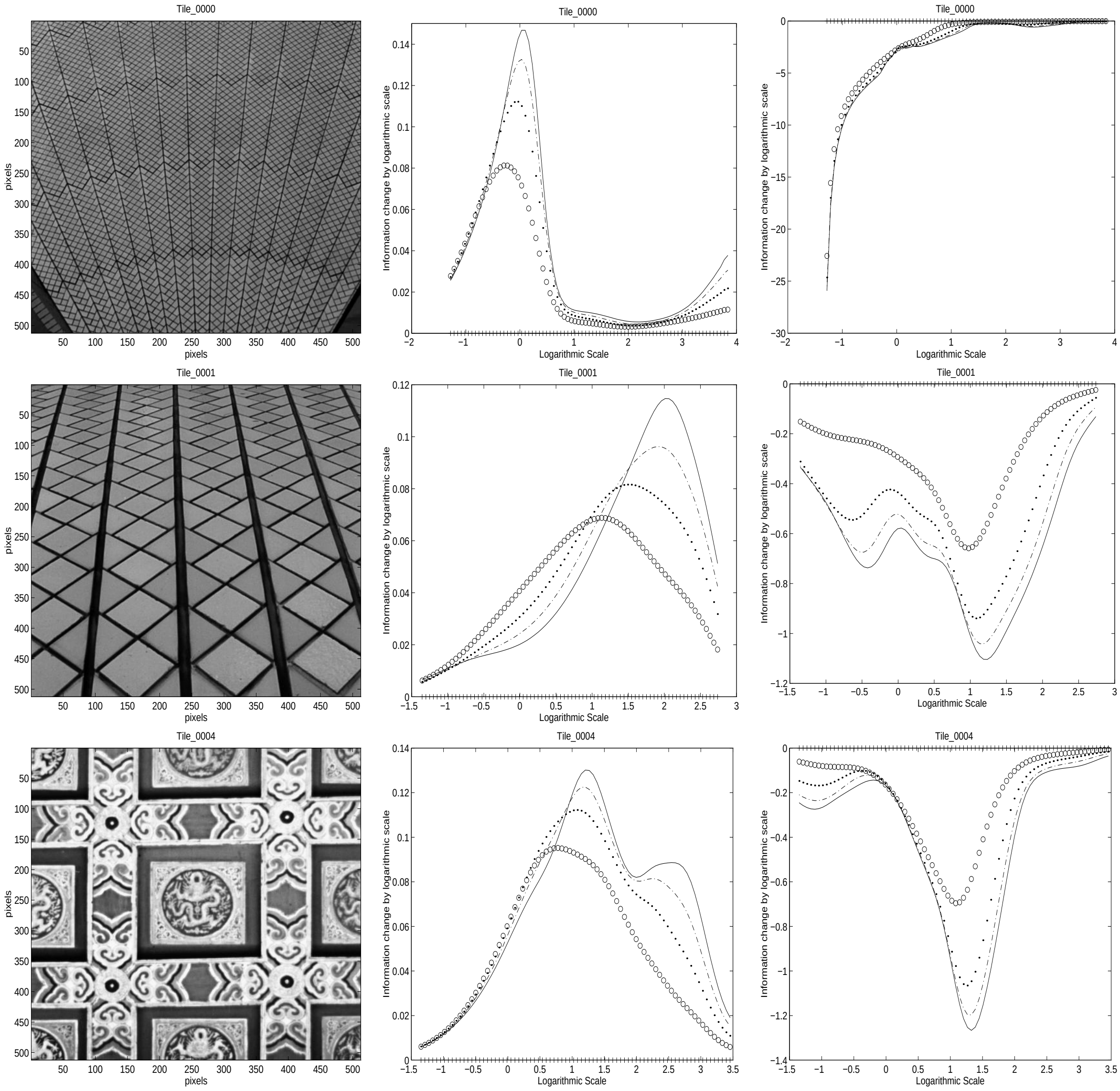

Figure 5.3: Yet more texture images and their entropy changes. See Figure 5.1 for an explanaition of the graphs.

Chapter 6

Programming Entropies in Matlab

With the exception of the Scale-Space implementation, all the experiments make use of programs written in the Matlab language. We have chosen to include the source code in the following for two reasons: to document the experiments and to demonstrate the use of Matlab.

Matlab is a matrix language, and the code below is optimized for matrix calculations, still should all functions except the last, scale.m, be understandable and fairly well documented.

All the programs below has been installed in /usr/local/image/src/matlab, and can be used at DIKU by including

```
setenv MATLABPATH /usr/local/image/src/matlab
```

in your .tcshrc file.

6.1 spectrum.m

'Spectrum' calculates the multifractal spectrum of an image. As you may note, it is only a wrapping of the two functions below.

```
function [F,A] = spectrum(I,t,q);
%SPECTRUM Calculate multifractal spectrum as function of exponents and scale
%
%       [F,A] = spectrum(I,s,q)
%         F - The multifractal spectrum
%         A - The corresponding scaling exponents
%         I - The image
%         t - The list of scales used (2*variance)
%         q - The list of information order used
%
%       This function evaluates the multifractal spectrum from the
%       image.

L = information_scale(I,t,q);
[F,A] = inf2spect(L,sqrt(12)/size(I,1),q);
```

6.2 inf2spect.m

Given an generalized entropy as function of scale and order, 'inf2spect' calculates the spectrum and the corresponding exponents.

```
function [F,A] = inf2spect(L,l,q);
%INF2SPECT Calculate multifractal spectrum as function of exponents and scale
```

```
%
%       [F,A] = inf2spect(L,l,q)
%         F - The multifractal spectrum
%         A - The corresponding scaling exponents
%         L - The generalized entropy (from e.g. information_scale)
%         l - The list of lengths used in the measurement usually (const. 1/N)
%         q - The list of information order used
%
%       This function evaluates the multifractal spectrum from the
%       generalized entropy

% The generalized dimensions
D = -L./log(l);

% The scaling exponents
T = ((q'-1)*ones(1,size(L,1)))'.*D;
A = (T(:,3:size(q,2))-T(:,1:size(q,2)-2))/(q(3)-q(1));
dq = q(2:size(q,2)-1);

% The spectrum
F = (dq'*ones(1,size(L,1)))'.*A - T(:,2:size(T,2)-1);
```

6.3 information_scale.m

To generate the generalized entropy as function of scale and order has 'information_scale' been used. It is optimized together with 'scale' such that the image is only Fourier Transformed once.

```
function L = information_scale(I,s,q);
%INFORMATION_SCALE Calculate the information as function of order and scale
%
%       L = information_scale(I,s,q)
%         L - The information as function of order and scale
%         I - The image
%         s - The list of scales (2*variance)
%         q - The list of information order
%
%       This function evaluates the generalized entropy or information for
%       an image.  The rows corresponds to the scales and the columns to the
%       information order. Be warned: This is a slow function!

L = zeros(size(s,2),size(q,2));
FI = fft2(extend(I,2,2));
for i = 1:size(s,2)
  I2 = real(ifft2(scale(FI,sqrt(s(i)/2),0,0)));
  I2 = I2(1:size(I,1),1:size(I,2));
  L(i,:) = information(I2/sum(sum(I2)),q);
end
```

6.4 information.m

'Information' is the procedure used to calculate the generalized entropy of a distribution. In a neighborhood around order 1 has a Taylor expansion to first order for both the enumerator and denominator been used such that the limit of order going to 1 will cause the generalized entropy to converge to Shannon's entropy.

```
function E = information(P,q);
%INFORMATION The information theoretic generalized entropy
%
%       E = information(P,q)
%         E - the entropy of order q (in bits, i.e. log2 is used)
%         P - a 1D or 2D distribution
%         q - the order, if q is a matrix then a matrix is returned
%             where each point is the corresponding generalized
%             entropy.
%
%       This function calculates the information of order q of a 1D or
%       2D distribution P, i.e. for every point 0 <= P(x) <= 1,
%       and sum(sum(P)) = 1. The log2 unit is used.  This measure is
%       not defined for q = 1, but it can be shown, that the limit for
%       q going to 1 is Shannon's Entropy.  Thus a first order
%       approximation is used close to 1.  At infinum the logarithm of
%       the max- or min-norm is returned
%
%       The complexity of the algorithm is O(n*m) where n and m are the
%       dimensions of P.
%
%       To calculate the entropy of a uniform distribution
%       (as specified) use,
%         E = information(ones(10,10)/100,0);
%
%                                           Jon Sporring, June 13, 1996

E = zeros(size(q));
for i = 1:size(q,1)
  for j = 1:size(q,2)
    if q(i,j) == Inf
      % Essentially the logarithm of the max-norm
      E(i,j) = -log2(max(max(P)));
    else
      if q(i,j) == -Inf
        % Essentially the logarithm of the min-norm
        E(i,j) = -log2(min(min(P+(P<=0))));
      else
        Pq = P.^q(i,j);
        X = sum(sum(Pq));
        if abs(q(i,j)-1) < 0.5
          % E(P,q) -> entropy(P) for q -> 1
          E(i,j) = -sum(sum(Pq.*log2(P+(P<=0))))/X;
        else
          E(i,j) = log2(X)/(1-q(i,j));
        end
      end
    end
  end
end
```

6.5 scale.m

Although this implementation has been abandoned in an early state of our experiments, this implementation of Scale-Space could have been used. It implements Scale-Space in the Fourier Transform, and is optimized in the sense that the Gaussian function is transformed analytically and only calculated in one quadrant. Be warned, this is not a simple algorithm to read, but it is about the fastest that can be done for Fourier Transform implementations, which again is the fastest possible Scale-Space implementation for large scales in general.

```
function Is = scale(I,s,dr,dc);
%SCALE  Gaussian Scale-Space using the Fourier Domain
%
%       Is = scale(I,s,dr,dc)
%         I - the Fourier transform of a matrix or an image
%         s - the standard deviation of the Gaussian (must be larger than 0)
%         dr - the derivative order in the direction of the rows
%         dc - the derivative order in the direction of the columns
%         Is - the Fourier transform of the scale image/matrix
%
%       This is an implementation of the Gaussian scale space on matrices.
%       The convolution is implemented in the Fourier domain and for that
%       reason the number of rows and columns of the matrix must be powers
%       of 2.
%       Fractional valued dr and dc are possible, but be warned the result
%       will probably be complex.
%       The complexity of this algorithm is O(n) where n is the total number
%       of elements of I.
%
%       To calculate an image of scale (variance) 2^2 use,
%         I2 = real(ifft2(scale(fft2(I),2,0,0)));
%       To derive an image once in the direction of rows at scale 1^2 do,
%         I2 = real(ifft2(scale(fft2(I),1,1,0)));
%
%                                             Jon Sporring, January 1, 1996

if s == 0
  Is = I;
else
  if (s < 0)
    error('s must be larger than zero');
  else
    rows = size(I,1);
    cols = size(I,2);

    if(rem(log2(rows),1) ~= 0 | rem(log2(cols),1) ~= 0)
      error('The image must have side lengths of power of 2');
    else
      % Calculate the Fourier transform of a gaussian fct.
      G = zeros(rows,cols);
      if (rows > 1) & (cols > 1)
        G(1:rows/2+1, 1:cols/2+1) = exp(-(([0:rows/2]'*ones(1,cols/2+1)/(rows-1)).^2 ...
 + (ones(rows/2+1,1)*[0:cols/2]/(cols-1)).^2)*(s*2*pi)^2/2);
      else
        if rows > 1
          G(1:rows/2+1, 1) = exp(-([0:rows/2]'/(rows-1)).^2*(s*2*pi)^2/2);
```

```
        else
          G(1,1:cols/2+1) = exp(-([0:cols/2]/(cols-1)).^2*(s*2*pi)^2/2);
        end
      end
      G(rows/2+1:rows, 1:cols/2+1) = flipud(G(2:rows/2+1, 1:cols/2+1));
      G(1:rows/2+1, cols/2+1:cols) = fliplr(G(1:rows/2+1, 2:cols/2+1));
      G(rows/2+1:rows, cols/2+1:cols) = fliplr(flipud(G(2:rows/2+1, 2:cols/2+1)));
  %    G = G*rows*cols/(sum(sum(G))*2*pi*s^2);

      % Calculate the Differentiation matrix
      j = sqrt(-1);
      if (rows > 1) & (cols > 1)
        x = [0:rows/2-1,-rows/2:-1];
        y = [0:cols/2-1,-cols/2:-1];
        DG = ((x.^dr)'*(y.^dc)*(j*2*pi)^(dr+dc)/(rows^dr*cols^dc));
      else
        if rows > 1
          x = [0:rows/2-1,-rows/2:-1];
          DG = (j*2*pi*x'/rows).^dr;
        else
          y = [0:cols/2-1,-cols/2:-1];
          DG = (j*2*pi*y/cols).^dc;
        end
      end

      Is = I.*G.*DG;
    end
  end
end
```

Chapter 7

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 and Information Theory and the like.

Bibliography

[1] J. Aczel. On mean values. *Bull. Amer. Math. Soc.*, 54:392–400, 1948.

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

[3] Robert A. Alberty. *Physical Chemistry.* John Wiley & Sons, 1987.

[4] P. W. Atkins. *Physical Chemistry.* Oxford University Press, 1990.

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

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

[7] James Damon. Local morse theory for gaussian blurred functions. In Jon Sporring, editor, *Gaussian Scale-Space Theory*, Computational Imaging and Vision. Kluwer, 1996. (In Press).

[8] 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*, 4:325–351, 1994.

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

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

[11] 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.

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

[13] 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.

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

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

[16] 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*, pages 529–540, 1990.

[17] 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*, DIKU report Nr. 96/19, ISSN 01078283, pages 48–57, Universitetsparken 1, DK-2100 Copenhagen, Denmark, May 1996. Department of Computer Science, University of Copenhagen.

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

[19] Rosalind Picard, Chris Graczyk, Steve Mann, Josh Wachman, Len Picard, and Lee Campbell. Vistex. via ftp:whitechapel.media.mit.edu, 1995. Copyright 1995 Massachusetts Institute of Technology.

[20] 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).

[21] 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).

[22] Erwin Schrödinger. *Statistical Thermodynamics.* Dover Publications, Inc., 1989. A reprint of 2nd ed. from Cambridge University Press, 1952.

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

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

[25] Claude E. Shannon and Warren Weaver. *The Mathematical Theory of Communication.* The University of Illinois Press, 1949.

[26] Jeffrey S. Simonoff. *Smoothing Methods in Statistics.* Springer Series in Statistics. Springer, 1996.

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

[28] Bart M. ter Haar Romeny (Ed.), editor. *Geometry-Driven Diffusion in Computer Vision.* Kluwer Academic Publishers, The Netherlands, 1994.

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

[30] J. Weickert. *Anisotropic diffusion in image processing.* PhD thesis, Dept. of Mathematics, University of Kaiserslautern, Germany, 1996.

[31] Joachim Weickert and Atsushi Imiya. On the history of gaussian scale-space axiomatics. In Jon Sporring, editor, *Gaussian Scale-Space Theory*, Computational Imaging and Vision. Kluwer, 1996. (In Press).

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

[33] A. P. Witkin. Scale space filtering. In *Proc. of International Joint Conference on Artificial Intelligence (IJCAI)*, Karlsruhe, Germany, 1983.