Effective Scale: A Natural Unit for Measuring Scale-Space Lifetime

Tony Lindeberg

Computational Vision and Active Perception Laboratory (CVAP)

Department of Numerical Analysis and Computing Science

Royal Institute of Technology, S-100 44 Stockholm, Sweden

Email: tony@bion.kth.se

IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 15, pp. 1068–1074, 1993.

Abstract

We develop how a notion of effective scale can be introduced in a formal way. For continuous signals a scaling argument directly gives that a natural unit for measuring scale-space lifetime is in terms of the logarithm of the ordinary scale parameter. That approach is, however, not appropriate for discrete signals, since then an infinite lifetime would be assigned to structures existing in the original signal. Here we show how such an effective scale parameter can be defined as to give consistent results for both discrete and continuous signals. The treatment is based upon the assumption that the probability that a local extremum disappears during a short scale interval should not vary with scale. As a tool for the analysis we give estimates of how the density of local extrema can be expected to vary with scale in the scale-space representation of different random noise signals, both in the continuous and discrete cases.

Keywords: scale-space, effective scale, scale-space lifetime, discrete smoothing transformations, density of local extrema, multi-scale representation, computer vision, digital signal processing

1 Introduction

When Witkin [19] coined the term *scale-space*, he empirically observed a "marked correspondence" between the perceptual salience of structures in a signal and the lengths of the intervals during which those structures exist in scale-space:

> ... those intervals that survive over a broad range of scales tend to leap out at the eye ...

This observation indicates that the *lifetime of structures*[1] in scale-space may be an important property to take into account when extracting information from signals. However, if we are to base a significance measure on such an entity, it is of crucial importance to measure it in proper units, so that significance values of structures existing at different scales can be appropriately compared. In principle, the ordinary scale parameter t, as obtained from the diffusion equation

$$\frac{\partial L}{\partial t} = \frac{1}{2}\nabla^2 L \tag{1}$$

with initial condition $L(\cdot;\ 0) = f$ defining the scale-space representation $L : R^N \times R_+ \rightarrow R$ of a continuous signal $f : R^N \rightarrow R$ [19, 7, 20, 8], is arbitrary and could be transformed by any strictly increasing change of variables. In other words, given a structure existing during a scale interval $[t_1, t_2]$ in scale-space, we could in principle for any strictly increasing function $h : R \rightarrow R$ imagine defining the lifetime of that structure as $t_{life} = h(t_2) - h(t_1)$.

The goal of this paper is to introduce such a transformed scale parameter, which will be termed *effective scale* and denoted by τ, so that scale-space lifetime can be measured as plain differences of scale values expressed in this unit.

For *continuous* signals the problem is almost trivial. A scaling argument directly implies (see Appendix A.2) that a natural way to define scale-space lifetime is by letting the effective scale parameter be the logarithm of the ordinary scale parameter. This relation is well-known

[1] See Appendix A.1 for a discussion about scale-space lifetime and further motivations to this work.

and has been used for example in pyramid representations, which usually comprise a logarithmic sampling along the scale direction (see e.g. Burt [3] or Crowley [5]).

However, for *discrete* signals the situation is no longer as simple, at least not if we are interested in a scale-space representation with a *continuous scale parameter*[2] [8]. The scaling argument cannot be carried out due to the finite *inner scale* [7] of the sampled data. If we anyway would apply the continuous results to discrete signals and naively measure the scale-space lifetime of a structure existing between scales t_1 and t_2 by $t_{life} = \log t_2 - \log t_1$ it is clear that the lifetime of fine scale structures would be substantially overestimated compared to the lifetime of coarse scale structures. For example, a structure existing in the original signal (corresponding to the scale parameter equal to zero) would be assigned an infinite lifetime. On the other hand, if we would measure the scale-space lifetime by $t_{life} = t_2 - t_1$ then instead the lifetime of coarse scale structures would be overestimated, since it is well-known that (at least at coarse scales) "things happen approximately logarithmically with scale".

In this presentation we will develop how such a transformed scale parameter valid also for discrete signals can be introduced in a formal manner. The treatment is based on the assumption that the relative decay rate of local extrema should be constant across scales for certain reference data. As reference data we take random noise signals from different distributions. We will demonstrate that with this formulation, the effective scale concept for continuous signals will be equivalent to the logarithmic transformation induced by the scaling argument. We will also show that with increasing scale, the effective scale concept for discrete signals approaches the effective scale concept for continuous signals. As a tool for the analysis we will derive estimates of how the number of local extrema in a signal can be expected to vary with scale for random noise data of different normal distributions. Special attention will be given to the transition phenomena at fine scales due to the finite sampling density of discrete signals.

[2] In a pyramid representation, which is one type of multi-scale representation with a discrete scale parameter, scale-space lifetime can obviously be measured as the number of resolution layers during which a structure exists. (However, the quantization in the set of possible scale values will be quite coarse).

Although it would be of interest to relate this subject to notions such as "significant" indicated above, we will not make any claims about how the unit system suggested here relates to biological perception. The treatment given will be strictly mathematical, aimed at addressing a technical problem with a formal treatment. We have, however, demonstrated with experiments on different types of real imagery that the effective scale concept developed here gives intuitively reasonable results when used in conjunction with the measurements of significance of blob-like structures performed in the scale-space primal sketch, see [9, 10, 13].

2 Transformation of the Scale Parameter: Effective Scale

At first glance the problem of transforming the scale parameter may seem somewhat ad hoc. What properties do we want from an "effective scale parameter"? Intuitively, we would like structures at different scales to be treated in a way as uniform as possible so that neither the lifetime of fine scale structures is overestimated compared to the lifetime of coarse scale structures nor the opposite. How should this property be formalized? The approach we will take here is to assume that the expected remaining lifetime of a local extremum should not vary with scale. More precisely, we will assume that the probability that a *certain* local extremum disappears[3] after a small amount of smoothing $\Delta\tau$, expressed in effective scale, should remain constant over scale, i.e., *the relative decay rate of local extrema should be constant over scales.*

[3] For one-dimensional signals the number of local extrema in a signal is guaranteed to decrease monotonically with scale. In two and higher dimensions the situation is more complicated, since the number of local extrema can in fact increase (locally) with scale-space smoothing due to creations of saddle-extremum pairs [11]. However, the expected number of local extrema, as an average over many signals, can always be expected to decrease.

2.1 Definition and Derivation

Assume that we know how the expected number of extremum points per unit length varies with scale. In other words, assume that we know[4] how

$$p(t) = \{\text{the expected density of extremum points at scale } t\} \quad (2)$$

varies with t. What we want to define is a transformation function h such that the effective scale can be written $\tau = h(t)$. The decay rate requirement can be stated as:

Requirement 1. **(Uniform relative decay rate for local extrema)**

The probability that a certain extremum point (or equivalently a certain blob) disappears after a small increment $d\tau$ in effective scale should be independent of both the effective scale τ and the current number of local extrema in the signal. That is

$$\frac{\frac{dp}{d\tau}}{p} = \frac{d(\log p)}{d\tau} = C_1 = constant \quad (3)$$

Integration of (3) gives:

$$\log p = C_1 \tau + C_2 \quad (4)$$

for some arbitrary C_2. By introducing new arbitrary constants A and B, we can conclude

Proposition 1. **(Effective scale)**

Assume that we know how the expected density of local extrema p behaves as a function of scale t and let τ be the effective scale parameter given by Req. 1. Then, for some arbitrary constants A and $B > 0$, the effective scale as function of the ordinary scale parameter is given by

$$\tau = h(t) = A + B \log p(t) \quad (5)$$

[4] In the discrete case the entity $p(t)$ can also be interpreted as the probability that a certain spatial point x is a local extremum in the smoothed grey-level image at scale t.

The actual values of A and B are, of course, unimportant. Without loss of generality A can be set to zero. Its interpretation is just as an arbitrary offset coordinate and does not affect the scale-space lifetime. Similarly, B just corresponds to an arbitrary but unessential linear rescaling of the effective scale parameter.

So far no assumptions have been made about the dimensionality of the signal or whether it is continuous or discrete. What is left to determine is how the density of extrema can be expected to behave with scale. Both theoretical and experimental results will be given below. However, first we will illustrate some immediate consequences of the stated definition.

2.2 Examples and Experimental Results

For continuous signals it is known that the number of local extrema in a signal decreases approximately as t^α with scale. This relation has been discussed by e.g. Müssigmann [15] and can also be motivated theoretically (see Sec. 3). Hence, we have $p(t) = \text{constant}/t^\alpha$ which means that

$$\tau(t) = A + B \log p(t) = A + B \log \text{constant} - \alpha B \log t \tag{6}$$

and a graph showing the number of local extrema as a function of scale will be a straight line in a log-log-diagram. This indicates that this definition of effective scale given by Requirement 1 is qualitatively similar to a definition of effective scale based on the scaling argument.

For discrete signals the number of extrema will also show the same qualitative behaviour at coarse levels of scale, where the grid effects are negligible. However, at fine levels of scale the $t^{-\alpha}$-behaviour cannot hold, since it is based on the assumption that the original signal contains equal amount of structure over all levels of scale. The discrete signal is limited by its finite sampling density.

These ideas are illustrated in Fig. 1, where we show the logarithm of the number of extrema as a function of the logarithm of the scale parameter. The left diagram shows simulated results for a large number of point noise images generated from three different distributions; normal

Figure 1: Experimental results showing the number of local extrema as function of the scale parameter t in log-log scale (a) measured values (b) accumulated mean values. Note that a straight-line approximation is valid only in the interior part of the scale interval. At the lower end point of the interval we have interference with the inner scale, given by the sampling density of the image, and the higher end point there is interference with the outer scale, given by the size of the image.

distribution, rectangle distribution and exponential distribution. The right curve shows the average of these results. Note that the straight line approximation is valid only in an interior scale interval. At fine scales we have interference with the *inner scale*, given by the sampling density of the image, and at coarse scales there is interference with the *outer scale* [7], given by the size of the image.

The idea behind the notion of effective scale is to take the inner scale into account and guarantee a precise definition of scale-space lifetime also at fine levels of scale. Combined with the notion of a scale-space for discrete signals [8], which takes the discrete nature of implementation into account, it gives us the necessary tool to investigate the fine scale structures.

In this presentation we have chosen not to treat the behaviour of finite images at very coarse levels of scale, since in such situations the treatment of the image boundaries will substantially affect the scale-space behaviour. Instead, we argue that if one really wants to study objects at such a coarse scale that the boundary effects become important, then the problem is to a large extent undefined and one should rather try to acquire additional image data in a region around the current image, so that the scale-space smoothing becomes well-defined up to the prescribed accuracy. This can be easily accomplished in an active vision situation.

3 Density of Local Extrema as Function of Scale

Of course, the question concerning how the density of local extrema can be expected to vary with scale seems to be very difficult or even impossible to answer to generally, since such a quantity can be expected to vary substantially from one image to another. How should we then be able to talk about "expected behaviour"? Should we consider all possible (realistic) signals/images, study how this measure evolves with scale and then form some kind of average?

In this section we will perform a simple one-dimensional study. We will consider random noise data with normal distribution. Under these assumptions it is possible to derive a compact closed form expression for this quantity. We will base the analysis on a treatment by Rice [17] about the expected density of local maxima of stationary normal processes (see also Papoulis [16] or Cramer and Leadbetter [4]).

3.1 Continuous Analysis

The density of local maxima μ for a stationary normal process can be expressed in terms of the second and fourth order derivatives of the autocorrelation function R or equivalently in terms of the second and fourth order moments of the spectral density S (the Fourier transform of R):

$$\mu = \frac{1}{2\pi}\sqrt{-\frac{R^{(4)}(0)}{R''(0)}} = \frac{1}{2\pi}\sqrt{\frac{\int_{-\infty}^{\infty} \omega^4 S(\omega) d\omega}{\int_{-\infty}^{\infty} \omega^2 S(\omega) d\omega}} \tag{7}$$

Since the scale-space representation L is generated from the input signal f by a linear transformation, the spectral density of L, denoted S_L, is given by

$$S_L(\omega) = |H(\omega)|^2 S_f(\omega) \tag{8}$$

where S_f is the spectral density of f and $H(\omega)$ the Fourier transform of the impulse response h

$$H(\omega) = \int_{-\infty}^{\infty} h(t) e^{-i\omega t} dt \tag{9}$$

In our scale-space case, h is of course the Gaussian kernel g with the Fourier transform G

$$g(\xi;\ t) = \frac{1}{\sqrt{2\pi t}} e^{-\xi^2/2t}; \quad G(\omega;\ t) = \frac{1}{2} e^{-\omega^2 t/2} \tag{10}$$

Assuming that f is generated by white noise with $S_f(w) = 1$ this gives

$$S_L(\omega) = \frac{1}{4} e^{-\omega^2 t} \tag{11}$$

Using the formula (see e.g. Spiegel [18] 15.77)

$$\int_0^\infty x^m e^{-ax^2} dx = \frac{\Gamma(\frac{m+1}{2})}{2a^{\frac{m+1}{2}}} \tag{12}$$

we obtain a closed form expression for the density of local maxima of a continuous signal, $p_c(t)$:

$$p_c(t) = \frac{1}{2\pi} \sqrt{\frac{\int_{-\infty}^{\infty} \omega^4 \frac{1}{4} e^{-\omega^2 t} d\omega}{\int_{-\infty}^{\infty} \omega^2 \frac{1}{4} e^{-\omega^2 t} d\omega}} = \frac{1}{2\pi} \sqrt{\frac{2\frac{\Gamma(\frac{5}{2})}{2t^{\frac{5}{2}}}}{2\frac{\Gamma(\frac{3}{2})}{2t^{\frac{3}{2}}}}} = \frac{1}{2\pi} \sqrt{\frac{3}{2}} \frac{1}{\sqrt{t}} \tag{13}$$

Of course, an identical result applies[5] to local minima. To summarize,

Proposition 2. **(Density of local extrema in scale-space (white noise, 1D))**

In the scale-space representation of a one-dimensional continuous signal generated by a white noise stationary normal process, the expected density of local maxima (minima) in a smoothed signal at a certain scale decreases with scale as $t^{-\frac{1}{2}}$.

[5] Observe that the same type of qualitative behaviour ($p_c(t) \sim t^{-\frac{1}{2}}$) applies also to the local extrema in the *spatial derivatives* of the scale-space representation (just replace $H = G$ by $H = (i\omega)^n G$ in the previous analysis).

This scale dependence implies that *a graph showing the density of local maxima (minima) as function of scale can be expected*[6] *to be a straight line in a log-log diagram.*

$$\log(p_c(t)) = \frac{1}{2}\log(\frac{3}{2}) - \log(2\pi) - \frac{1}{2}\log(t) = constant - \frac{1}{2}\log(t) \tag{14}$$

By combining Proposition 2 with Proposition 1 we get

Corollary 3. **(Effective scale for continuous signals (1D))**

For continuous one-dimensional signals the effective scale parameter τ_c as function of the ordinary scale parameter t is (up to an arbitrary affine transformation, i.e., for some arbitrary constants A' and $B' > 0$) given by a logarithmic transformation

$$\tau_c(t) = A' + B'\log(t) \tag{15}$$

An interesting question concerns what will happen if the uncorrelated white noise model for the input signal is changed. A spectral density that has been applied to e.g. fractals (see e.g. Barnsley et.al. [2] or Gårding [6]) is given by $S_f(w) = w^{-\beta}$. For one-dimensional signals, reasonable values of β are obtained between 1 and 3 [2]. Of course, such a distribution is somewhat non-physical, since $S_f(w)$ will tend to infinity as t tends to zero and neither one of the spectral moments is convergent. However, when multiplied by a Gaussian function the second and fourth order moments in (7) will converge provided that $\beta < 3$. We obtain,

$$p_{c,\beta}(t) = \frac{1}{2\pi}\sqrt{\frac{\int_{-\infty}^{\infty}\omega^4\frac{1}{4}e^{-\omega^2 t}\omega^{-\beta}d\omega}{\int_{-\infty}^{\infty}\omega^2\frac{1}{4}e^{-\omega^2 t}\omega^{-\beta}d\omega}} = \ldots = \frac{1}{2\pi}\sqrt{\frac{3-\beta}{2}}\frac{1}{\sqrt{t}} \quad (\beta < 3) \tag{16}$$

Proposition 4. **(Density of local extrema in scale-space (fractal noise, 1D))**

In the scale-space representation of a one-dimensional continuous signal generated by a stationary

[6] Of course, we cannot expect that a graph showing this curve for a particular signal to be a straight line, since this would require some type of ergodicity assumption that in general will not be satisfied. However, the average behaviour over many different types of imagery can be expected to be close to this situation.

normal process with spectral density $\omega^{-\beta}$*, the expected density of local maxima (minima) in a smoothed signal at a certain scale decreases with scale as* $t^{-\frac{1}{2}}$*.*

Note that also this graph will be a straight line in a log-log diagram.

3.2 Discrete Analysis

From the previous continuous analysis we have that the density of local extrema may tend to infinity as the scale parameter tends to zero. As earlier indicated, this result is not applicable to discrete signals, since in this case the density of local extrema will have an upper bound because of the finite sampling. Hence, in order to to capture what happens in the discrete case, a genuinely discrete treatment is necessary. We will base the analysis on the discrete scale-space concept developed in [8]. Given a discrete signal $f : Z \to R$ the scale-space representation $L : Z \times R_+ \to R$ is defined by

$$L(x;\ t) = \sum_{n=-\infty}^{\infty} T(n;\ t) f(x-n) \tag{17}$$

where $T(n;\ t) = e^{-t} I_n(t)$ is the discrete analogue of the Gaussian kernel and I_n are the modified Bessel functions of integer order [1]. Equivalently, this scale-space family can be defined in terms of a semi-discretized version of the diffusion equation [8].

Consider the scale-space representation of a signal generated by a random noise signal. The probability that a point at a certain scale is say a local maximum point is equal to the probability that its value is greater than (or possibly equal to)[7] the values of its nearest neighbours:

$$P(x_i \text{ is a local maximum at scale } t) =$$
$$P((L(x_i;\ t) \geq L(x_{i-1};\ t)) \wedge (L(x_i;\ t) \geq L(x_{i+1};\ t))) \tag{18}$$

[7] Although there are several possible ways to define a local extremum of a discrete signal using different combinations of ">" and "≥", these definitions will yield the same result with respect to this application.

If we assume that the input signal f is generated by a stationary normal process then also L will be a stationary normal process and the distribution of any triple $(L_{i-1}, L_i, L_{i+1})^T$, from now on denoted by $\xi = (\xi_1, \xi_2, \xi_3)^T$, will be jointly normal, which means that its statistics will be completely determined by the mean vector and the autocovariance matrix. Trivially, we have that the mean of ξ is zero provided that the mean of f is zero. Since the transformation from f to L is linear, the autocovariance C_L for the smoothed signal L will be given by

$$C_L(\cdot;\ t) = T(\cdot;\ t) * T(\cdot;\ t) * C_f(\cdot) = T(\cdot;\ 2t) * C_f(\cdot) \tag{19}$$

where C_f denotes the autocovariance of f. In the last equality we have made use of the semigroup property $T(\cdot;\ s) * T(\cdot;\ t) = T(\cdot;\ s+t)$ for the family of convolution kernels. If the input signal consists of white noise then C_f will be the discrete delta function and $C_L(\cdot;\ t) = T(\cdot;\ 2t)$. Taking the symmetry property $T(-n;\ t) = T(n;\ t)$ into account, the distribution of ξ will be jointly normal with mean vector m_{3D} and covariance matrix C_{3D} given by:

$$m_{3D} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}; \qquad C_{3D} = \begin{pmatrix} T(0;\ 2t) & T(1;\ 2t) & T(2;\ 2t) \\ T(1;\ 2t) & T(0;\ 2t) & T(1;\ 2t) \\ T(2;\ 2t) & T(1;\ 2t) & T(0;\ 2t) \end{pmatrix} \tag{20}$$

By introducing new variables $\eta_1 = \xi_2 - \xi_1$ and $\eta_2 = \xi_2 - \xi_3$ we have that $\eta = (\eta_1, \eta_2)^T$ will be jointly normal and its statistics completely determined by

$$m_{2D} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}; \qquad C_{2D} = \begin{pmatrix} a_0(t) & a_1(t) \\ a_1(t) & a_0(t) \end{pmatrix} \tag{21}$$

From well-known rules for the covariance $C(\cdot, \cdot)$ of a linear combination it follows that

$$a_0(t) = C(\eta_1, \eta_1) = C(\eta_2, \eta_2) = 2(T(0;\ 2t) - T(1;\ 2t)) \tag{22}$$

$$a_1(t) = C(\eta_1, \eta_2) = C(\eta_2, \eta_1) = T(0;\ 2t) - 2T(1;\ 2t) + T(2;\ 2t) \tag{23}$$

From $a_0(t) - a_1(t) = T(0;\ t) - T(2;\ t)$ and the unimodality property of T ($T(i;\ t) \geq T(j;\ t)$ if $|i| > |j|$) it follows that $a_0(t) > a_1(t)$ and trivially $a_0(t) > 0$ for all t. Now $p_d(t)$ can be expressed in terms of a two-dimensional integral

$$p_d(t) = \int\int_{\{\eta=(\eta_1,\eta_2):(\eta_1\geq 0)\wedge(\eta_2\geq 0)\}} \frac{1}{\sqrt{(2\pi)^2|C_{2D}|}} e^{-\frac{1}{2}\eta^T C_{2D}^{-1}\eta} d\eta_1 d\eta_2 \tag{24}$$

After some calculations (see [9] App. A.5.4) it follows that

$$p_d(t) = \frac{1}{4} + \frac{1}{2\pi}\arctan\left(\frac{a_1(t)}{\sqrt{a_0^2(t) - a_1^2(t)}}\right) \tag{25}$$

Observe that for any $a_0(t)$ and $a_1(t)$ this value is guaranteed to never be outside the interval $[0, \frac{1}{2}]$. With our expressions for $a_0(t)$ and $a_1(t)$, given by smoothing with the discrete analogue of the Gaussian kernel, the maximum value over variations in t is obtained for $t = 0$:

$$p_d(0) = \frac{1}{3} \tag{26}$$

Proposition 5. **(Density of local extrema in discrete scale-space (1D))**
In the scale-space representation (17) of a one-dimensional discrete signal generated by a white noise stationary normal process, the expected density of local maxima (minima) in a smoothed signal at a certain scale t is given by (25) with $a_0(t)$ and $a_1(t)$ according to (22) and (23).

It is interesting to compare the discrete expression (25) with the earlier continuous result (13). The scale value where the continuous estimate gives a density equal to the discrete density at $t = 0$ is given by the equation $p_c(t) = p_d(0)$, that is by

$$\frac{1}{2\pi}\sqrt{\frac{3}{2}}\frac{1}{\sqrt{t}} = \frac{1}{3} \tag{27}$$

which has the solution

$$t_{c-d} = \frac{27}{8\pi^2} \approx 0.3420 \tag{28}$$

This corresponds to a σ-value of about 0.5848. Below this scale value the continuous analysis is, from that point of view, definitely not a valid approximation of what will happen to discrete signals. By combining Proposition 5 with Proposition 1 we get

Corollary 6. **(Effective scale for discrete signals (1D))**

For discrete one-dimensional signals the effective scale parameter τ_d as function of the ordinary scale parameter t is given by

$$\tau_d(t) = A'' + B'' \log\left(\frac{4\pi}{3\pi + 6arctan\left(\frac{a_1(t)}{\sqrt{a_0^2(t) - a_1^2(t)}}\right)}\right) \tag{29}$$

for some arbitrary constants A'' and $B'' > 0$ with $a_0(t)$ and $a_1(t)$ are given by (22) and (23).

When defining the effective scale τ_d for discrete signals it is natural to let $t = 0$ correspond to $\tau_d = 0$. In that case A'' will be zero. WLOG we will from now on set $A'' = 0$ and $B = 1$.

3.3 Asymptotic Behaviour at Fine and Coarse Scales

A second order MacLaurin expansion of $p_d(t)$ (see [9] App. A.5.5) yields

$$p_d(t) = \frac{1}{3} - \frac{1}{2\sqrt{3}\pi}t + \frac{1}{6\sqrt{3}\pi}t^2 + O(t^3) \tag{30}$$

This means that the effective scale $\tau_d(t)$ can be MacLaurin expanded (see [9] App. A.5.5)

$$\tau_d(t) = \log\left(\frac{p_d(0)}{p_d(t)}\right) = \frac{\sqrt{3}}{2\pi}t + \left(\frac{1}{2\sqrt{3}\pi} + \frac{3}{8\pi^2}\right)t^2 + O(t^3) \tag{31}$$

Corollary 7. **(Effective scale at fine scales (1D))**

At fine scales the effective scale τ for one-dimensional discrete signals is approximately an affine *function of the ordinary scale parameter t.*

A Taylor expansion of $p_d(t)$ at coarse scales (see [9] App. A.5.6) gives

$$p_d(t) = \frac{1}{2\pi}\sqrt{\frac{3}{2}}\frac{1}{\sqrt{t}}\left(1 + \frac{1}{8t} + O(\frac{1}{t^2})\right) \tag{32}$$

which asymptotically agrees with the continuous result in (13). By inserting this expression into the expression for effective scale and using $p_d(0) = \frac{1}{3}$ we get

$$\tau_d(t) = \log\left(\frac{p_d(0)}{p_d(t)}\right) = \log\left(\frac{2\pi}{3}\sqrt{\frac{2}{3}}\right) + \frac{1}{2}\log(t) + \log\left(1 - \frac{1}{8t} + O(\frac{1}{t^2})\right) \tag{33}$$

Corollary 8. **(Effective scale at coarse scales (1D))**

At coarse scales the effective scale τ for one-dimensional discrete signals is approximately (up to an arbitrary affine transformation) a logarithmic function of the ordinary scale parameter t.

The term $\log(1 - \frac{1}{8t} + O(\frac{1}{t^2}))$ expresses how much the effective scale derived for discrete signals differs from the effective scale derived for continuous signals, provided that the same values of the (arbitrary) constants A and B are selected in both cases.

3.4 Comparisons Between the Continuous and Discrete Results

To illustrate the difference between the density of local maxima in the scale-space representation of a continuous and a discrete signal we show the graphs of p_c and p_d in Fig. 2 (linear scale) and Fig. 3 (log-log scale). As expected, the curves differ significantly for small t and approach each other as t increases.

Numerical values quantifying this difference for a few values of t are given in Table 1. We have tabulated the ratio

$$\tau_{diff}(t) = \frac{\tau_d(t) - \tau_c(t)}{\tau_c(2t) - \tau_c(t)} = \frac{\tau_d(t) - \tau_c(t)}{\frac{\log(2)}{2}} \tag{34}$$

Figure 2: The density of local maxima of a discrete signal as function of the ordinary scale parameter t in linear scale. (a) Graph for $t \in [0, 100]$. (b) Enlargement of the interval $t \in [0, 10]$. For comparison the graphs showing the density of local extrema for a continuous signal $p_c(t)$ and the second order Taylor expansion of $p_d(t)$ around $t = 0$ have also been drawn. As expected, the continuous and discrete results differ significantly for small values of t but approach each other as t increases. The MacLaurin expansion is a valid approximation only in a very short interval around $t = 0$.

Figure 3: The density of local maxima of a continuous and a discrete signal as function of the ordinary scale parameter t in log-log scale ($t \in [0, 100]$). The straight line shows $p_c(t)$ and the other curve $p_d(t)$. We observe that p_c and p_d approach each other as the scale parameter increases. When t tends to zero, $p_c(t)$ tends to infinity while $p_d(t)$ tends to a constant ($\frac{1}{3}$).

which is a natural measure for how much the effective scale obtained from a continuous analysis differs from a discretely determined effective scale. The quantity is normalized so that one unit in τ_{diff} corresponds to the increase in τ_c induced by an increase in t with a factor of two.

4 Summary and Discussion

We have developed how a concept called effective scale, can be defined in a formal way for both continuous and discrete signals. The treatment is based on the assumption that local extrema at different scales should be treated similarly over scales in the sense that the probability that

t	$\tau_{diff}(t)$
0	∞
0.0625	250.30 %
0.25	67.46 %
1.0	-41.82 %
4.0	-10.47 %
16.0	-2.32 %
64.0	-0.56 %
256.0	-0.14 %
∞	0

Table 1: Indications about how the effective scale obtained from a discrete analysis differs from the effective scale given by the continuous scale-space theory. The quantity $\tau_{diff}(t)$ expresses the difference between $\tau_d(t)$ and $\tau_c(t)$ normalized such that one unit (100 %) in $\tau_{diff}(t)$ corresponds to the increase in τ_c induced by an increase in t with a factor of two.

a certain local extremum existing at a certain scale should disappear after a small amount of smoothing $\Delta\tau$, expressed in effective scale, should not depend on scale. From this postulate we have in the one-dimensional case derived closed form expressions for the effective scale as function of the ordinary scale parameter, related this effective scale concept to the one obtained from a scaling argument and made comparisons between the continuous and discrete treatments.

The same type of analysis can, in principle, be carried out also for two-dimensional discrete signals. The probability that a specific point at a certain scale is a local maximum point is again equal to the probability that its value is greater than the values of its neighbours. Depending on the connectivity concept (four-connectivity or eight-connectivity on a square grid) we then obtain either a four-dimensional or an eight-dimensional integral to solve. However, because of the dimensionality of the integrals we have not made any attempts to calculate explicit expressions for the variation of the density as function of scale. Instead, for implementational purpose, the behaviour over scale has been simulated for various uncorrelated random noise signals (see Sec. 2.2). From those experiments it has been empirically demonstrated that the $t^{-\alpha}$ dependence (with[8] $\alpha \approx 1.0$) of the density of local extrema as function of scale constitutes a reasonable

[8] The reason why the exponent α changes from 0.5 to 1.0 when going from one to two dimensions can intuitively be understood by a dimensional analysis: Assume (as in Appendix A.2) that the standard deviation of the Gaussian kernel, $\sigma = \sqrt{t}$, can be linearly related to a characteristic length, x, in the scale-space representation of an N-dimensional signal at scale t. Moreover, assume that a characteristic distance d between the local extrema in that signal is linearly related to x. Then, the density of local extrema will be proportional to $d^{-N} \sim x^{-N} \sim \sigma^{-N}$, that is to $t^{-N/2}$.

approximation at coarse levels of scale.

5 Acknowledgments

I would like to thank Jan-Olof Eklundh for continuous support and encouragement, Lars Holst for useful discussions about the statistical analysis and Jonas Gårding for comments on fractals.

The support from the Swedish National Board for Industrial and Technical Development, NUTEK, is gratefully acknowledged.

References

[1] Abramowitz M., Stegun I.A. (1964) *Handbook of Mathematical Functions*, Appl. Math. Ser., **55**, National Bureau of Standards.

[2] Barnsley M.F., Devaney R.L., Mandelbrot B.B., Peitgen H-O., Saupe D., Voss R.F. (1988) *The Science of Fractals*, Springer-Verlag, New York, USA.

[3] Burt P.J., Adelson E.H. (1983) "The Laplacian Pyramid as a Compact Image Code", *IEEE Trans. Comm.*, Vol. 9, No. 4, pp532-540.

[4] Cramer H., Leadbetter M.R. (1967) *Stationary and Related Stochastic Processes*, John Wiley and Sons, New York.

[5] Crowley J.L., Stern R.M. (1984) "Fast Computation of the Difference of Low Pass Transform", *IEEE Trans. Patt. Anal. Machine Intell.*, Vol. 6, pp212-222.

[6] Gårding J. (1988) "Properties of Fractal Intensity Surfaces", *Patt. Recogn. Lett.*, Vol. 8, pp319-324.

[7] Koenderink J.J. (1984) "The Structure of Images", *Biol. Cyb.*, Vol. 50, pp363-370.

[8] Lindeberg T.P. (1990) "Scale-Space for Discrete Signals", *IEEE Trans. Patt. Anal. Machine Intell.*, Vol. 12, No. 3, pp234-254.

[9] Lindeberg T.P. (1991) *Discrete Scale-Space Theory and the Scale-Space Primal Sketch*, Ph.D. Thesis, ISRN KTH/NA/P–91/8–SE, Royal Institute of Technology, S-100 44 Stockholm, Sweden.

[10] Lindeberg T.P., Eklundh J.O. (1991) "On the Computation of a Scale-Space Primal Sketch", *J. Visual Comm. Image Repr.*, Vol. 2, No. 1, pp55-78, 1991.

[11] Lindeberg T.P. (1992) "Scale-Space Behaviour of Local Extrema and Blobs", *J. Math. Imaging and Vision*, to appear.

[12] Lindeberg T.P., Eklundh J.-O. (1992) "The Scale-Space Primal Sketch: Construction and Experiments" , *Image and Vision Comp.*, to appear.

[13] Lindeberg T.P. (1992) "Detecting Salient Blob-Like Image Structures with a Scale-Space Primal Sketch — A Method for Focus-of-Attention", *Int. J. Comp. Vision*, to appear.

[14] Lindeberg T.P. (1991) "Guiding Early Visual Processing with Qualitative Scale and Region Information", *Submitted.*

[15] Müssigmann U. (1989) "Texture Analysis, Fractals and Scale-Space Filtering", *Proc. 6th Scand. Conf. Image Anal.*, Oulu, Finland, Jun 19-22, pp987-993.

[16] Papoulis A. (1972) *Probability, Random Variables and Stochastic Processes*, McGraw-Hill.

[17] Rice S.O. (1945) "Mathematical Analysis of Random Noise", *Bell Syst. Tech. J.*, Vol. XXIV, No. 1, pp46-156.

[18] Spiegel M.R. (1968) *Mathematical Handbook of Formulas and Tables*, Schaum's Outline Series in Mathematics, McGraw-Hill.

[19] Witkin A.P. (1983) "Scale-Space Filtering", *Proc. 8th Int. Joint Conf. on Art. Intell.*, Karlsruhe, Germany, Aug 8-12, pp1019-1022.

[20] Yuille A., Poggio T. (1986) "Scaling Theorems for Zero-Crossings", *IEEE Trans. Patt. Anal. Machine Intell.*, Vol. 9, No. 1, pp15-25.

A Appendix

A.1 On Scale-Space Lifetime of Structures and Significance Measurements

To exemplify what we mean by the scale-space lifetime of a structure, let us mention that one type of structures that we have considered in earlier work [9, 10, 13] are objects called *scale-space blobs*, which in general only exist over certain intervals in scale. More precisely, a scale-space blob consists of a set of objects called *grey-level blobs* linked across scales. Every such grey-level blob exists at a single level of scale and is associated with one local extremum and one saddle point in the smoothed grey-level image at that level of scale. The linking across scales proceeds as long as neither the extremum nor the saddle point is involved in any bifurcations [11]. In this way, the extent of the scale-space blob will be delimited by two scale values (at which bifurcations take place), from which the scale-space lifetime can be defined.

More generally, we could also imagine measuring the lifetime any type of structure that can be defined from a signal at any level of scale and be linked across scales in a well-defined manner. Witkin [19] considered the trajectories of the zero-crossings of the Laplacian and called them "fingerprints". Another example could be edges for which the connectivity remains the same across scales.

Concerning the relation between scale-space lifetime and significance, let us remark that in [9, 10, 13] we have been using the (4D) *volume* of these scale-space blobs as a significance measure for extracting blob-like structures from image data. Obviously, this ranking depends on the actual parametrization of the four coordinates (one of those dimensions is scale; the other ones are the (2D) image space and grey-level coordinates)[9]. Therefore, one may speculate whether there exists any "natural coordinate system" for measuring the scale-space blob volume, so that significance values of blobs at different scales can be readily compared. In our previous work we have been using the effective scale concept developed here for transforming the scale parameter and combined this with a statistical treatment of the other three coordinates. Of course, it is very hard if not impossible to give a rigorous theoretical justification for this particular way of computing the significance measure. Ultimately, it is based on a number of assumptions for which there are no proof. However, by experiments [10, 12] we have demonstrated that the approach gives intuitively reasonable results when applied to different types of real images and also that it generates output results useful for further processing [14].

A.2 Scaling Argument in the Continuous Case

The scaling argument showing that, in the continuous case the effective scale parameter, τ, as function of the ordinary scale parameter, t, is given by a logarithmic transformation, can be carried out as follows:

Consider a structure existing at a certain scale and assume that the structure can be associated with a characteristic length[10] x. If a similar structure existing at a different level of scale is to be treated in a similar manner, then the *relative* change in characteristic length, Δx, of that structure caused a by some amount of smoothing, $\Delta\tau$, (expressed in effective scale) should be independent of both the size of that

[9] The reason why we have included also the spatial and grey-level coordinates in the significance measure is because we have noted that small blobs with weak contrast can survive for a substantial amount of time in scale-space if they are located in regions with slowly varying intensity.

[10] Similar to a coarse characteristic length descriptor as used in dimensional analysis in physics.

structure and the current level of scale. In other words, the following relation must hold:

$$\frac{\Delta x}{x} = C_1 \Delta \tau \tag{35}$$

for some arbitrary (non-zero) constant C_1. Assuming that the standard deviation of the Gaussian kernel, $\sigma = \sqrt{t}$, can be linearly related to a characteristic length in a grey-level image at that scale we can write:

$$\frac{\Delta \sigma}{\sigma} = C_1 \Delta \tau \tag{36}$$

By taking the limit of this expression as $\Delta\sigma$ and Δx simultaneously tend to zero and then integrating we obtain

$$\tau = C_2 + \frac{1}{C_1} \log \sigma = C_2 + \frac{1}{2C_1} \log t \tag{37}$$

for some arbitrary integration constant C_2. This shows that for continuous signals the natural scale parameter is essentially the logarithm of the ordinary scale parameter.