

# Stochastic Analysis of Scale-space Smoothing

Kalle Åström, Anders Heyden  
Dept of Mathematics, Lund University  
Box 118, S-221 00 Lund, Sweden  
email: kalle@maths.lth.se heyden@maths.lth.se

## Abstract

*In the high-level operations of computer vision it is taken for granted that image features have been reliably detected. This paper addresses the problem of feature extraction by scale-space methods. This paper is based on two key ideas: to investigate the stochastic properties of scale-space representations and to investigate the interplay between discrete and continuous images. These investigations are then used to predict the stochastic properties of sub-pixel feature detectors.*

## 1 Introduction

Low level image processing is often used to detect and localise features such as edges and corners. It is also used to correlate or match small parts of one image with parts in another. Methods for doing this have been developed for some time, see [6, 9, 10, 11, 14, 18, 19, 22]. However, the stochastic analysis of these algorithms have often been based upon poorly motivated stochastic models. In particular, the effects of image discretisation, interpolation and scale-space smoothing is often neglected or not analysed in detail.

In this paper, image acquisition, interpolation and scale-space smoothing are modelled into some detail. *Image acquisition* is viewed as a composition of blurring, ideal sampling and added noise, similar to [20]. The discrete signal is analysed after *interpolation*. This makes it possible to detect features on a sub-pixel basis. Averaging or *scale-space smoothing* is used to reduce the effects of noise. To understand feature detection in this framework, one has to analyse the effect of noise on interpolated and smoothed signals. In doing so a theory is obtained that connects the discrete and continuous scale-space theories.

The paper is organised as follows. Section 2 treats the image acquisition model. In Section 3 a method is proposed where the discrete scale-space is induced from the continuous scale-space theory. The stochastic properties of the intensity error field are discussed in Section 4. A short introduction

to stationary random fields is given and some important results that are relevant for our model are demonstrated. The ideas are verified with numerical experiments on real images. The sub-pixel edge detector is studied in Sections 5. Sub-pixel correlation is investigated in Section 6.

## 2 Image acquisition

To model the image acquisition process, the intensity distribution that would be caught by an ideal camera is first affected by aberrations in the optics of the real camera, e.g. blurring caused by spherical aberration, coma and astigmatism. Other aberrations deform the image, like Petzval field curvature and distortion, see [13]. Such distortion can typically be handled by geometric considerations in mid-level vision and will not be commented upon here. One way to model camera blur is to convolve the ideal intensity distribution with a kernel corresponding to the smoothing caused by the camera optics. This process also removes some amount of the high spatial frequencies.

In a video-camera, the blurred image intensity distribution is typically measured by a CCD array. One can think of each pixel intensity as the weighted mean of the intensity distribution in a window around the ideal pixel position. Taking the weighted mean around a position is equivalent to first convolving with the weighting kernel and then ideal sampling. Finally, due to quantisation and other errors, stochastic errors are introduced.

Led by this discussion we will use the following image acquisition model:

$$W_{\text{ideal}} \xrightarrow{\text{blur}} W \xrightarrow{\text{sampling}} w_0 \xrightarrow{\text{noise}} v_0, \quad (1)$$

where *upper case letters*,  $W$ , denote signals with *continuous* parameters, whereas *lower case letters*,  $w$ , denote *discrete* signals. Here, and often in the sequel, we use the word signal synonymously with function, and discrete signal synonymously with sequence or function defined on  $\mathbb{Z}^n$ , for some  $n$ . These three steps of blurring, sampling and noise will now be discussed in a little more detail.

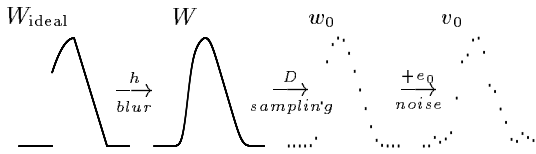


Figure 1: Illustration of the image acquisition model. The original intensity distribution  $W_{\text{ideal}}$  is blurred by the operator  $h$  corresponding to camera optics blur and digitisation blur. The blurred signal  $W$  is then sampled to form a discrete signal  $w_0$ . Finally noise is added. The result is the measured discrete signal  $v_0 = w_0 + e_0$ .

Here *blurring* is modelled as an abstract operator  $h$ , such that  $W = h(W_{\text{ideal}})$ . We assume that no aliasing effects are present, when the function  $W$  is sampled at integer positions. This makes it possible to reconstruct  $W$  from the sampled data.

**Assumption 2.1.** *All energy in the high spatial frequencies is cancelled before sampling. The function  $W$  is band-limited, i.e.  $W \in \mathcal{B}(\mathbb{R}^n)$ , where*

$$\mathcal{B}(\mathbb{R}^n) = \{W \in L_2(\mathbb{R}^n) \mid \text{supp } \mathcal{F}W \subset (-1/2, 1/2)^n\}$$

In the definition of the Fourier transform, we use the formula

$$\mathcal{F}W(f) = \int_{\mathbb{R}^n} W(\tau) e^{-i2\pi f \cdot \tau} d\tau, \quad (2)$$

where  $f \cdot \tau$  denotes scalar product.

The *sampling* is assumed to be ideal. Introduce the *sampling* or *discretisation* operator,  $D : \mathcal{B} \rightarrow l_2$ ,

$$w(i, j) = (DW)(i, j) = W(i, j). \quad (3)$$

Note that the sampling operator maps a continuous signal  $W$  onto a discrete signal  $w$ .

Finally *noise* is assumed to be an additive stationary random field. Experimentally it is verified that the errors in individual pixel intensities often can be modelled as independent random variables with similar distribution.

These assumptions will serve as an initial model. Further improvements can be made by a more detailed camera acquisition model. Nevertheless, these assumptions will help us to model and analyse the next stage, namely estimating the continuous image intensity distribution from the discrete image. Obviously, it is impossible to reconstruct the original intensity distribution  $W_{\text{ideal}}$  without prior knowledge. It is, however, reasonable to try to estimate the blurred and distorted intensity distribution  $W$ , or to estimate an even more blurred version.

### 3 Interpolation and smoothing

Scale-space theory and its application to computer vision is discussed briefly in this section. A more thorough treatment is given in [16]. The idea is to associate to each signal a family of signals smoothed to different degrees. Each such signal captures the behaviour of the signal at one scale. The idea of smoothing is useful to attenuate high-frequency noise without disturbing the low-frequency components of the signal. There is a trade-off in choosing the smoothing parameter. The real strength in using the scale-space approach is the possibility to study the whole scale-space representation. This will, however, not be pursued in this paper. The emphasis will be made to study the stochastic properties of each scale-space representation separately.

In the continuous case, smoothing with the *Gaussian kernel*

$$G_b(x) = \frac{1}{\sqrt{2\pi b^2}} e^{-|x|^2/2b^2} \quad (4)$$

is very natural. In fact, under some consistency conditions (symmetry, semi-group property, non-creation of local extrema), the Gaussian kernel is the only choice that gives a consistent scale-space theory, cf. [5, 15, 16, 23]. The *smoothing* operator  $S_b$  represents convolution with the Gaussian kernel  $G_b$ . A signal  $W$  is represented at scale  $b$  by its smoothed version  $W_b$ :

$$W_b = S_b(W) = G_b * W. \quad (5)$$

The signal  $W_b$  is called the *scale-space representation* of  $W$ , at scale  $b$ . In the sequel subscripts are used to denote different scales. This scale-space representation has several advantages. Local structure decreases as scale increases. No local extrema are created. Another nice feature is that the smoothed function  $W_b$  has continuous derivatives of arbitrary order. A third useful property is that the high frequency components of the noise are attenuated as scale increases. By using multidimensional Gaussians, there is a natural generalisation to functions  $W$  of several variables. However, some of the nice properties are lost when doing so.

Scale-space theory in the discrete time case has been investigated in [16]. It turns out that just by sampling a continuous scale-space kernel, one obtains a discrete scale-space kernel. However, in doing so one does not obtain a scale-space theory with all the nice features of the continuous scale-space theory. There are difficulties with fine scales. In particular it is difficult to define higher order derivatives at fine scale levels. For the same reason it is difficult to define local extremum and zero crossings for fine scales. The semi-group property is lost. These questions are discussed in [16].

## Interpolation and smoothing

The main idea of our approach is to induce the discrete signal, the scale spaces, etc. from the associated interpolated quantities. By an *interpolation* or *restoration* method we mean an operator that maps a discrete signal,  $w$ , to a continuous one,  $W$ . The following types of interpolation operators  $I_F$  will be used:

$$W(s) = (I_F w)(s) = \sum_i F(s-i)w(i) . \quad (6)$$

The ideal interpolation operator  $I = I_{\text{sinc}}$  is of special interest.

We propose to use ideal low-pass interpolation  $I$ , and discretisation  $D$  as mappings between the continuous and discrete signals to solve the restoration and discrete scale-space problems. In other words we relate the discrete and continuous signals through the operations of discretisation and ideal low-pass interpolation. This is illustrated by the diagram:

$$W \begin{array}{c} \xleftarrow{I} \\ \xrightarrow{D} \end{array} w , \quad (7)$$

where  $D$  is the discretisation operator and  $I$  is the ideal interpolation operator.

Note that if the camera induced blur cancels the high frequency components in  $W$  as in Assumption 2.1, the deterministic restoration  $W_0 = I(w_0)$  is equal to  $W$ .

Using these definitions, the discrete and continuous scale-space representations can be defined simultaneously and consistently. We propose the following:

1. If the primary interest is the interpolated continuous signal, then *restore* the scale-space smoothed continuous signal  $W_b$  from the discrete signal  $w_0$  first using ideal interpolation and then continuous scale-space smoothing.
2. If the primary interest is a discrete scale-space representation, then use the induced representation from the continuous scale-space, as defined in (7).

The procedure is illustrated by the diagram:

$$\begin{array}{ccc} W_0 & \xleftarrow{I} & w_0 \\ s_b \downarrow & & \downarrow s_b \\ W_b & \xrightarrow{D} & w_b \end{array} \quad (8)$$

Thus, from the discrete signal  $w_0$ , the *continuous* scale-space smoothed signal  $W_b$  is obtained as  $W_b = S_b(I(w_0))$ . The *discrete* scale-space signal  $w_b = s_b(w_0)$ , is induced from the continuous scale-space signal, i.e.

$$w_b = s_b(w_0) \stackrel{\text{def}}{=} D(S_b(I(w_0))) , \quad (9)$$

where  $s_b$  is introduced as the discrete scale-space smoothing operator. Notice that  $s_b$  is a convolution with a kernel  $g_b$ ,

$$g_b = D(G_b * \text{sinc}) . \quad (10)$$

The differences between this approach and others, like the sampled Gaussian approach, is very small for large scales but significant for small scales. In fact it can be shown that

$$\|\text{sinc} * G_b - G_b\|_2^2 \leq \frac{1}{b\sqrt{\pi}} \Phi(-\pi b\sqrt{2}) , \quad (11)$$

where  $\Phi$  is the normal cumulative distribution function. Notice that the right hand side is small when  $b$  is large. The sampled Gaussian approach is also equivalent to using interpolation with the delta distribution followed by Gaussian smoothing. This is illustrated in Figure 2. The main motivation for using ideal low-pass interpolation is, however, that the approach is well suited for stochastic analysis as will be shown later. Observe that the interpolated signal  $W$  is smooth. Therefore, there is no difficulty in defining higher order derivatives.

This scale-space theory has several theoretical advantages: It works for all scales. The semi-group property,  $s_{\sqrt{a}} s_{\sqrt{b}} = s_{\sqrt{a+b}}$ , holds. The coupling to continuous scale-space theory gives a natural way to interpolate in the discrete space. There are no difficulties in defining derivatives at arbitrary scales. It is possible to calculate derivatives at arbitrary interpolated positions. Operators which commute in the continuous theory automatically commute in the discrete theory. The effect of additive stationary noise can easily be modelled. It makes it possible to compare the real intensity distribution with the interpolated distribution. There is, however, a price to pay. The discrete scale-space smoothing operator  $s_b$  is a convolution with the discrete function

$$g_b = D(\text{sinc} * G_b) ,$$

i.e.  $s_b(w) = g_b * w$ . In practice this scale-space theory is difficult to use for small scale parameters, because of the large tail of the sinc function. However, the function  $\text{sinc} * G_b$  has a very small tail for larger scales. In practice one may use the approximation  $\text{sinc} * G_b \approx G_b$  for large scales, according to (11). This simplifies implementation substantially.

## 4 The random field model

The discrete image  $v_0 = w_0 + e_0$  is analysed directly or through scale-space smoothing, as illustrated by the diagram:

$$\begin{array}{ccc} W_0 + E_0 & \xleftarrow{I} & w_0 + e_0 \\ s_b \downarrow & & \downarrow s_b \\ W_b + E_b & \xrightarrow{D} & w_b + e_b \end{array} \quad (12)$$

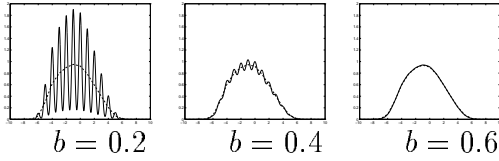


Figure 2: Illustration of interpolation using the sampled Gaussian function. In the figure the signals  $W_b$  (dashed) and  $S_b(I_\delta(w_0))$  (full) is shown for  $b = 0.2, 0.4$  and  $0.6$ . The approximation  $S_b(I_\delta(w_0)) \approx W_b$  is poor, for small scale parameter  $b$ . The approximation is, however, good for large scale parameter.

Note that all operations are linear. The stochastic and deterministic properties can, therefore, be studied separately and the final result is obtained by superposition. Thus with an a priori model on  $W_{\text{ideal}}$ , for example an ideal edge or corner, it is possible to predict the deterministic parts  $W_b$  and  $w_b$ . The stochastic properties of the error fields  $e_0, e_b, E_0$  and  $E_b$ , will now be studied.

## Stationary random fields

The theory of random fields is a simple and powerful way to model noise in signals and images. Stationary or wide sense stationary random fields are particularly easy to use. Denote by  $\mathcal{E}$  the expectation value of a random variable.

**Definition 4.1.** A random field  $X(t)$  with  $t \in \mathbb{R}^n$  is called *stationary* or *wide sense stationary*, if its *mean*  $m(t) = m_X(t) = \mathcal{E}[X(t)]$  is constant and if its *covariance function*  $r_X(t_1, t_2) = \mathcal{E}[(X(t_1) - m(t_1))(X(t_2) - m(t_2))]$  only depends on the the difference  $\tau = t_1 - t_2$ . A random field  $X(t)$  with  $t \in \mathbb{R}^n$  is called *strictly stationary* if for all  $(t_1, \dots, t_n)$  and all  $\tau$  the stochastic variable  $(X(t_1), \dots, X(t_n))$  has the same probability distribution as  $(X(t_1 + \tau), \dots, X(t_n + \tau))$ . ■

For stationary fields we will use  $r_X(s, t)$  and  $r_X(s - t)$  interchangeably as the *covariance function*. The analogous definition is used for a stationary field in discrete parameters. The notion of *spectral density*

$$R_X(f) = (\mathcal{F}r_X)(f) = \int r_X(\tau) e^{-i2\pi f \cdot \tau} d\tau \quad (13)$$

is also important. Again the same definition can be used for random fields with discrete parameters  $s \in \mathbb{Z}^n$ , but whereas the spectral density for random fields with continuous parameters is defined for all frequencies  $f$ , the spectral density of discrete random fields is only defined on an interval  $f \in [-1/2, 1/2]^n$ . Introductions to the theory of random processes and random fields are given in

[1, 7, 8]. In these books you will find that convolution, discretisation and derivation preserves stationarity.

$$\begin{aligned} w = D(W) &\Rightarrow r_w = D(r_W) \\ Y = h * X &\Rightarrow R_Y = R_X | \mathcal{F}h |^2 \\ Y = X' &\Rightarrow r_Y = -r_X'' \end{aligned}$$

We will now show that the ideal interpolation  $I$  preserves stationarity as well. First we will analyse the one-dimensional case. To do this we need a lemma concerning an infinite series:

**Lemma 4.1.**

$$\sum_i \text{sinc}(s - i) \text{sinc}(t - i) = \text{sinc}(s - t) \quad (14)$$

*Proof.* The proof follows from a simple calculations and a formula for summation of a standard series see [21]. Hence,

$$\begin{aligned} &\sum_i \text{sinc}(s - i) \text{sinc}(t - i) = \\ &= \sum_i \frac{\sin(\pi(s - i)) \sin(\pi(t - i))}{\pi^2(s - i)(t - i)} = \\ &= \sum_i \frac{(-1)^{2i} \sin(\pi s) \sin(\pi t)}{\pi^2(s - i)(t - i)} = \\ &= \frac{\sin(\pi s) \sin(\pi t)}{\pi^2} \sum_i \frac{1}{(s - i)(t - i)} = \\ &= \frac{\sin(\pi s) \sin(\pi t)}{\pi^2} \frac{\pi}{s - t} (\cot(\pi t) - \cot(\pi s)) = \\ &= \frac{\sin(\pi s) \cos(\pi t) - \sin(\pi t) \cos(\pi s)}{\pi(s - t)} = \\ &= \frac{\sin(\pi(s - t))}{\pi(s - t)} = \text{sinc}(s - t) \quad \blacksquare \end{aligned} \quad (15)$$

This lemma will now be used in the proof of the following theorem, which describes the stochastic properties of the restored signal at scale zero.

**Theorem 4.1.** If  $e(i)$  is a stationary discrete stochastic process with zero mean and covariance function

$$r_e(i, j) = r_e(i - j) \quad ,$$

such that  $r_e \in l^p$ , for some  $p < \infty$ , then the ideal interpolation at scale zero,

$$E(s) = \sum_i \text{sinc}(s - i) e(i) \quad , \quad (16)$$

is a well defined random process, with convergence in quadratic mean. Moreover,  $E$  is stationary with covariance function

$$r_E(\tau) = I(r_e)(\tau) = \sum_k r_e(k) \text{sinc}(\tau - k) \quad . \quad (17)$$

*Proof.* To prove that  $E(s)$  is well defined we need to prove that  $\sum_i \text{sinc}(s-i)e(i)$  converges in the quadratic mean for every  $s$ . Let

$$S_m = \sum_{|i| < m} \text{sinc}(s-i)e(i) .$$

Convergence in quadratic mean can be established using the Cauchy criterion by showing that

$$\mathcal{E}[|S_m - S_n|^2] \rightarrow 0, \quad \text{as } m, n \rightarrow \infty .$$

Here

$$\begin{aligned} \mathcal{E}[|S_m - S_n|^2] &= \\ &= \sum_{\substack{n < |i| \leq m \\ n < |j| \leq m}} \text{sinc}(s-i)\text{sinc}(s-j)\mathcal{E}[e_i e_j] = \\ &= \sum_{\substack{n < |i| \leq m \\ n < |j| \leq m}} \text{sinc}(s-i)\text{sinc}(s-j)r_e(i-j) . \end{aligned} \quad (18)$$

This tends to zero as  $m, n \rightarrow \infty$  if the double sum

$$\sum_{i,j} \text{sinc}(s-i)\text{sinc}(s-j)r_e(i-j)$$

is absolutely convergent. Making the change of variables,  $k = i - j, l = s - i$ , the double sum can be rewritten as

$$\begin{aligned} \sum_{i,j} |\text{sinc}(s-i)\text{sinc}(s-j)r_e(i-j)| &= \\ &= \sum_{\substack{k=i-j \\ l=s-i}} |\text{sinc}(l)\text{sinc}(k-l)r_e(k)| \leq \\ &\leq \sum_k |r_e(k)| \sum_l |\text{sinc}(l)||\text{sinc}(k-l)| . \end{aligned} \quad (19)$$

Here  $i, j$  and  $k$  are integers whereas  $s$  and therefore also  $l$  are real numbers. The second sum is the discrete convolution of  $|\text{sinc}(s+\cdot)|$  and  $|\text{sinc}(s+\cdot)|$ . Both sequences lie in  $l^p$  for every  $p > 1$ . Since, by Young's inequality, see [12], the convolution of two functions of type  $l^p$  and  $l^q$  is  $l^r$  with  $1/p + 1/q = 1 + 1/r$ , the convolution is of type  $l^r$  for every  $r = p/(2-p)$ , with  $p > 1$ . Hence the sequence  $f$  given by

$$f(k) = \sum_l |\text{sinc}(l)||\text{sinc}(k-l)|$$

belongs to  $l^p$  for every  $p > 1$ . Hölder's inequality then gives that

$$\sum_k |r_e(k)||f(k)| \quad (20)$$

is absolutely convergent if  $r_e \in l^q$  for some  $q < \infty$ .

It now follows that  $m_E(s) = \mathcal{E}[\sum_i \text{sinc}(s-i)e(i)] = \sum_i \text{sinc}(s-i)\mathcal{E}[e(i)] = 0$ . To prove that

$E(s)$  is stationary we need to prove that the covariance  $r_E(s, t)$  only depends on the difference  $s - t$ . The covariance of  $E(s)$  and  $E(t)$  is given by

$$\begin{aligned} r_E(s, t) &= \mathcal{E}[E(s)E(t)] = \\ &= \sum_{i,j} \text{sinc}(s-i)\text{sinc}(t-j)r_e(i-j) = \\ &= \sum_{k=\overset{i}{i-j}} \text{sinc}(s-i)\text{sinc}(t+k-i)r_e(k) = \\ &= \sum_k r_e(k)\text{sinc}(s-t-k) = I(r_e)(s-t), \end{aligned} \quad (21)$$

where we have used Lemma 4.1 to obtain the last but one equality. Thus the continuous random process  $E(s)$  is stationary with covariance function as described. ■

The corresponding theorem in higher dimensions can be proved in exactly the same manner.

Thus, all operations in the commutative diagram (12) preserve stationarity. This simplifies the modelling of errors in scale-space theory. The effects of the operators  $I, D, S_b$  and  $s_b$  on covariance  $r$  and spectral density  $R$  are all known by now.

It is often convenient to assume that the discrete noise  $e_0$  can be modelled as white noise, i.e.

$$r_e(k) = \begin{cases} \epsilon^2, & \text{if } k = 0, \\ 0, & \text{if } k \neq 0. \end{cases}$$

It can then be shown that the covariance function of the interpolated and smoothed error field is

$$r_{E_b} = \epsilon^2 \text{sinc} * G_{b\sqrt{2}} . \quad (22)$$

**Remark.** The restored image intensity distribution  $V_b$  is a sum of a deterministic part  $W_b$  and a stationary random field  $E_b$ . Notice that the restoration and the residual are invariant of the position of the discretisation grid. The effect of discretisation is thus removed. ■

## 5 Edge detection

Stochastic analysis of sub-pixel edge detectors, is one application of the defined theory. To do this we have to model the edge  $W_{\text{ideal}}$  and the blur operator  $h$ . For simplicity, the edge modelled as an ideal step function, i.e. as the Heaviside function with height  $A$  and the smoothing operator  $h$  is modelled as convolution with kernel  $h$ , which is assumed to fulfill Assumption 2.1, and to be approximately a Gaussian of width  $a$ ,

$$h_a \approx G_a . \quad (23)$$

One-dimensional edges are found as local maxima of the derivative of the interpolated and smoothed

signal

$$V_b = G_b * I(e_0 + D(h_a * W_{\text{ideal}})) .$$

According to the above theory  $V_b$  is a sum of a deterministic part  $W_b = G_b * h_a * W_{\text{ideal}}$  and a stochastic part  $E_b$  with covariance function  $R_E = R_e |FG_b|^2$ . Using a model also of  $r_e$ , in this case as white noise with standard deviation  $\epsilon$ , it is thus possible to calculate the distribution of the estimated edge position. This has to be done numerically in most cases, but using reasonable approximations and simplifications the distribution can be approximated as a normal distribution with zero mean and variance, see [4]

$$\sigma^2 \approx \epsilon^2 \frac{3(a^2 + b^2)^3 \sqrt{\pi}}{4A^2 b^5} . \quad (24)$$

These approximation are only valid when  $a$  and  $b$  are fairly large. It is straightforward to calculate the distribution of estimated edge positions numerically and to get results which are valid for a greater range of the parameters. Nevertheless, (24) is a short analytical expression which describes the localisation variance for a large range of the parameters. The variance is inversely proportional to the square of the edge height  $A$ . The variance also decreases with increasing scale parameter  $b$  at low levels and increases at high levels. This is illustrated in Figure 3. In fact the choice of scale parameter  $b$  that minimises the variance is  $b = a\sqrt{5}$ .

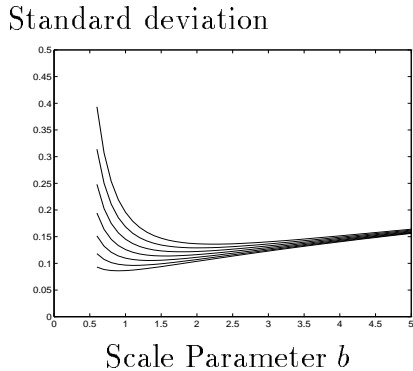


Figure 3: Theoretical standard deviation of edge localisation error versus scale parameter  $b$  at different camera blur parameter  $a = 0.4, 0.5, \dots, 1.0$ . The intensity jump is  $A = 50$  and the standard deviation in each pixel is  $\epsilon = 3$ . The approximations in the calculations are good when  $b$  is larger than 0.8.

The one-dimensional analysis can be generalised to two dimensions in a straightforward manner. As before image acquisition is modelled as convolution with kernel  $h$  followed by discretisation and added noise:  $v_0 = D(h * W_{\text{ideal}}) + e_0$ . The discrete image is then analysed through ideal interpolation

and smoothing:  $V_b = S_b(I(v_0))$ . It is quite popular to define edges as points where the gradient magnitude is maximal in the direction of the gradient, i.e.

$$(\nabla V_b)^T (\nabla^2 V_b) \nabla V_b = 0 , \quad (25)$$

cf. [16]. Several simplifications will be made. We will study the stability of edges with respect to a given search direction  $\tilde{n}$ , i.e. edges are defined as points where

$$(\tilde{n})^T (\nabla^2 V_b) \tilde{n} = 0 . \quad (26)$$

Let the true edge  $\gamma$  be parametrised by curve parameter  $\tau$ . Apply the edge detector in search direction  $\tilde{n}$  from every point  $\gamma(\tau)$ . The detected edge can then be parametrised as  $\tilde{\gamma}(\tau) = \gamma(\tau) + z(\tau)\tilde{n}$ , where  $z$  describes the deviation of the detected edge from the true edge.

It turns out that the errors  $z(\tau)$  along the edge is approximately a stationary process with respect to curve parameter  $\tau$ , with covariance function

$$r_z(\tau) = \epsilon^2 \frac{2(a^2 + b^2)^3}{A^2 \cos^6 \alpha} e^{-\tau^2/(4b^2)} \left( \frac{3}{16b^6} - \frac{3}{16b^8} (\tau \sin(\alpha))^2 + \frac{1}{64b^{10}} (\tau \sin(\alpha))^4 \right) . \quad (27)$$

This could be used to extract the mean value of the random process related to the line. The mean value can be used as the estimated location of the line. If we assume that the search direction differs at most 5 degrees from the perpendicular direction to the edge we get the following estimate of the covariance function

$$r_z(\tau) \approx 1.1 \epsilon^2 \frac{3(a^2 + b^2)^3}{8A^2 b^6} e^{-\tau^2/(4b^2)} . \quad (28)$$

Notice that the parameter  $\tau$  is measured as the arclength along the edge. This has been verified both in simulation, Figure 4 and with real data, Figure 5. In the experiment with real data, five images were taken of the same planar curve. The edge curve was then extracted from each of the five images and aligned. The residuals  $z_i$  in the normal direction with respect to arclength  $\tau$  is shown in Figure 5.

Since the edge is detected as the solution to the equation

$$W_b'' = 0 ,$$

we can regard the edge as a level set to  $W''$ . This makes it possible to use a more refined analysis than the approximation with the tangent line described above. This is discussed in detail in [17].

## 6 Sub-pixel correlation

Analysis of sub-pixel correlation is another application of our scale-space theory. Correlation is usually

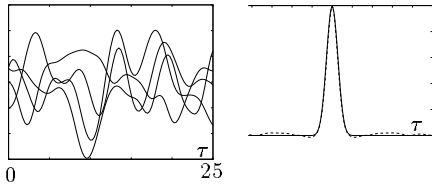


Figure 4: Results from two-dimensional edge detection with simulated data. Left: Edge position errors in a direction roughly perpendicular to the edge at different positions along the edge. The result from four simulations are shown. Distant errors are not correlated but there is a high correlation between edge position errors at spatially close positions. The residuals can be modelled as samples of a random process with respect to the parametrisation of the curve. Right: Theoretical and estimated covariance functions for the residual error process.

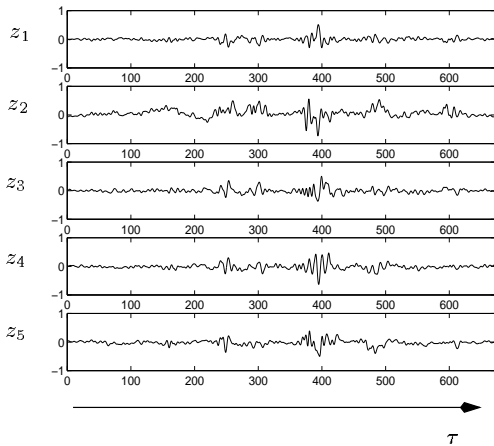


Figure 5: Five edge residuals  $z_i(\tau)$ , empirically estimated after projective alignment of the extracted contours from five different images of the same curve.

done on pixel level, where a regions of one image is translated in whole pixel units and matched to parts of a second image so that the sum of squared differences are minimised. The stochastic errors of pixel correlation is difficult to analyse, mainly because the translation between the regions in the two images usually is of sub-pixel type.

A substantial improvement is obtained by using scale-space restoration of continuous images. This makes it possible to correlate regions in two images with sub-pixel translations with much higher precision than obtained by ordinary methods. Furthermore, a proper modelling of the residual field makes it possible to analyse the stochastic properties of the localisation error. The idea is that, at least locally, the images only differ by an unknown translation  $\rho$ . Denote by  $V = W + E$  and  $\bar{V} = \bar{W} + \bar{E}$  the resto-

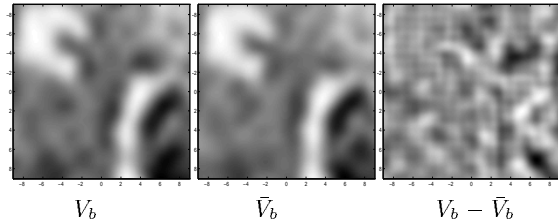


Figure 6: Regions in two images are correlated with sub-pixel translations using least squares of the residuals of the restored continuous scale-space representation at scale  $b = 0.9$ . The accuracy of the sub-pixel translation can be analysed through stochastic models of the residual field. The residual  $V_b - \bar{V}_b$ , shown in a different scale, can also be used to empirically estimate the stochastic properties of the error field  $E_b$ .

red intensity fields in two images for a fixed scale  $b$ . The deterministic functions are identical except for a translation. For a fixed translation  $\rho_0 = (\rho_1, \rho_2)$ , we thus have

$$W(t) = \bar{W}(t + \rho_0), \quad \forall t .$$

To determine the translation  $h$  with sub-pixel accuracy a least squares integral is minimised,

$$F(\rho) = \int_{t \in \Omega} (V(t) - \bar{V}(t + \rho))^2 dt .$$

The result of such a minimisation is shown in Figure 6.

Furthermore, the residual field  $V(t) - \bar{V}(t + \rho)$  can be used to empirically study the stochastic properties of the camera noise  $e_0$ .

The quality of the estimated sub-pixel translation,

$$\hat{\rho} = \operatorname{argmin} F(\rho) ,$$

can be analysed using the statistical model given above. Let  $X = \hat{\rho} - \rho_0$  be the error in estimated translation. By linearising the function  $F$  it can be shown, see [2, 3], that the probability distribution of  $X$  can be approximated with a normal distribution with zero mean and covariance matrix given by

$$C = \mathcal{C}[X] \approx A^{-1} B A^{-T} , \quad (29)$$

$$A = 2 \int_{t_1 \in \Omega} (V \bar{W} \bar{W}^T)(t_1) dt_1 , \quad (30)$$

$$B = \int_{t_1 \in \Omega} ((V \bar{W}) * r_{E - \bar{E}})(t_1) (V \bar{W})(t_1) dt_1 . \quad (31)$$

## 7 Conclusions

In this paper we have modelled the image acquisition process, taking into account both the deterministic and stochastic aspects. In particular the

discretisation process is modeled in detail. This interplay between the continuous signal and its discretisation is very fruitful and the increased knowledge sheds light on scale-space theory, feature detection and stochastic modelling of errors.

The relation between the continuous signal and its discretisation is used to obtain an alternative scale-space theory for discrete signals. It is also used to derive methods of restoring the continuous scale-space representation from the discrete representation. This enables us to calculate derivatives at any position and of any scale.

Furthermore, the stochastic errors in images are modelled and new results are given that show how these errors influence the continuous and discrete scale-space representations and their derivatives. This information is crucial in understanding the stochastic behaviour of scale-space representations as well as fundamental properties of feature detectors. In particular, we have analysed a simple sub-pixel edge detector and a sub-pixel correlator in detail.

From the covariance function it is possible to give confidence interval of the detected position of the edge. We have shown that the location of the edge at different positions along the edge can be regarded as a random process. Furthermore, the covariance function of this random process can be calculated and expressed in terms of the variance of the noise, the widths of the Gaussian kernels and the search angle relative to the true normal of the line.

In order to validate the theory, experiments and simulations both on real and simulated data have been presented. Good agreement with the theoretical model is achieved.

The work can be extended in several directions. Edges were modelled as straight ideal step edges. It would be interesting to study the effect (the bias) caused by other types of edges and the effect of high curvature edges. The model of image acquisition, interpolation and scale space smoothing can also be used to analyse other feature detectors.

## Acknowledgements

The authors thank Georg Lindgren for inspiration and valuable discussions.

## References

- [1] A. Adler. *The Geometry of Random Fields*. Wiley, New York, 1985.
- [2] K. Åström. *Invariancy Methods for Points, Curves and Surfaces in Computational Vision*. PhD thesis, Dept of Mathematics, Lund University, Sweden, 1996.
- [3] K. Åström and A. Heyden. Stochastic analysis of scale-space smoothing. In *Proc. International Conference on Pattern Recognition, Vienna, Austria, 1996*.
- [4] K. Åström and A. Heyden. Stochastic analysis of sub-pixel edge detection. In *Proc. International Conference on Pattern Recognition, Vienna, Austria, 1996*.
- [5] J. Babaud, A. P. Witkin, M. Baudin, and R. O. Duda. Uniqueness of the Gaussian kernel for scale-space filtering. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 8(1):26–33, 1986.
- [6] F. Canny. A computational approach to edge detection. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 8(6):676–698, 1986.
- [7] H. Cramér and M. R. Leadbetter. *Stationary and Related Stochastic Processes*. Wiley, New York, 1967.
- [8] N. A. C. Cressie. *Statistics for Spatial Data*. Wiley, New York, 1991.
- [9] E. De Michelli, B. Caprile, P. Ottonello, and V. Torre. Localization and noise in edge detection. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 10:1106–1117, 1989.
- [10] R. Deriche. Using Canny’s criteria to derive an optimal edge detector recursively implemented. *Int. Journal of Computer Vision*, 1:167–187, 1987.
- [11] O. Faugeras. *Three-Dimensional Computer Vision*. MIT Press, Cambridge, Mass, 1993.
- [12] G. H. Hardy, J. E. Littlewood, and G. Polya. *Inequalities*. Cambridge University Press, third edition, 1959.
- [13] E. Hecht. *Optics*. Addison-Wesley, Reading, Mass., 1987.
- [14] E. Hildreth and D. Marr. Theory of edge detection. *Proceedings of Royal Society of London*, 207:187–217, 1980.
- [15] J. J. Koenderink. The structure of images. *Biological Cybernetics*, 50:363–370, 1984.
- [16] T. Lindeberg. *Scale-Space Theory in Computer Vision*. Kluwer Academic Publishers, 1994.
- [17] G. Lindgren and I. Rychlik. How reliable are contour curves - confidence sets for level contours. *Bernoulli*, 1(2), 1995.
- [18] D. Marr. *Vision*. W. H. Freeman and Company, 1982.
- [19] V. S. Nalwa and T. O. Binford. On detecting edges. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 8:699–714, 1986.
- [20] B. Pratt. *Digital Image Processing*. Wiley-Interscience, 1978.
- [21] A. P. Prudnikov, Y. A. Brychkov, and O. I. Marichev. *Integrals and Series: Volume 1: Elementary Functions*. Gordon and Breach Science Publications, 1986.
- [22] V. Torre and A. Poggio. On edge detection. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 8:147–163, 1986.
- [23] A. P. Witkin. Scale-space filtering. In *Proc. Eighth International Joint Conference on Artificial Intelligence, Karlsruhe, West Germany*, pages 1019–1022, 1983.