

Stochastic Analysis of Image Acquisition, Interpolation and 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. There has been a strong development in scale-space theory and its applications to low-level vision in the last couple of years. Scale-space theory for continuous signals is on a firm theoretical basis. However, discrete scale-space theory is known to be quite tricky, particularly for low levels of scale-space smoothing. The 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.

The modeling of image acquisition, image interpolation and scale-space smoothing is discussed, with particular emphasis on the influence of random errors and the interplay between the discrete and continuous representations. In doing so, new results are given on the stochastic properties of discrete and continuous random fields. A new discrete scale-space theory is also developed. In practice this approach differs little from the traditional approach at coarser scales, but the new formulation is better suited for the stochastic analysis of sub-pixel feature detectors. The interpolated images can then be analyzed independently of the position and spacing of the underlying discretisation grid. This leads to simpler analysis of sub-pixel feature detectors. The analysis is illustrated for edge detection and correlation. The stochastic model is validated both by simulations and by the analysis of real images.

Mathematics Subject Classification: 68U10, 60D05, 60G60

1 Introduction

This work is motivated by an attempt to understand how well edges can be located in images. Study Figure 1a, which contains a digital image taken with an ordinary CCD-camera. The scene contains a sharp discontinuity in

*This work has been done within then ESPRIT Reactive LTR project 21914, CUMULI and the Swedish Research Council for Engineering Sciences (TFR), project 95-64-222

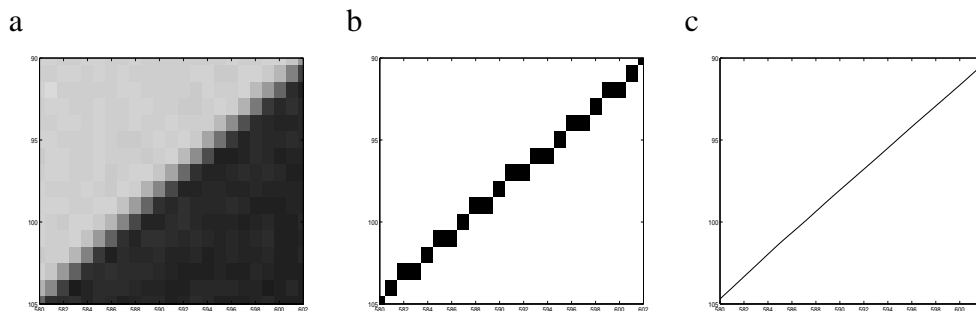


Figure 1: Figure 1a illustrate a part of a digital image containing a high contrast edge. Figure 1b illustrate the result from a traditional edge-detector. Figure 1c illustrate the edge as localized using the new methodology presented in this paper. The result is not only the edge but also an estimate of its localization variance.

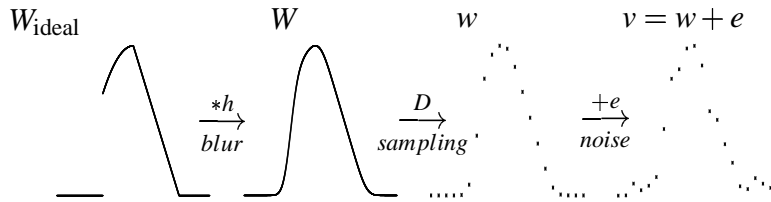


Figure 2: Illustration of the image acquisition model. The original intensity distribution W_{ideal} is blurred by the kernel $h = (h_2 * h_1)$, corresponding to camera optics blur h_1 and digitization blur h_2 . The blurred signal W is then sampled to form a discrete signal w . Finally noise is added. The result is the measured discrete signal $v = w + e$.

contrast called an edge. It is the image of a white paper on a black background. Figure 1a contains a small section of that high contrast edge. We would like to know how well that edge can be located. What does the magnitude of these errors depend on?

A common misconception is that the edge can only be located on a pixel level, that the reason we cannot locate the edge exactly is that we have only got a discrete set of measurements. In fact, most 'edge detector' algorithms [5, 8, 9, 10, 13, 20, 21, 27] as they are called in the image analysis literature give as output a binary image where all the edge pixels are marked without any estimate of edge localization error, see Figure 1b.

In this paper we will try to present a simple model for *image acquisition* and *image interpretation*. This model can be used to understand what is measured in digital images, how to construct simple feature detectors and how to analyze their performance. Thus it will be shown that edges can be detected with sub-pixel accuracy and the edge localization error can be predicted and estimated using image data. Figure 1c illustrates the edge as localized using the new methodology presented in this paper. The edge localization errors are in the order of 1/100'th to 1/10'th of a pixel.

Image acquisition is viewed as a composition of blurring, ideal sampling and added noise, similar to [24]. The discrete signal is analyzed 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 analyze 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 organized as follows. Section 2 treats the image acquisition model. In Section 3 a method for *image interpretation* is proposed. The discrete scale-space is induced from the continuous scale-space theory. This solves the problem of using fine scales and it also simplifies the theory. The main motivation for our formulation of scale-space theory is, however, to simplify the stochastic analysis of low-level feature detectors. 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 and 6. Sub-pixel correlation is investigated in Section 7.

A word should be said about the methodology of this paper. The image acquisition process and the properties of image features are difficult, or impossible, to fully model a priori. Here we make a mathematical engineering approach, starting with some simple assumptions from which the analysis starts. In the results obtained, further approximations and assumptions will be made and motivated. These are thus part of the modeling, and make it possible to give a stochastic analysis of two major problems in computer vision, edge-detection and correlation. The validity of the modeling is tested afterward in simulations and experiments. The results turn out to justify the theory.

2 Image acquisition

To model the image acquisition, the intensity distribution W_{ideal} 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 [12]. 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 convoluting with the weighting kernel

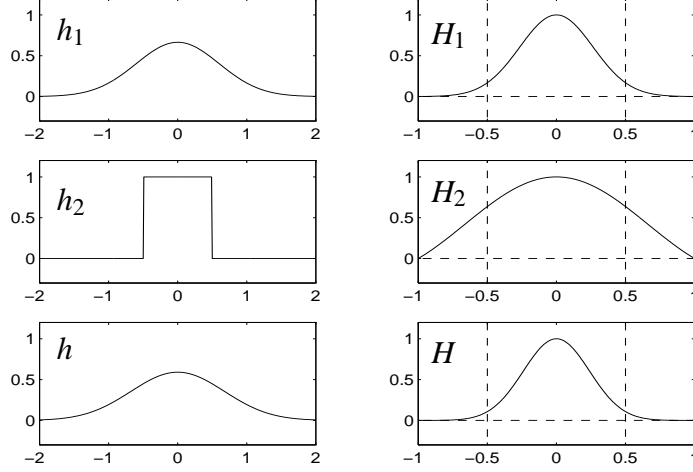


Figure 3: The figure illustrates the kernels h_1 , h_2 and $h = h_1 * h_2$, representing blurring and windowing, and their corresponding Fourier transforms.

and then ideal sampling. This is illustrated in Figure 2. The ideal intensity distribution W_{ideal} is first blurred to form the smooth distribution W . This is sampled to form the discrete signal w . Finally, due to quantization and other errors, stochastic errors are introduced in the measured discrete signal v .

The estimation of the original continuous intensity distribution W_{ideal} using only the discretized and noisy image v is a severely ill-posed inverse problem, that has to be regularised.

The following assumptions and notations will be used.

Assumption 2.1. *The blur caused by camera optics can be modeled as convolution with a kernel h_1 , and the “window blur” caused by weighted sampling, modeled as convolution with kernel h_2 , where both operations are of low-pass type. Denote by h the convolution of these two kernels, $h = h_1 * h_2$, where $*$ denotes convolution. Thus $W = W_{ideal} * h$. ■*

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

$$\mathcal{F} W(f) = \int_{\mathbb{R}^n} W(\tau) e^{-i2\pi f \cdot \tau} d\tau, \quad (1)$$

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

Assumption 2.2. *All energy in the high spatial frequencies is cancelled before discretization. The function $W = W_{ideal} * h$ is band-limited, i.e. the Fourier transform $\mathcal{F} W$ is zero outside a bounded interval. ■*

The Assumption 2.2 says that we have no aliasing effects, when the function W is sampled. In the sequel the function W will be sampled at integer positions. To avoid aliasing effects we will assume that W is band-limited within frequency interval $(-1/2, 1/2)^n$. This makes it possible to reconstruct W from the sampled data, as will be discussed later. For this purpose we introduce the class of functions

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

where L^2 denotes the class of square integrable functions, i.e. functions W such that $\int |W(x)|^2 dx < \infty$.

Assumption 2.3. *The error can be modeled by the addition, after discretization, of a stationary, discrete random field, defined in Section 4. ■*

Experimentally it is verified that the errors in individual pixel intensities often can be modeled as independent random variables with similar distribution.

Aliasing

Some interesting questions are: What is a reasonable model of the blur caused by camera optics and discretization, i.e. what are the functions h_1 and h_2 above? Are high frequencies suppressed before discretization? A crude

model of the camera and discretization blur will be presented, to obtain a feeling for how high frequencies are attenuated before discretization. This is motivated in the one-dimensional case. The kernels in the general case can be obtained by tensor multiplication.

The discussion that follows only serve the purpose of discussing and justifying Assumption 2.2. The exact form of h_1 and h_2 below is not important and will not be used in the sequel.

The camera blur or point-spread function as it is also called is discussed in some detail in [12]. In a well corrected system it can be modeled by the *Airy irradiance distribution*, see [12, p. 485]. Distortions in the camera make it reasonable to model the camera blur as the convolution with a Gaussian kernel with some width a , i.e.

$$h_1(t) = \frac{1}{\sqrt{2\pi}a} e^{-t^2/2a^2} . \quad (3)$$

The corresponding Fourier transform is

$$H_1(f) = e^{-2\pi^2 a^2 f^2} . \quad (4)$$

The measurement or discretization kernel at each pixel is modeled with a rectangular kernel,

$$h_2(t) = \theta(t - 1/2) - \theta(t + 1/2) ,$$

where θ denotes the Heaviside function

$$\theta(x) = \begin{cases} 0, & \text{if } x < 0, \\ 1, & \text{if } x \geq 0. \end{cases} \quad (5)$$

The Fourier transform of h_2 is

$$H_2(f) = \text{sinc}(f) ,$$

where sinc is the function

$$\text{sinc}(s) = \begin{cases} \frac{\sin(\pi s)}{\pi s}, & \text{if } s \neq 0, \\ 1, & \text{if } s = 0. \end{cases}$$

The corresponding interpolation operator is denoted T . Thus the total blur caused by optics and discretization is modeled as a convolution with the kernel

$$h = h_1 * h_2 ,$$

whose Fourier transform is

$$H = H_1 H_2 .$$

The kernels and their transforms are illustrated in Figure 3 where the parameter a in (3) has the value 0.6. The result is clearly a filter of low-pass type. To quantify this we use the squared norms:

$$A_{\text{low}} = \int_{-1/2}^{1/2} |H|^2 df ,$$

$$A_{\text{high}} = \int_{-\infty}^{-1/2} |H|^2 df + \int_{1/2}^{\infty} |H|^2 df .$$

These quantities represents how the filter h attenuates energy at high and low spatial frequencies. For a width of $a = 0.6$ pixels, we have $A_{\text{low}} = 0.42$ and $A_{\text{high}} = 1.2 \cdot 10^{-3}$. Thus the energy in the low frequencies is roughly halved while the energy in the high frequencies is attenuated by a factor of 1000. To avoid aliasing, high frequency components should be attenuated before discretization. Notice that this is done to a reasonable extent by the camera. Aliasing effects are thus substantially reduced by the normal pre-smoothing in the camera system. The effect is still larger when the kernel width a is increased, as is shown in the Table 1.

Thus pre-filtering before sampling is inherent in a camera system. The way it is done depends critically on the properties of each individual camera and the resolution of the digitizer. To get the best effect, the camera blur and the discretization grid should be chosen carefully to match each other. This is analogous to the design of anti-aliasing filters in signal processing, see [4, p. 468]. Notice in particular that too sharp an image will give alias effects which cannot be removed by the subsequent signal processing.

a	A_{low}	A_{high}
0.6	0.42	$1.2 \cdot 10^{-3}$
1.0	0.27	$1.1 \cdot 10^{-6}$
1.5	0.18	$3.0 \cdot 10^{-12}$
2.0	0.14	$7.1 \cdot 10^{-20}$

Table 1: The table illustrates the attenuation of low versus high spatial frequencies for different levels of camera blur.

3 Sampling, interpolation and smoothing

To formalize the image acquisition discussion of the previous section, some notation will be introduced. Let *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 .

Definition 3.1. The **discretisation operator**, or **sampling operator**, $D : \mathcal{B}(\mathbb{R}^2) \rightarrow l^2$, is

$$w(i) = (DW)(i) = W(i) , \quad \forall i \in \mathbb{Z}^n \quad (6)$$

where l^2 denotes square summable sequences, i.e. sequences such that $\sum_{i \in \mathbb{Z}^n} |w(i)|^2 < \infty$. Note that the sampling operator maps a continuous signal W onto a discrete signal w . ■

By assumptions 2.1 and 2.3, image acquisition is modeled as a composition of convolution, ideal discretization and addition of noise:

$$W_{\text{ideal}} \xrightarrow{\text{blur}} W = h_2 * (h_1 * W_{\text{ideal}}) \xrightarrow{\text{sampling}} w = DW \xrightarrow{\text{noise}} v = w + e . \quad (7)$$

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 analyze 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_{ideal} without some a priori knowledge. Compare with blind deconvolution in which W_{ideal} is estimated using the assumption that the convolution kernel has compact support, [17].

It is, however, reasonable to try to estimate the blurred and distorted intensity distribution

$$W_0 = h_2 * (h_1 * W_{\text{ideal}}) = h * W_{\text{ideal}},$$

or to estimate an even more blurred version.

Scale-space smoothing

Scale-space theory and its application to computer vision is discussed briefly in this section. A more thorough treatment is given in [18]. The idea is to associate to each signal $f : \mathbb{R}^n \rightarrow \mathbb{R}$ a family $\{T_t f | t \geq 0\}$ of gradually smoother signals. The original signal corresponds to the scale $t = 0$ and increasing the scale t should simplify the image without creating spurious structure.

Each such signal captures the behaviour of the signal at one scale. This is illustrated in Figure 4. It is natural to view image acquisition as a process that removes fine details. It will therefore be difficult (impossible) to reconstruct the whole scale-space representation of W_{ideal} . It is however possible to reconstruct the signal at courser scales. Scale-space is therefore a natural tool for image interpretation.

Smoothing is also useful in order 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**

$$G_b(x) = \frac{1}{\sqrt{2\pi b^2}} e^{-x^2/2b^2} \quad (8)$$

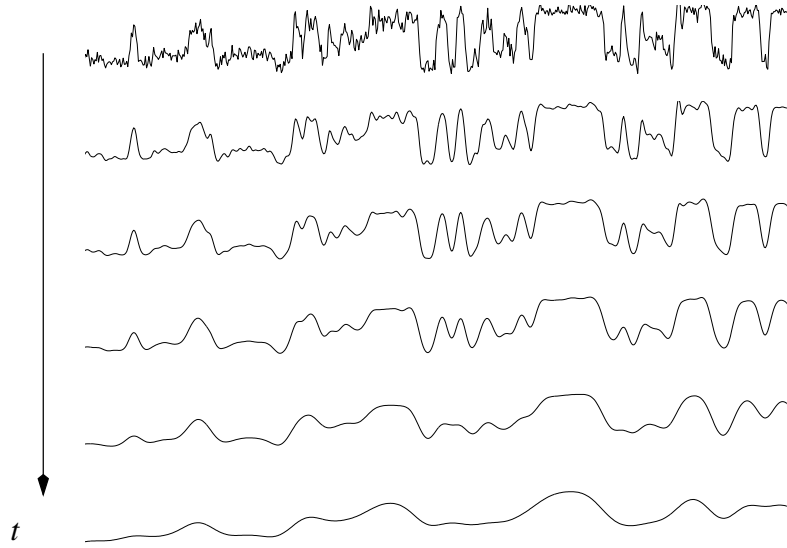


Figure 4: Scale-space representation of a one-dimensional signal. The signal at different scales can be thought to represent the signal at different levels of resolution or detail.

is very natural. In fact if we want to define a scale-space representation $T_t f$ of a one-dimensional signal f with the following properties:

- T_t is a **linear** and **translation invariant** operator for every t ,
- **Scale invariance.** If the signal is enlarged by some factor λ , i.e. $g(x) = f(x/\lambda)$ then there exists a scale $t' = t'(t, \lambda)$ such that $T_t g(x) = T_{t'} f(x/\lambda)$,
- **Generalized Semi-group property:** $T_{t_1}(T_{t_2} f) = T_{t_1+t_2} f$,
- **Positivity preserving:** $f > 0 \implies T_t f > 0$,

then the only choice is to define $T_t f$ according to

$$T_t f = f * G_{\sqrt{t}} ,$$

cf. [14]. Alternate definitions and proofs can be found in [3, 15, 18, 28].

Here we will use an alternative scale parameter $b = \sqrt{t}$.

Definition 3.2. 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 :

$$W_b = S_b W = (G_b * W) . \tag{9}$$

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. It can be shown that structure decrease as scale parameter increase in the sense that local extrema are never enhanced, i.e. $\partial_t(T_t f)(x_0) < 0$ if x_0 is a maximum and $\partial_t(T_t f)(x_0) > 0$ if x_0 is a minimum. 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 generalization to functions W of several variables.

Scale-space theory in the discrete time case has been investigated in [18]. It turns out that just by sampling a continuous scale-space kernel, one obtains a discrete scale-space kernel. Sampling of the Gaussian kernel can thus be used to obtain 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 [18].

Interpolation and scale-space 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 ideal interpolation operator is of special interest:

Definition 3.3. Ideal interpolation $I : l^2 \ni w \rightarrow W \in L^2$ is defined by

$$W(s) = (Iw)(s) = \sum_i \text{sinc}(s-i)w(i) , \quad (10)$$

■

We propose to use ideal interpolation I , and discretization 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 discretization and ideal low-pass interpolation. This is illustrated by the diagram:

$$W \underset{D}{\overset{I}{\rightleftharpoons}} w , \quad (11)$$

where D is the discretization 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 2.2, the deterministic restoration W_0 is equal to W , i.e. $W_0 = W$. This can be stated more precisely using the sampling theorem.

Theorem 3.1. A continuous time signal W with a Fourier transform with support within the interval $(-1/2, 1/2)^n$ is given uniquely by sampled signal $w = D(W)$. The continuous time signal is obtained by ideal interpolation $W = I(w)$.

Proofs and further reading can be found in [16, 22, 26]. Thus ideal interpolation is the pseudo-inverse of discretisation, i.e.

$$D = DID , \quad (12)$$

$$I = IDI . \quad (13)$$

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 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 (11).

The procedure is illustrated by the diagram:

$$\begin{array}{ccc} W_0 & \xleftarrow{I} & w_0 \\ S_b \downarrow & & \downarrow S_b \\ W_b & \xrightarrow{D} & w_b \end{array} \quad (14)$$

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.

$$w_b = s_b(w) \stackrel{\text{def}}{=} D(S_b(I(w_0))) , \quad (15)$$

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

$$g_b = D(G_b * \text{sinc}) . \quad (16)$$

The differences between this approach and others, like the sampled Gaussian approach, is very small for large scales but significant for small scales. This fact is quantified in the following lemma.

Lemma 3.1. *For large scale parameter b , $\text{sinc} * G_b$ is approximately equal to G_b , in the sense that*

$$\|\text{sinc} * G_b - G_b\|_2^2 \leq \frac{1}{b\sqrt{\pi}} \Phi(-\pi b\sqrt{2}) , \quad (17)$$

where Φ denotes the one-dimensional normal cumulative distribution function.

Proof. Consider the Fourier transforms

$$\mathcal{F} G_b(f) = e^{-2\pi^2 b^2 f^2}$$

and

$$\mathcal{F} (\text{sinc} * G_b) = \chi_{[-1/2, 1/2]} \mathcal{F} G_b .$$

These are almost equal because $\mathcal{F} G_b$ is small outside the interval $[-1/2, 1/2]$ when b is large. By Parseval's Theorem we have

$$\|\text{sinc} * G_b - G_b\|_2^2 = \|\chi_{[-1/2, 1/2]} \mathcal{F} G_b - \mathcal{F} G_b\|_2^2 = \int_{|f| > 1/2} (e^{-2\pi^2 b^2 f^2})^2 df = \frac{1}{b\sqrt{\pi}} \Phi(-\pi b\sqrt{2}) \quad (18)$$

■

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:

1. It works for all scales.
2. The semi-group property, $s_{\sqrt{a}} s_{\sqrt{b}} = s_{\sqrt{a+b}}$, holds.
3. The coupling to continuous scale-space theory gives a natural way to interpolate in the discrete space.
4. There are no difficulties in defining derivatives at arbitrary scales.
5. It is possible to calculate derivatives at arbitrary interpolated positions.
6. Operators which commute in the continuous theory automatically commute in the discrete theory.
7. The effect of additive stationary noise can easily be modeled.
8. 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. $s_b(w) = g_b * w$. The discrete scale-space kernels g_b for some values of b are illustrated in Figure 5. 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. This makes it easy to implement. In practice one may use the approximation $\text{sinc} * G_b \approx G_b$ for large scales, according to Lemma 3.1. This simplifies implementation substantially.

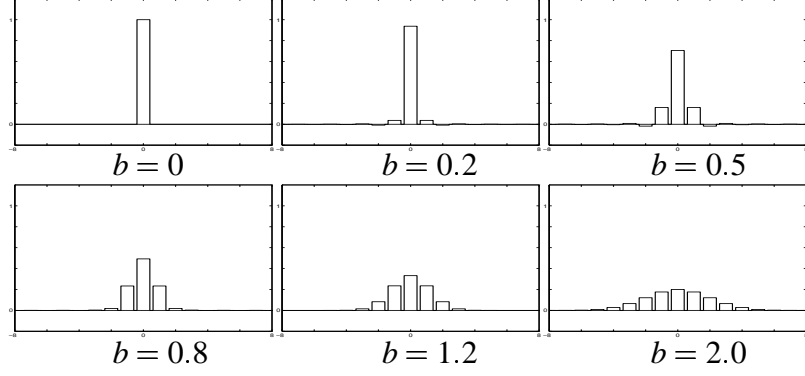


Figure 5: The discrete scale-space kernels $g_b = D(\text{sinc} * G_b)$ for different scales b .

Comparison with other approaches

Discrete scale-space kernels of slightly different types have been used in the literature. In this section, two of these will be compared to our approach.

Definition 3.4. The **sampled Gaussian** discrete scale-space kernel is defined as

$$g_b(j) = \frac{1}{\sqrt{2\pi b^2}} e^{-j^2/2b} .$$

■

Definition 3.5. The **integrated Gaussian** discrete scale-space kernel is defined as

$$g_b(j) = \int_{j-1/2}^{j+1/2} \frac{1}{\sqrt{2\pi b^2}} e^{-x^2/2b} dx .$$

■

It is well known that these approaches are difficult to use for fine scales. The discrete scale-space theories can be understood in the light of interpolation, smoothing and discretisation. We will try to analyse a choice of discrete scale-space smoothing s_b as a composition of some interpolation I_F , Gaussian smoothing S_b and discretisation D . We will use interpolations of type

$$W(s) = (I_F w)(s) = \sum_i F(s-i)w(i) , \quad (19)$$

where I_F is the interpolation operator using distribution F . The question is then what type of interpolation I_F corresponds to sampled Gaussian and interpolated Gaussian, i.e. How should F be chosen so that $s_b = DS_b I_F$:

$$\begin{array}{ccc} W_0 & \xleftarrow{I_F} & w_0 \\ S_b \downarrow & & \downarrow S_b \\ W_b & \xrightarrow{D} & w_b \end{array} \quad (20)$$

It is straight forward to see that the operator s_b is in fact discrete convolution with a kernel $g_b = D(G_b * F)$.

Example 3.1. Discrete scale space using the sampled Gaussian is equivalent to interpolation with δ -distribution, followed by Gaussian smoothing and discretization.

$$s_b = D \circ S_b \circ I_\delta .$$

$$g_b = D(G_b * \delta) .$$

This works well for large scales, but a poor approximation is obtained for small scales as is shown in Figure 6. ■

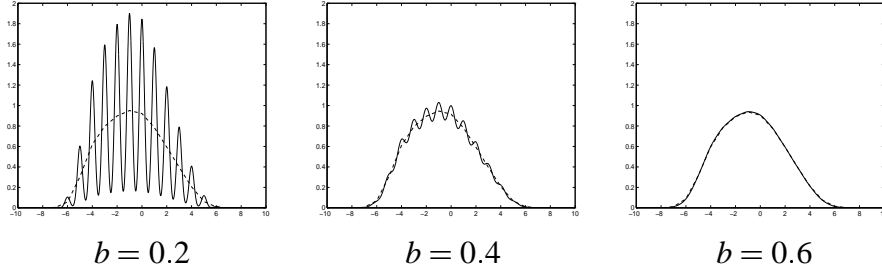


Figure 6: Illustration of interpolation using the sampled Gaussian function. In the figure the signals W_b (dashed) and $S_b(I_\delta(w_0))$ (full) is shown for $b = 0.2, 0.4$ and 0.6 . The approximation $S_b(I_\delta(w_0)) \approx W_b$ is poor, for small scale parameter b . The approximation is, however, good for large scale parameter.

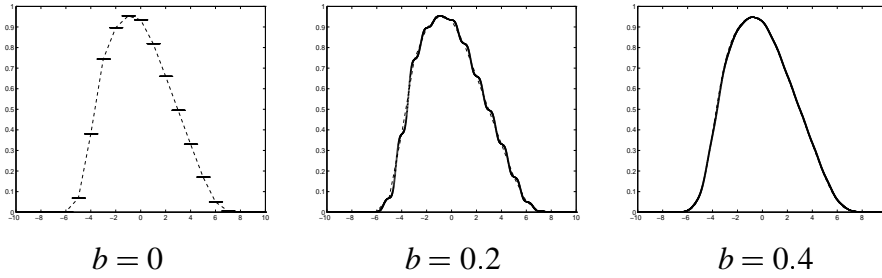


Figure 7: Illustration of interpolation using the integrated Gaussian function. In the figure the signals W_b (dashed) and $S_b(I_{\text{step}}(w_0))$ (full) is shown for $b = 0, 0.2$ and 0.4 . The approximation $S_b(I_{\text{step}}(w_0)) \approx W_b$ is poor, for small scale parameter b . The approximation is, however, good for large scale parameter.

Definition 3.6. *Piecewise constant interpolation* is defined as interpolation with the function.

$$\text{step}(s) = \begin{cases} 0, & |s| > 1/2, \\ 1, & |s| \leq 1/2. \end{cases}$$

■

Example 3.2. Discrete scale space using the integrated Gaussian is equivalent to piece-wise constant interpolation, followed by Gaussian smoothing and discretization, i.e.

$$s_b = D \circ S_b \circ I_{\text{step}} .$$

$$g_b = D(G_b * \text{step}) .$$

Although it gives better results than the sampled Gaussian it performs poorly at fine scales, cf. Figure 7. ■

Example 3.3. The proposed scale-space restoration is made by first doing ideal low-pass interpolation and then scale-space smoothing for continuous signals. This is illustrated in Figure 8. Notice the good approximation already at scale 0. Compare with Figures 6 and 7. ■

Remark. Observe that the reason why restoration in general, and restoration using the sinc function in particular, works, is that the sampled function W has high regularity. All high frequency components have been attenuated by camera and digitization blur. ■

4 The random field model

In this section the stochastic models are investigated. So far we have the following model for **image acquisition**

$$W_{\text{ideal}} \xrightarrow{\text{blur}} W \xrightarrow{D} w_0 \xrightarrow{\text{noise}} v_0 = w_0 + e_0 , \quad (21)$$

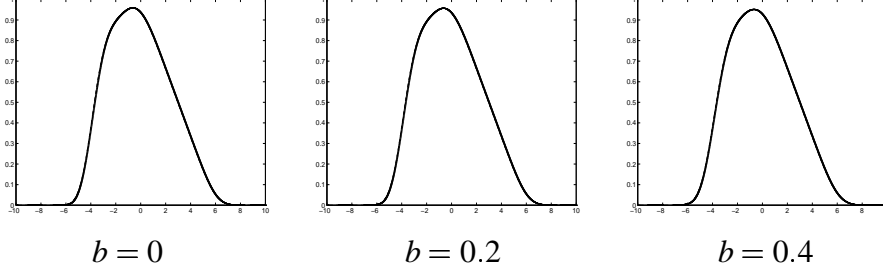


Figure 8: Illustration of ideal low-pass interpolation. The figure illustrates the signals W_b (dashed) and $S_b(I_{\text{sinc}}(w_0))$ (full) for $b = 0, 0.2$ and 0.4 . The approximation $S_b(I_{\text{sinc}}(w_0)) \approx W_b$ is good even for small scale parameter b . (In fact the two curves are indistinguishable in the figure.)

where W_{ideal} is the ideal intensity distribution entering the camera system. This intensity is first blurred. The result W is assumed to have no frequency components outside the interval $(-1/2, 1/2)^n$, see Assumption 2.2. This signal is then discretised and finally perturbed by additive noise.

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

$$\begin{array}{ccc}
 W_0 + E_0 & \xleftarrow{I} & w_0 + e_0 \\
 S_b \downarrow & & \downarrow s_b \\
 W_b + E_b & \xrightarrow{D} & w_b + e_b
 \end{array} \quad (22)$$

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_{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 difference $\tau = t_1 - t_2$. ■

This should be compared with the notion of strict stationarity.

Definition 4.2. 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 τ 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**

$$R_X(f) = (\mathcal{F} r_X)(f) = \int r_X(\tau) e^{-i2\pi f \cdot \tau} d\tau \quad (23)$$

is also important. Again the same definition can be used for random fields with discrete parameters $s \in \mathbb{Z}$, but whereas the spectral density for random fields with continuous parameters is defined for all frequencies f , the spectral density of discrete random fields

$$R_X(f) = \sum r_X(k) e^{-i2\pi f \cdot k} \quad (24)$$

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, 6, 7]. These also contain proofs and comments to the following two useful theorems.

Theorem 4.1 (Convolution of a random field). Let X be a stationary random field and let $Y = h * X$, with $h \in L^2$. Then Y is a stationary random field with covariance

$$r_Y(\tau) = \int_u \int_v h(u)h(v)r_X(\tau + u - v)dudv . \quad (25)$$

If H is the Fourier transform of h , then the spectral density function R_Y is

$$R_Y(f) = R_X(f)|H(f)|^2 . \quad (26)$$

Theorem 4.2 (Sampling of a random field). If $X(t)$ is a stationary random field with continuous parameter $t \in \mathbb{R}^n$ and if x is the result of sampling X at times $t \in \mathbb{Z}^n$, i.e. $x = D(X)$, then x is a stationary discrete random field, with covariance

$$r_x = Dr_X \quad (27)$$

and spectral density

$$R_x(f) = \sum_{k \in \mathbb{Z}^n} R_X(f + k) . \quad (28)$$

Consider the diagram:

$$\begin{array}{ccc} E_0 & \xleftarrow{I} & e_0 \\ S_b \downarrow & & \downarrow s_b \\ E_b & \xrightarrow{D} & e_b \end{array} \quad (29)$$

It follows from the Theorems 4.1 and 4.2 that the operators D , S_b and s_b preserve stationarity. We will now show that the ideal interpolation I preserves stationarity as well. First we will analyze the one-dimensional case. To do this we need a lemma concerning an infinite series:

Lemma 4.1.

$$\sum_i \text{sinc}(s - i) \text{sinc}(t - i) = \text{sinc}(s - t) . \quad (30)$$

Proof. The proof follows from a simple calculations and a formula for summation of a standard series

$$\pi \cot \pi z = \lim_{m \rightarrow \infty} \sum_{n=-m}^m \frac{1}{z - n} ,$$

see [2, p. 188]. Hence,

$$\begin{aligned} & \sum_i \text{sinc}(s - i) \text{sinc}(t - i) = \\ &= \sum_i \frac{\sin(\pi(s - i)) \sin(\pi(t - i))}{\pi^2(s - i)(t - i)} = \\ &= \sum_i \frac{(-1)^{2i} \sin(\pi s) \sin(\pi t)}{\pi^2(s - i)(t - i)} = \\ &= \frac{\sin(\pi s) \sin(\pi t)}{\pi^2} \sum_i \frac{1}{(s - i)(t - i)} = \\ &= \frac{\sin(\pi s) \sin(\pi t)}{\pi^2} \sum_i \left(\frac{1}{(s - t)(t - i)} - \frac{1}{(s - t)(s - i)} \right) = \\ &= \frac{\sin(\pi s) \sin(\pi t)}{\pi^2} \frac{\pi}{s - t} (\cot(\pi t) - \cot(\pi s)) = \\ &= \frac{\sin(\pi s) \cos(\pi t) - \sin(\pi t) \cos(\pi s)}{\pi(s - t)} = \\ &= \frac{\sin(\pi(s - t))}{\pi(s - t)} = \text{sinc}(s - t) . \end{aligned} \quad (31)$$

■

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. In the theorem l^p is used to denote sequences w such that $\sum |w(i)|^p < \infty$.

Theorem 4.3 (Interpolation of a random process). *If $e(i)$ is a stationary discrete stochastic process with zero mean and covariance function*

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

such that

$$r_e \in l^p, \quad \text{for some } p < \infty ,$$

then the ideal interpolation at scale zero,

$$E(s) = \sum_i \text{sinc}(s - i)e(i) , \quad (32)$$

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

$$r_E(\tau) = I(r_e)(\tau) = \sum_k r_e(k) \text{sinc}(\tau - k) . \quad (33)$$

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

$$\mathcal{E}[|S_m - S_n|^2] = \sum_{\substack{n < |i| \leq m \\ n < |j| \leq m}} \text{sinc}(s - i) \text{sinc}(s - j) \mathcal{E}[e_i e_j] = \sum_{\substack{n < |i| \leq m \\ n < |j| \leq m}} \text{sinc}(s - i) \text{sinc}(s - j) r_e(i - j) . \quad (34)$$

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

$$\sum_{i, j} |\text{sinc}(s - i) \text{sinc}(s - j) r_e(i - j)| = \sum_{\substack{k=i-j \\ l=s-i}} |\text{sinc}(l) \text{sinc}(k - l) r_e(k)| \leq \sum_k |r_e(k)| \sum_l |\text{sinc}(l)| |\text{sinc}(k - l)| . \quad (35)$$

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 [11], 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

$$\sum_k |r_e(k)| |f(k)| \quad (36)$$

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

$$\begin{aligned} r_E(s, t) &= \mathcal{E}[E(s)E(t)] = \sum_{i, j} \text{sinc}(s - i) \text{sinc}(t - j) \mathcal{E}[e_i e_j] = \\ &= \sum_{i, j} \text{sinc}(s - i) \text{sinc}(t - j) r_e(i - j) = \sum_{k=i-j} \text{sinc}(s - i) \text{sinc}(t + k - i) r_e(k) = \\ &= \sum_k r_e(k) \sum_i \text{sinc}(s - i) \text{sinc}(t + k - i) = \sum_k r_e(k) \text{sinc}(s - t - k) = \\ &= I(r_e)(s - t), \quad (37) \end{aligned}$$

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. ■

Remark. The covariance function is smooth, i.e. infinitely differentiable. Therefore the random field is \mathcal{C}^∞ in quadratic mean. ■

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

Theorem 4.4 (Interpolation of a random field). *Let $e(i_1, \dots, i_n)$ be a stationary discrete random field with zero mean and covariance function*

$$r_e((i_1, \dots, i_n), (j_1, \dots, j_n)) = r_e(i_1 - j_1, \dots, i_n - j_n) ,$$

such that

$$r_e \in l^p, \quad \text{for some } p < \infty .$$

Then the ideal interpolation (Definition 3.3) of the discrete random field,

$$E = I(e) , \tag{38}$$

is a well defined random field in quadratic mean and E is stationary with covariance function

$$r_E(\tau) = I(r_e)(\tau) . \tag{39}$$

Thus, all operations in the commutative diagram (29) preserve stationarity. This simplifies the modeling 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 modeled as white noise, i.e.

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

Theorem 4.5. *Assume that the discrete image v_0 has been obtained by ideal sampling of the blurred image W with added white noise e_0 , i.e.*

$$v_0 = D(W) + e_0 .$$

Also assume that $\text{supp}(\mathcal{F} W) \in (1/2, 1/2)^n$. Define the restored intensity distribution at scale b according to

$$V_b = (I v_0) * G_b ,$$

where I denotes ideal interpolation and G_b is the Gaussian kernel

$$G_b(s) = \frac{1}{\sqrt{2\pi b^2}} e^{-|s|^2/2b^2} .$$

Then V_b can be written as

$$V_b = (W * G_b) + E_b = W_b + E_b ,$$

where E_b is a stationary random field with covariance function

$$r_{E_b} = \text{sinc} * G_{b\sqrt{2}} . \tag{40}$$

Remark. The theorem states that it is possible to estimate the original continuous distribution W at scale b using our methodology. The error E_b in this estimate is a stationary random field with known covariance function. ■

Proof. It follows from the linearity of interpolation and convolution that

$$V_b = G_b * (I v_0) = G_b * (I w_0) + G_b * (I e_0)$$

It follows from Theorem 3.1 that $(I w_0) = W$. The interpolated error field E_0 has covariance function

$$r_{E_0}(\tau) = \sigma^2 \text{sinc}(\tau)$$

at scale 0 according to Theorem 4.4. Using Theorem 4.1 and the fact that G_b is symmetric, it follows that

$$\begin{aligned}
r_{E_b}(\tau) &= \sigma^2 \int_u \int_v G_b(u) G_b(v) \text{sinc}(\tau + u - v) dv du = \\
&= \sigma^2 \int_u G_b(-u) (G_b * \text{sinc})(\tau + u) du = \\
&= (G_b * (G_b * \text{sinc}))(\tau) = \\
&= ((G_b * G_b) * \text{sinc})(\tau) = (G_{b\sqrt{2}} * \text{sinc})(\tau) ,
\end{aligned} \tag{41}$$

■

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 discretization grid. The effect of discretization is thus removed. ■

One advantage of the idea of restoring a continuous scale-space representation using the discrete image is that it enables us to calculate high order derivatives of the image *at arbitrary positions and at any scale*. Local features can thus be defined in precise mathematical terms, and their position can be calculated with a high degree of numerical precision. The main motivation of our formulation is, however, that it simplifies the analysis of the stochastic properties of the feature location. To do this we must know the stochastic properties of derivatives of wide-sense stationary random fields. This problem is solved in standard texts on random fields such as [1]. We have:

Theorem 4.6 (Partial derivatives of a random field). *Let X be a stationary random field with twice differentiable covariance function r_X . Then X is differentiable in mean squares sense. If $Y = \partial_{t_1} X$ is a partial derivative of X , then Y is also a stationary random field with covariance*

$$r_Y(\tau) = -\partial_{t_1}^2 r_X(\tau) . \tag{42}$$

The spectral density function R_Y is

$$R_Y(f) = R_X(f)(2\pi f_1)^2 . \tag{43}$$

5 One-dimensional edge detection

Stochastic analysis of sub-pixel edge detectors, is one application of this theory. One-dimensional edge detection is treated in this section.

The problem we want to attack is the following:

Problem 5.1. (One-dimensional edge detection).

*Let W_{ideal} be a one-dimensional signal which is smooth except for a finite set of step-discontinuities. Let $v_0 = D(h * W_{\text{ideal}}) + e_0$ be the result after smoothing with kernel h , discretization and added noise. Estimate the position of the discontinuity and the uncertainty in this estimate, using the discrete signal v_0 .*

The analysis below can be extended to any function W_{ideal} . For simplicity, it will be modeled as a ideal step function, i.e. as the Heaviside function, cf. (5).

According to the previous sections, the scale-space analysis is *invariant* to the position of the discretization grid. Thus, without loss of generality, we can assume that the discontinuity is at position 0. The kernel h is assumed to fulfill Assumption 2.2, and to be approximately a Gaussian of width a ,

$$h \approx G_a . \tag{44}$$

Here, we have implicitly assumed that a is large. The noise will be modeled as discrete white noise with standard deviation ϵ . Some of the results that holds exactly using ideal interpolation, can be shown to hold approximately using the sampled Gaussian approach, for larger scales, see Lemma 3.1. The following sub-pixel edge detector will be analyzed.

Definition 5.1 (Edge detection). Using the discrete signal v_0 as defined in Problem 5.1. Define edge positions as points where the magnitude of the derivative

$$\left| \frac{\partial}{\partial x} V_b \right| = \left| \frac{\partial}{\partial x} S_b(I(v_0)) \right| ,$$

attains a local maximum. ■

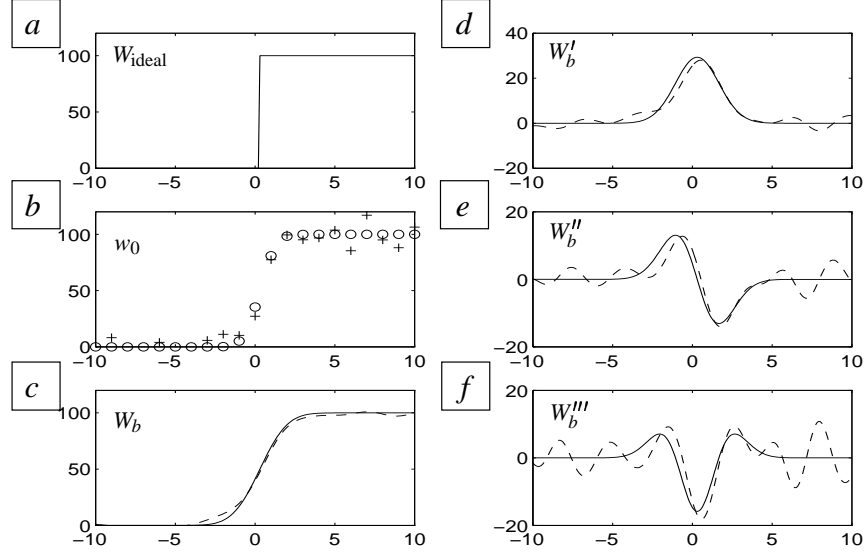


Figure 9: Illustration of one-dimensional sub-pixel edge detection, using scale-space smoothing. a: The ideal step edge. b: The discretized 'image' without noise w_0 (o) and with noise $v_0 = w_0 + e_0$ (+). c-f: The scale space interpolations at scale b and their first three derivatives. The deterministic signal is shown as full lines, whereas the perturbed signal is shown as a broken line. The sub-pixel edge position is defined as the position of the maximum of the first derivative.

As pointed out in the previous section the problem can be decomposed into a deterministic part and a random part. The calculations are then used to predict the stochastic properties of one-dimensional edge detectors. The results are then verified both with simulated and real data.

Deterministic part

First consider the deterministic part of image acquisition. The gray level profile that hits the CCD-array can be modelled as a gray-level function.

$$\mathbb{R} \ni x \mapsto W_{ideal}(x) = A \theta(x) \in \mathbb{R}, \quad (45)$$

where θ denotes the Heaviside function, cf. (5).

According to the image acquisition model the measured deterministic signal is

$$w_0 = D(h * W_{ideal}) \quad (46)$$

Using (44) we get $w_0 = D(h * W_{ideal}) \approx DS_a W_{ideal}$:

$$\mathbb{Z} \ni i \mapsto w_0(i) \approx A \Phi_a(i) \in \mathbb{R}, \quad (47)$$

where Φ_a is the cumulative distribution function of the normal probability density function with variance a^2 .

The discrete signal w_0 is analyzed with interpolation and scale-space smoothing. This gives the continuous signal $W_b = S_b(I(w_0))$. It was assumed that W has no energy outside the frequency interval $(-1/2, 1/2)$. Thus we have $W_b = S_b(W_0) = S_b(h * W_{ideal}) \approx S_b S_a W_{ideal} = S_c W_{ideal}$, where $c = \sqrt{a^2 + b^2}$.

Remark. The approximation $W_b \approx W_{ideal} * G_c$ holds for large b even if we use the *sampled Gaussian* approach. This can be justified using a Riemann sum as an approximation of an integral.

$$\begin{aligned} W_b(x) &= \sum_i w_0(i) G_b(x-i) = \sum_i (W_{ideal} * G_a)(i) G_b(x-i) \approx ((W_{ideal} * G_a) * G_b)(x) = \\ &= (W_{ideal} * (G_a * G_b))(x) = (W_{ideal} * G_c)(x). \end{aligned} \quad (48)$$

■

Again notice that after scale-space interpolation of the discretized image w_0 , we obtain the same result as by smoothing the original intensity W_{ideal} distribution with G_c . *The result is invariant of the position of the discretization grid.* In the following analysis we will need W_b and its first three derivatives:

$$W_b(x) = (W * G_c)(x) = A \Phi_c(x-d) , \quad (49)$$

$$W_b'(x) = A G_c(x-d) , \quad (50)$$

$$W_b''(x) = -A \frac{x-d}{c^2} G_c(x-d) , \quad (51)$$

$$W_b'''(x) = A \left(-\frac{1}{c^2} + \frac{(x-d)^2}{c^4} \right) G_c(x-d) . \quad (52)$$

Note that

$$W_b''(d) = 0 , \quad (53)$$

which implies that W_b has a local extremum at $x = d$. We will need the quantity

$$W_b'''(d) = -\frac{A}{c^3 \sqrt{2\pi}} \quad (54)$$

to determine the stability of this local extrema.

Random part

In this subsection we will derive the effect of additive random noise to the signal. The noise e_0 , is modeled as a white noise Gaussian process with mean 0 and variance ε^2 . That is

$$e_i \in N(0, \varepsilon) , \quad (55)$$

$$C(e_i, e_j) = \begin{cases} \varepsilon^2, & \text{if } i = j , \\ 0, & \text{if } i \neq j . \end{cases} \quad (56)$$

After scale space interpolation we obtain

$$E_b = S_b(I(e_0)) . \quad (57)$$

According to Theorem 4.5 this is a stationary process with with zero mean and covariance function

$$r_{E_b}(\tau) = \varepsilon^2 \text{sinc} * G_{b\sqrt{2}} . \quad (58)$$

According to Lemma 3.1, for coarse scales this can be approximated as

$$r_{E_b}(\tau) = \varepsilon^2 \text{sinc} * G_{b\sqrt{2}} \approx \varepsilon^2 \frac{1}{2b\sqrt{\pi}} e^{-(\tau)^2/(4b^2)} , \quad (59)$$

according to Lemma 3.1.

Remark. If we use the *sampled Gaussian* scale-space approach, the approximation

$$r_{E_b}(\tau) \approx \varepsilon^2 \frac{1}{2b\sqrt{\pi}} e^{-(\tau)^2/(4b^2)}$$

can be justified with explicit calculations.

$$\begin{aligned} C(s, t) &= \mathcal{E} [(X(s) - \mathcal{E}[X(s)]) (X(t) - \mathcal{E}[X(t)])] = \\ &= \mathcal{E} \left[\left(\sum_i \frac{1}{b\sqrt{2\pi}} e^{-(s-i)^2/(2b^2)} e_i \right) \left(\sum_j \frac{1}{b\sqrt{2\pi}} e^{-(t-j)^2/(2b^2)} e_j \right) \right] = \\ &= \sum_{i,j} \frac{1}{b\sqrt{2\pi}} e^{-(s-i)^2/(2b^2)} \frac{1}{\sqrt{2\pi}b} e^{-(t-j)^2/(2b^2)} C(e_i, e_j) = \\ &= \sum_i \frac{1}{b\sqrt{2\pi}} e^{-(s-i)^2/(2b^2)} \frac{1}{b\sqrt{2\pi}} e^{-(t-i)^2/(2b^2)} \varepsilon^2 = \\ &= \varepsilon^2 \frac{1}{2\pi b^2} e^{-(s-t)^2/4b^2} \sum_i e^{-(i-\frac{s+t}{2})^2/b^2} . \end{aligned} \quad (60)$$

As expected, the sampled Gaussian approach gives an error process E_b which is non-stationary. However, if b is large we can approximate the last term

$$\Theta\left(\frac{s+t}{2}\right) = \sum_i e^{-(i-\frac{s+t}{2})^2/b^2}$$

(a Jacobi theta-function) and regard it as a Riemann sum of the integral,

$$\Theta\left(\frac{s+t}{2}, b\right) = b \sum_i \frac{1}{b} e^{-(i/b-\frac{s+t}{2b})^2} \approx b \int_{-\infty}^{+\infty} e^{-(z-\frac{s+t}{2b})^2} dz = b\sqrt{\pi} . \quad (61)$$

The value of the integral is obtained by variable substitution. The function $\Theta(x, b)$ is periodic in x with period one. The maximum is obtained for $x = \mathbb{Z}$ and the minimum for $x = \mathbb{Z} + 1/2$. The approximation is reasonable for $b > 0.5$ and good for $b > 1$. We have

$$\begin{aligned} \left| \frac{\Theta(x, 0.5)}{b\sqrt{\pi}} - 1 \right| &< 0.17 , \\ \left| \frac{\Theta(x, 1)}{b\sqrt{\pi}} - 1 \right| &< 2 \cdot 10^{-4} , \\ \left| \frac{\Theta(x, 1.5)}{b\sqrt{\pi}} - 1 \right| &< 10^{-9} . \end{aligned}$$

Combining (60) and (61) gives

$$C(s, t) \approx \varepsilon^2 \frac{1}{2\pi b^2} e^{-(s-t)^2/(4b^2)} b\sqrt{\pi} = \varepsilon^2 \frac{1}{2b\sqrt{\pi}} e^{-(s-t)^2/(4b^2)} . \quad (62)$$

This is a covariance function of a stationary random process, E_b , since it only depends on $s - t$, cf. (59). ■

Scale-space interpolation of the discrete random process is a stationary random process. The stationarity means that the interpolated error process is independent of the underlying discretization grid. Using Theorem 4.6 the covariance function of the first three derivatives of E_b are given by

$$\begin{cases} r_{E_b'}(t) = -r_{E_b}''(t), \\ r_{E_b}''(t) = r_{E_b}^{(iv)}(t), \\ r_{E_b}'''(t) = -r_{E_b}^{(vi)}(t), \end{cases} \quad (63)$$

where $r_E(t) = C(E(s), E(s+t))$, cf. [1]. Calculating the derivatives of

$$r_E(t) = \varepsilon^2 \frac{1}{2b\sqrt{\pi}} e^{-t^2/(4b^2)} \quad (64)$$

gives

$$\begin{aligned} r_{E_b'}(t) &= \varepsilon^2 \frac{1}{\sqrt{\pi}} e^{-t^2/(4b^2)} \left(\frac{1}{4b^3} - \frac{1}{8b^5} t^2 \right) , \\ r_{E_b}''(t) &= \varepsilon^2 \frac{1}{\sqrt{\pi}} e^{-t^2/(4b^2)} \left(\frac{3}{8b^5} - \frac{3}{8b^7} t^2 + \frac{1}{32b^9} t^4 \right) , \\ r_{E_b}'''(t) &= \varepsilon^2 \frac{1}{\sqrt{\pi}} e^{-t^2/(4b^2)} \left(\frac{15}{16b^7} - \frac{45}{32b^9} t^2 + \frac{15}{64b^{11}} t^4 + \frac{1}{128b^{13}} t^6 \right) . \end{aligned} \quad (65)$$

The variances which are needed later are given by

$$\begin{cases} r_{E_b'}(0) = \varepsilon^2 \frac{1}{4b^3\sqrt{\pi}} , \\ r_{E_b}''(0) = \varepsilon^2 \frac{3}{8b^5\sqrt{\pi}} , \\ r_{E_b}'''(0) = \varepsilon^2 \frac{15}{16b^7\sqrt{\pi}} . \end{cases} \quad (66)$$

Observe that the random fields of E_b' , E_b'' and E_b''' all have zero mean. It can also be shown that $E_b''(0)$ and $E_b'''(0)$ are independent, cf. [1].

Analysis of one-dimensional edges

The deterministic and stochastic analysis will now be used to analyze the distribution the edge-location in the one-dimensional case. As before the measured signal v_0 is modeled as a sum of a deterministic part w_0 and a random error process e_0 , i.e.

$$v_0 = w_0 + e_0 .$$

The edge position is defined as the position where the scale-space interpolation $V_b = S_b(I(v_0))$ has maximal derivative, i.e. as the zero crossing of $V_b''(x)$. Again, due to the *invariancy* of this approach, we can assume, without loss of generality, that the true edge is located at $d = 0$. Using (53) and (54), the undisturbed second order derivative of the edge, $W_b''(x)$, can be approximated near the zero crossing by the linear function

$$W_b''(x) \approx W_b'''(0)x = -\frac{A}{c^3\sqrt{2\pi}}x . \quad (67)$$

Near the zero crossing, the second order derivative calculated from the noisy image is approximated by the line

$$V_b''(x) = Kx + M , \quad (68)$$

where

$$\begin{cases} K = V_b'''(0), \\ M = V_b''(0) \end{cases} \quad (69)$$

are random variables whose first and second moments are given by

$$\begin{cases} \mathcal{E}[K] = W_b'''(0) = -\frac{A}{c^3\sqrt{2\pi}}, \\ \mathcal{V}[K] = r_{E'''}(0) = \varepsilon^2 \frac{15}{16b^7\sqrt{\pi}}, \\ \mathcal{E}[M] = W_b''(0) = 0, \\ \mathcal{V}[M] = r_{E''}(0) = \varepsilon^2 \frac{3}{8b^5\sqrt{\pi}} . \end{cases} \quad (70)$$

The line given by (68) has the zero-crossing

$$X = -\frac{M}{K} , \quad (71)$$

which again is a random variable. The probability distribution of X can be calculated explicitly or numerically. Notice that the mean and the variance of X is undefined. Nevertheless the probability distribution of X can be approximated by a normal distribution $N(m, \sigma)$ with

$$\begin{aligned} m &= -\frac{\mathcal{E}[M]}{\mathcal{E}[K]} = 0 , \\ \sigma^2 &= \frac{\mathcal{V}[M]}{\mathcal{E}[K]^2} . \end{aligned} \quad (72)$$

Here Gauss' approximation formulas are used together with the fact that M and K are independent, and that $\mathcal{E}[M] = 0$ according to (70). Combining (70) and (72) gives

$$\sigma^2 \approx \mathcal{V}[M]/\mathcal{E}[K]^2 = \varepsilon^2 \frac{3(a^2 + b^2)^3\sqrt{\pi}}{4A^2b^5} \quad (73)$$

which is the estimated variance of the detected edge. The approximations were made to obtain a simple expression for the edge localization variance. Notice that (73) is only valid when a and b are fairly large and when $\sigma[M] \ll \mathcal{E}[K]$. It is straightforward to calculate the distribution of X numerically and to get results which are valid for a greater range in the parameters. Nevertheless, (73) is a short analytical expression which describes the localization 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 10. In fact the choice of scale parameter b that minimizes the variance is $b = a\sqrt{5}$.

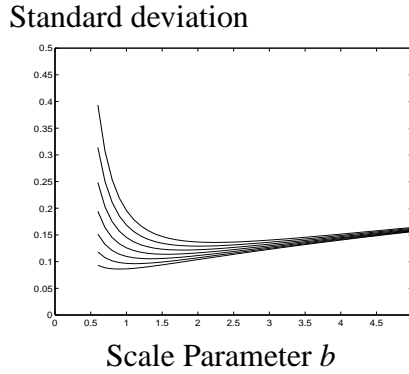


Figure 10: Theoretical standard deviation of edge localization error versus scale parameter b at different camera blur parameter $a = 0.4, 0.5, \dots, 1.0$. The intensity jump is $A = 50$ and the standard deviation in each pixel is $\varepsilon = 3$. The approximations in the calculations are good when b is larger than 0.8.

Simulations and experiments

The whole process of image formation and edge detection was simulated using the model developed in the previous sections. The detected edge positions for 500 different realizations are shown in Figure 11. This simulation indicates that the assumptions and the approximations used are reasonably valid. Preliminary results also indicate that sharp edges can be detected with a standard deviation of about $1/20$ 'th of a pixel.

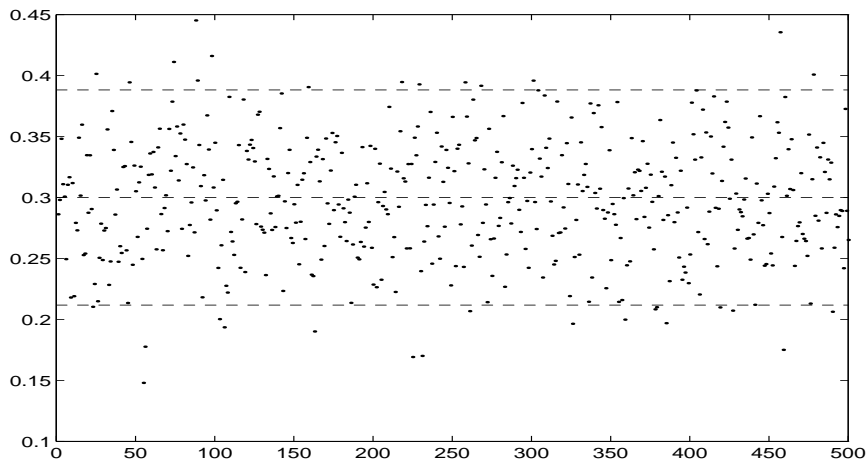


Figure 11: Result after simulating and detecting 500 one-dimensional edges using parameters $A = 100$, $a = 0.8$, $b = 1.1$, $\varepsilon = 2$. The theoretical mean and the middle 95% of the theoretical distribution are shown. This band was derived from the calculations above. In this simulation 23 of 500 detected edge positions fall outside the this region.

The one-dimensional analysis above can be applied directly to horizontal and vertical edges in a two-dimensional image. Each row or column is analyzed separately. This is illustrated in Figure 12, in which a real image containing a sharp edge between a light paper and a dark background has been studied. The sub-pixel edge location was found for each of 170 rows, independently of the others. The x positions for the different rows are denoted $x(k)$, $k = 230, \dots, 400$. Since we a priori know that the edge has small curvature we can use standard regression techniques to fit a low order polynomial $(x_0 + x_1k + x_2k^2)$ to the sub-pixel edge positions $x(k)$. The residual

$$r(k) = x(k) - (x_0 + x_1k + x_2k^2) ,$$

can then be used to estimate the standard deviation of the edge locations. This was repeated for several choices of scale parameter b . This empirical standard deviation will now be compared to the theoretical. Some of the parameters of (73) are easy to estimate, e.g. the intensity jump A . In this image the jump was approximately $A =$

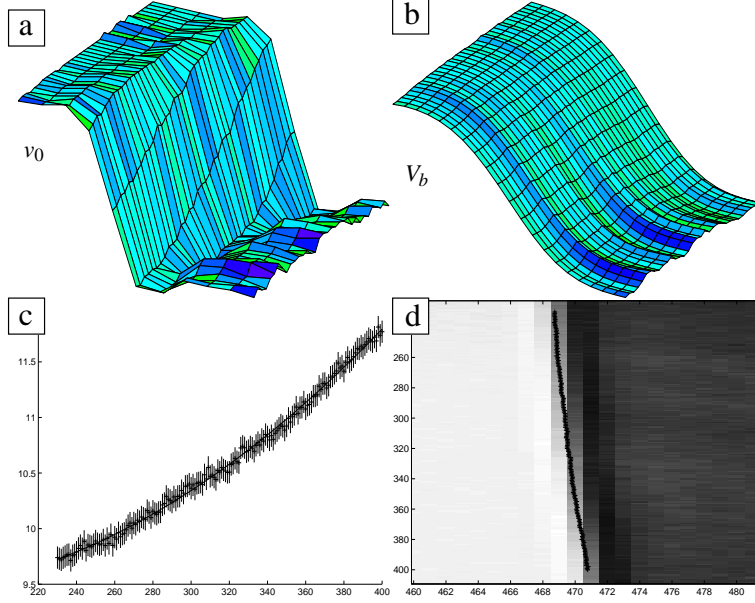


Figure 12: The figure illustrates the application of the one-dimensional edge detector to an almost vertical edge in a real image of a white paper against a dark background. The one-dimensional analysis is applied to each row, $230 \leq k \leq 400$, in the image independent of the others. a: A mesh plot of the original discrete image v_0 . b: A continuous interpolation v_b is done for each row. c: For each row the x coordinate of the edge is estimated. A low order polynomial $\tilde{x}(k)$ is fitted to the edge positions $x(k)$. using $a \approx 0.6$, $b = 2$, $A \approx 150$ and $\varepsilon \approx 3$. The figure shows the polynomial $\tilde{x}(k)$, the estimated positions $x(k)$ with a 95% confidence interval for each row k . d: The figure shows the original image and the estimated edge-points positions.

150. The standard deviation of image intensity can also be estimated in regions where the intensity is approximately constant. In this image we had $\varepsilon \approx 3$. The camera blur parameter a is more difficult to estimate. Using $a = 0.6$, a good fit is obtained between the standard deviation computed from (73) and its empirical value. This is shown in Figure 13. This strongly supports that the model and the approximations are valid. It is also one way to estimate the unknown parameter a in the image acquisition model.

6 Two-dimensional edge detection

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

$$(\nabla V_b)^T (\nabla^2 V_b) \nabla V_b = 0, \quad (74)$$

cf. [18]. 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

$$(\tilde{n})^T (\nabla^2 V_b) \tilde{n} = 0. \quad (75)$$

Using Assumptions 2.2 and 2.3, the smoothed signal V_b can be written as sum of a deterministic signal W_b and a stationary random field E_b . These parts are studied separately.

Deterministic part

An ideal edge is modeled as

$$\mathbb{R}^2 \ni (x, y) \mapsto W_{\text{ideal}}(x, y) = A\theta(x) \in \mathbb{R}. \quad (76)$$

Standard deviation

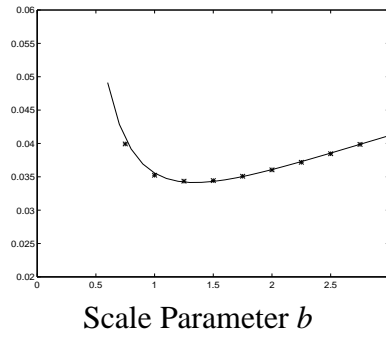


Figure 13: Edge localization error $\sigma[X]$ versus scale parameter b . The curve shows the theoretical standard deviation, cf. Equation 73 with $a = 0.6$, $A = 150$ and $\varepsilon = 3$. The stars show the empirical estimate using residuals from a real image at smoothing scales $b = 0.75$ in steps of 0.25 up to $b = 3$. Notice the close fit.

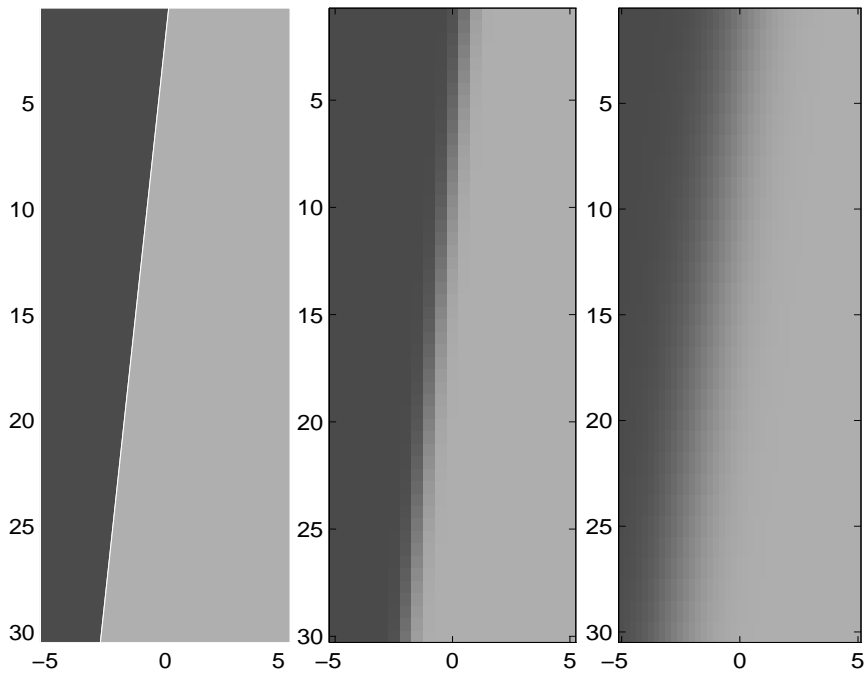


Figure 14: In this figure we illustrate the ideal intensity function W_{ideal} defined at all points, the discrete image w_0 defined at integer points and the scale space smoothed intensity distribution defined at all points W .

The deterministic discretized image w_0 of this edge is

$$\mathbb{Z}^2 \ni (i, j) \mapsto w_0 = D(h * W_{\text{ideal}})(i, j) \approx A\Phi_a(i) \in \mathbb{R} , \quad (77)$$

where Φ_a is the one-dimensional normal cumulative distribution function. Below the scale-space interpolation $W_b = S_b(I(w_0))$ will be studied.

$$W_b = S_b(h * W_{\text{ideal}}) \approx S_c W_{\text{ideal}} , \quad (78)$$

where $c = \sqrt{a^2 + b^2}$. In the two-dimensional case we estimate the edge as the locus of the points where the directional derivative has a local maxima on a line with direction $\tilde{\mathbf{n}}$. If we approximate W_b as above we find that it is constant along the edge. Consider the derivatives in the direction $\tilde{\mathbf{n}} = (\cos(\alpha), \sin(\alpha))$, where α denotes the angular difference between the search direction $\tilde{\mathbf{n}}$ and the normal $\mathbf{n} = (1, 0)$ to the edge. Introduce

$$F(t) = W_b(t \cos(\alpha), t \sin(\alpha)) .$$

In the following analysis we will need F and its first three derivatives:

$$\begin{aligned} F(t) &= (W * G_c)(t \cos(\alpha)) , \\ F'(t) &= A \cos(\alpha) G_c(t \cos(\alpha)) , \\ F''(t) &= A \cos^2(\alpha) \left(-\frac{t \cos(\alpha)}{c^2}\right) G_c(t \cos(\alpha)) , \\ F'''(t) &= A \cos^3(\alpha) \left(-\frac{1}{c^2} + \frac{(t \cos(\alpha))^2}{c^4}\right) G_c(t \cos(\alpha)) . \end{aligned} \quad (79)$$

Notice that the first derivative has a maximum for $t = 0$, independently of α , since $F'(0) = 0$ and the slope of the second order derivative at the zero crossing is

$$F'''(0) = -\frac{A \cos^3(\alpha)}{c^3 \sqrt{2\pi}} < 0 . \quad (80)$$

Random part

Assume that discrete white noise e_0 is added to the image. According to Theorem 4.5, the scale-space interpolated error field E_b then is stationary with covariance

$$r_{E_b}(\tau) = \text{sinc} * G_{b\sqrt{2}} . \quad (81)$$

Lemma 3.1 is used to simplify the calculations:

$$\text{sinc} * G_{b\sqrt{2}} \approx \varepsilon^2 \frac{1}{4\pi b^2} e^{-|\tau|^2/4b^2} . \quad (82)$$

The covariance function is the covariance between the intensity at two positions (x, y) and $(x + \tau_x, y + \tau_y)$. We also need the covariance functions of the first three directional derivatives of $E_b(s, t)$. Since this random field is approximately isotropic for large b , it is sufficient to calculate the directional derivatives in the s -direction. Theorem 4.6 gives

$$\begin{cases} r_{E'_s}(\rho) = -\frac{\partial^2 r_E}{\partial^2 s}(s, t), \\ r_{E''_{ss}}(\rho) = \frac{\partial^4 r_E}{\partial^4 s}(s, t), \\ r_{E'''_{sss}}(\rho) = -\frac{\partial^6 r_E}{\partial^6 s}(s, t). \end{cases} \quad (83)$$

Calculating the derivatives gives

$$\begin{aligned} r_{E'_s}(s, t) &= \varepsilon^2 \frac{1}{\pi} e^{s^2/4b^2} \left(\frac{1}{8b^4} - \frac{1}{16b^6} s^2 \right) e^{-t^2/4b^2} , \\ r_{E''_{ss}}(s, t) &= \varepsilon^2 \frac{1}{\pi} e^{-s^2/4b^2} \left(\frac{3}{16b^6} - \frac{3}{16b^8} s^2 + \frac{1}{64b^{10}} s^4 \right) e^{-t^2/4b^2} , \\ r_{E'''_{sss}}(s, t) &= \varepsilon^2 \frac{1}{\pi} e^{-s^2/4b^2} \left(\frac{15}{32b^8} - \frac{45}{64b^{10}} s^2 + \frac{15}{128b^{12}} s^4 + \frac{1}{256b^{14}} s^6 \right) e^{-t^2/4b^2} . \end{aligned} \quad (84)$$

The variances are given by

$$\begin{cases} r_{E_s'}(0,0) = \varepsilon^2 \frac{1}{8b^4\pi}, \\ r_{E_{ss}''}(0,0) = \varepsilon^2 \frac{3}{16b^6\pi}, \\ r_{E_{sss}'''}(0,0) = \varepsilon^2 \frac{15}{32b^8\pi}. \end{cases} \quad (85)$$

The random fields E_s' , E_{ss}'' and E_{sss}''' have zero mean. Furthermore, $E_{ss}''(0,0)$ and $E_{sss}'''(0,0)$ are independent, cf. [1]. The analysis above is only valid when we search in a direction perpendicular to the edge. If the search direction forms the angle α with the edge normal, we have to evaluate $r_{E_{ss}''}(s,t)$ and the other directional derivatives in $(s,t) = (\tau \cos(\alpha), \tau \sin(\alpha))$, just as in the deterministic case. This gives the following expressions for the dependencies of the directional derivatives along the line

$$\begin{aligned} r_{E_s'}(\tau) &= \varepsilon^2 \frac{1}{\pi} e^{-\tau^2/4b^2} \left(\frac{1}{8b^4} - \frac{1}{16b^6} \tau^2 \sin^2 \alpha \right), \\ r_{E_{ss}''}(\tau) &= \varepsilon^2 \frac{1}{\pi} e^{-\tau^2/4b^2} \left(\frac{3}{16b^6} - \frac{3}{16b^8} \tau^2 \sin^2 \alpha + \frac{1}{64b^{10}} \tau^4 \sin^4 \alpha \right), \\ r_{E_{sss}'''}(\tau) &= \varepsilon^2 \frac{1}{\pi} e^{-\tau^2/4b^2} \left(\frac{15}{32b^8} - \frac{45\tau^2 \sin^2 \alpha}{64b^{10}} + \frac{15\tau^4 \sin^4 \alpha}{128b^{12}} + \frac{\tau^6 \sin^6 \alpha}{256b^{14}} \right). \end{aligned} \quad (86)$$

Analysis of two-dimensional edges

Similar to the one-dimensional case we now can calculate the distribution of the edge location. This time the second order directional derivative, $F''(x)$, also depends on the angle α between the search direction and the edge normal. Close to the edge position, $F''(x)$ can be approximated by the line

$$y = kx = F''(0)x = -\frac{A \cos^3(\alpha)}{c^3 \sqrt{2\pi}} x. \quad (87)$$

The second order derivative calculated from the noisy image is again approximated near the zero crossing by a line

$$V_b''(x) = Kx + M, \quad (88)$$

where

$$\begin{cases} K = V_b'''(0,0), \\ M = V_b''(0,0) \end{cases} \quad (89)$$

are random variables with

$$\begin{cases} \mathcal{E}[K] = W_b'''(0) = -\frac{A \cos^3(\alpha)}{c^3 \sqrt{2\pi}}, \\ \mathcal{V}[K] = r_{E_{sss}'''}(0) = r_{E_{sss}'''}(0,0) = \varepsilon^2 \frac{15}{32b^8\pi}, \\ \mathcal{E}[M] = W_b''(0) = 0, \\ \mathcal{V}[M] = r_{E_{ss}''}(0,0) = \varepsilon^2 \frac{3}{16b^6\pi}. \end{cases} \quad (90)$$

This line (88) has the zero-crossing

$$X = -\frac{M}{K}, \quad (91)$$

which is a random variable. The probability distribution of X can be approximated by the normal distribution $N(m, \sigma)$ with

$$\begin{aligned} m &= -\frac{\mathcal{E}[M]}{\mathcal{E}[K]} = 0, \\ \sigma &= \frac{\mathcal{V}[M]}{\mathcal{E}[K]^2}, \end{aligned} \quad (92)$$

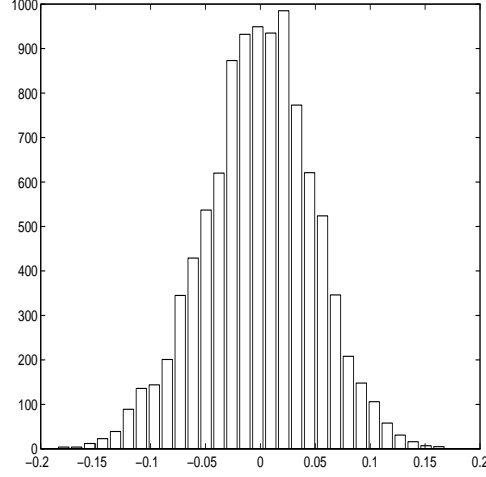


Figure 15: Resulting residual histogram for simulated data. The theoretical standard deviation of 0.0499 agrees with the experimentally estimated of 0.0498.

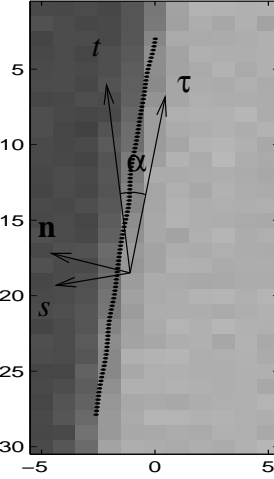


Figure 16: The figure illustrates some notations used in the analysis of two-dimensional edge detection.

where Gauss approximation formulas are used together with the fact that M and K are independent, and $\mathcal{E}[M] = 0$. Combining (90) and (91) gives

$$\mathcal{V}[X] \approx \mathcal{V}[M]/\mathcal{E}[K]^2 = \epsilon^2 \frac{3(a^2 + b^2)^3}{8A^2b^6 \cos^6(\alpha)} . \quad (93)$$

This is the estimated variance of the detected edge. Observe that the variance decreases with increasing height, A , of the edge. The variance also increases when α increases, that is when we do not search perpendicularly to the line.

The detected edge as a random process

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

$$z(\tau) = \frac{(V_b)''_{ss}(\gamma(\tau))}{(W_b)'''_{sss}(\gamma(\tau))} \quad (94)$$

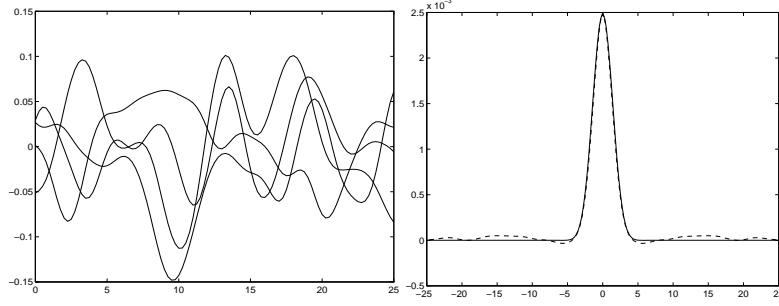


Figure 17: Results from two-dimensional edge detection with simulated data. Left: Edge position errors in a direction roughly perpendicular to the edge at different positions along the edge. The result from four simulations are shown. Distant errors are not correlated but there is a high correlation between edge position errors at spatially close positions. The residuals can be modeled as samples of a random process with respect to the parametrization of the curve. Right: Theoretical and estimated covariance functions for the residual error process.

where

$$(W_b)'''_{ss}(\gamma(\tau)) = -\frac{A \cos^3(\alpha)}{c^3 \sqrt{2\pi}} .$$

Then the covariance between the deviations $z(\tau_1)$ and $z(\tau_2)$ is

$$C[z(\tau_1), z(\tau_2)] = C[(V_b)''_{ss}(\gamma(\tau_1)), (V_b)''_{ss}(\gamma(\tau_2))] \frac{c^6 2\pi}{A^2 \cos^6 \alpha} , \quad (95)$$

where

$$C[(V_b)''_{ss}(\gamma(\tau_1)), (V_b)''_{ss}(\gamma(\tau_2))] = r_{E'''_{sss}}((\tau_1 - \tau_2) \sin \alpha, (\tau_1 - \tau_2) \cos \alpha)$$

in (86). Hence, z is a stationary process with covariance function

$$\begin{aligned} r_z(\tau) &= r_{E'''_{sss}}(\tau \sin(\alpha), \tau \cos(\alpha)) / F'''(0) = \\ &= \varepsilon^2 \frac{2c^6}{A^2 \cos^6 \alpha} e^{-\tau/(4b^2)} \left(\frac{3}{16b^6} - \frac{3}{16b^8} (\tau \sin(\alpha))^2 + \frac{1}{64b^{10}} (\tau \sin(\alpha))^4 \right) . \end{aligned} \quad (96)$$

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

$$r_z(\tau) \approx 1.1 \varepsilon^2 \frac{3(a^2 + b^2)^3}{8A^2 b^6} e^{-\tau^2/(4b^2)} . \quad (97)$$

Notice that the parameter τ is measured as the arclength along the line.

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 [19].

Implementation and experiments

The two-dimensional edge detector described above has been implemented. Its performance on both simulated and real images have been investigated. In the simulations the true edge was well defined. The deviations z were studied both with respect to different realizations but also as a random process along the edge.

Figure 14 illustrates the original intensity W , the discrete measured intensity w_0 and the smoothed image intensity W_b . The discrete image was disturbed with simulated Gaussian uncorrelated noise. This image was then used in the edge detection routines to calculate the edges along lines roughly perpendicular to the true edge. In these simulations the search line was 0.022 radians off the normal. This was done at several positions along the

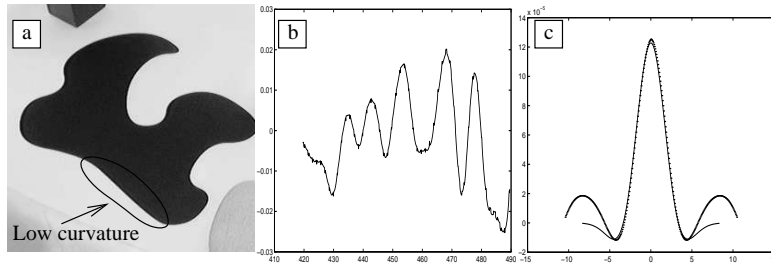


Figure 18: Implementation on real image. Along curved parts with small curvature, the assumption that the curve is approximately linear holds. The residual after fitting a low degree polynomial to the edge is an approximately random process along the curve. The empirical covariance function is reasonably close to the theoretical one.

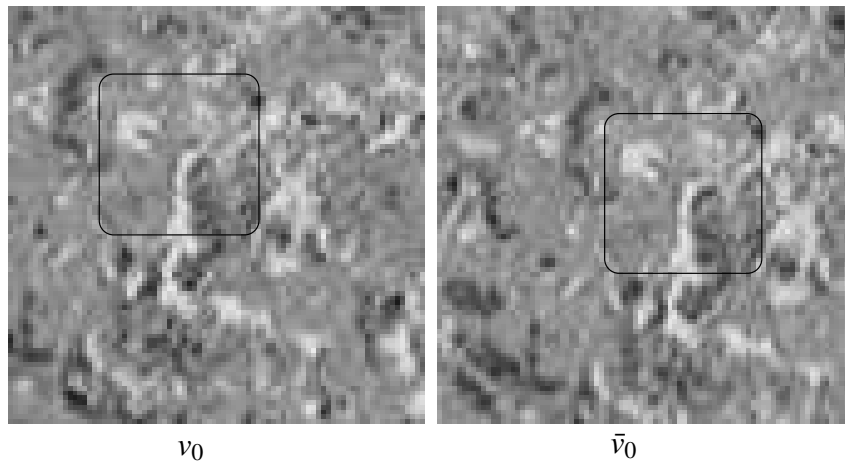


Figure 19: Two real images of a highly textured floor. These are used to illustrate correlation.

true edge. Figure 15 shows a histogram of the residuals at one position for several realisations. The theoretical and empirical standard deviations agree well. The figure also shows the edge positions with confidence intervals.

Figure 17 illustrates four different realizations of the deviation z in search direction, from the true edge as we move along the edge. The deviation $z(\tau)$ is a stationary random process. Edge position errors at distance τ have covariance $r_z(\tau)$. The theoretical and empirical covariance functions agree well as can be seen in the figure.

The edge detector was also applied to a real image of a dark object against a light background, cf. Figure 18.a. Part of the contour is smooth and has small curvature. The extracted curve at this patch was analyzed and the residuals along the curve were estimated in the following way. The curved part, approximately 100 pixels long, was rotated to become almost horizontal. The original object, which had been drawn on a high resolution laser printer, was known to have very small variations in curvature. This motivated fitting a third order polynomial to the curve. The residual is shown in Figure 18.b, and defines a stationary random process. The empirical covariance function was estimated from the residuals using standard techniques. Figure 18.c shows the resulting empirical covariance function together with the covariance computed from the model (96) using the following parameters: $A = 170$, $a = 0.7$, $b = 2$, $\varepsilon = 1.7$, and $\alpha = 30$ degrees. The theory agrees well with the empirical results in this case.

Our analysis was based on ideal, straight step discontinuities. In this case the edge detector is unbiased. Bias is expected for more realistic situations and at edge points with considerable curvature.

7 Sub-pixel correlation

Analysis of sub-pixel correlation is another application of the theory presented here. Correlation is usually done on pixel level, where a region 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 minimized. The result after least squares minimization in a typical case is illustrated in Figure 20. The stochastic errors of pixel correlation is difficult to analyze, mainly because the translation between the regions in the two images usually is of sub-pixel type.

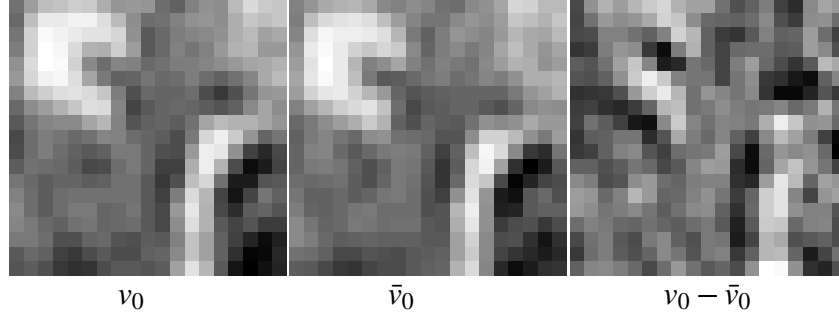


Figure 20: Regions in two images are correlated with pixel translations using least squares of the residuals.

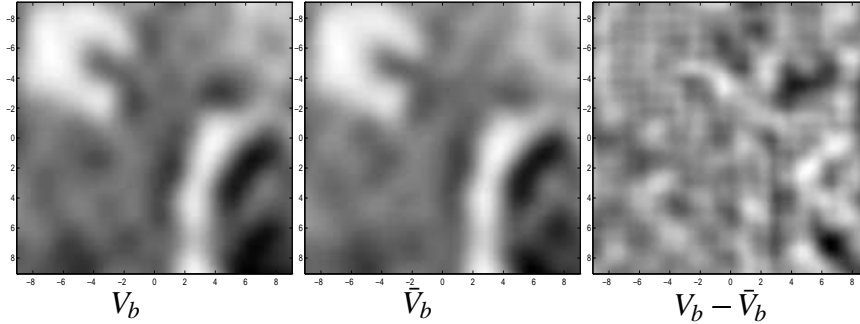


Figure 21: Regions in two images are correlated with sub-pixel translations using least squares of the residuals of the restored continuous scale-space representation at scale $b = 0.9$. The accuracy of the sub-pixel translation can be analyzed through stochastic models of the residual field. The residual $W - \bar{W}$ can also be used to empirically estimate the stochastic properties of the error field E_b . The gray level scale is different for the residual $W - \bar{W}$.

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 modeling of the residual field makes it possible to analyze the stochastic properties of the localization error. The idea is that, at least locally, the images only differ by an unknown translation h . Denote by $V = W + E$ and $\bar{V} = \bar{W} + \bar{E}$ the restored intensity fields in two images for a fixed scale b . The deterministic functions are identical except for a translation. For a fixed translation $h = (h_1, h_2)$, we thus have

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

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

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

The result of such a minimization is shown in Figure 21.

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

Statistical analysis

The quality of the estimated sub-pixel translation,

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

can be analyzed using the statistical model given above. Let $X = \hat{h} - h$ be the error in estimated translation. Without loss of generality we can assume that h is zero. The stochastic variable X is a result of minimising the function F , which in turn contains the random fields E and \bar{E} . What is the mean and covariance matrix of X ? Ideally the mean should be zero, indicating that \hat{h} is a non-biased estimate of h , and the covariance should hopefully be small. Linearising F around the true displacement $h = 0$, we find that the squared residual F can be approximated as

$$F(X) \approx X^T A X + bX + f , \quad (98)$$

where A , b and f are stochastic variables

$$\begin{aligned} f &= F(0) \ , \\ b &= \nabla_h F(0) \ , \\ A &= \nabla_h^2 F(0) \ . \end{aligned}$$

The minimum of (98) is attained for

$$X = -A^{-1}b \ .$$

Gauss' approximation formula should be appropriate in this situation, since b is small and A is large with small covariance. Hence

$$m = \mathcal{E}[X] \approx \mathcal{E}[A]^{-1} \mathcal{E}[b] \ .$$

The gradient b at the correct displacement x is

$$\begin{aligned} b &= \nabla \int_{t \in \Omega} (W(t) + E(t) - (\bar{W}(t+x) + \bar{E}(t+x)))^2 dt = \\ &= \nabla \int_{t \in \Omega} ((W - \bar{W}) + (E - \bar{E}))^2 dt = \\ &= \nabla \int_{t \in \Omega} ((W - \bar{W})^2 + (W - \bar{W})(E - \bar{E}) + (E - \bar{E})^2) dt = \\ &= \int_{t \in \Omega} -2(W - \bar{W})\nabla\bar{W} + (W - \bar{W})\nabla\bar{E} + \\ &\quad + (E - \bar{E})\nabla\bar{W} + 2(E - \bar{E})\nabla\bar{E} dt = \\ &= \int_{t \in \Omega} -((E - \bar{E})\nabla\bar{W} + 2(E - \bar{E})\nabla\bar{E}) dt \ . \end{aligned}$$

The expectation value of b is

$$\begin{aligned} \mathcal{E}[b] &= \mathcal{E} \left[\int_{t \in \Omega} -((E - \bar{E})\nabla\bar{W} + 2(E - \bar{E})\nabla\bar{E}) dt \right] = \\ &= \int_{t \in \Omega} -(\mathcal{E}[E - \bar{E}]\nabla\bar{W} + 2\mathcal{E}[(E - \bar{E})]\mathcal{E}[\nabla\bar{E}]) dt = 0 \ . \end{aligned}$$

This means that $m = 0$. Using the Gauss approximation formulas and the fact that $\mathcal{E}[b] = 0$, we obtain

$$C = C[X, X] \approx \mathcal{E}[A]^{-1} C[b] \mathcal{E}[A]^{-T} + \text{terms involving } \mathcal{E}[b] = \mathcal{E}[A]^{-1} C[b] \mathcal{E}[A]^{-T} \ .$$

The covariance of the gradient is

$$\begin{aligned} C[b] &= \mathcal{E}[bb^T] = \mathcal{E} \left[\int_{t_1 \in \Omega} V(t_1)(E - \bar{E})(\nabla\bar{W} + \nabla\bar{E}) dt_1 + \right. \\ &\quad \left. + \int_{t_2 \in \Omega} V(t_2)(E - \bar{E})(\nabla\bar{W} + \nabla\bar{E})^T dt_2 \right] = \\ &= \mathcal{E} \left[\int_{t_1 \in \Omega} \int_{t_2 \in \Omega} V(t_1)V(t_2)(E - \bar{E})(t_1)(E - \bar{E})(t_2) \right. \\ &\quad \left. (\nabla\bar{W} + \nabla\bar{E})(t_1)(\nabla\bar{W} + \nabla\bar{E})^T(t_2) dt_1 dt_2 \right] \approx \\ &\approx \int_{t_1 \in \Omega} \int_{t_2 \in \Omega} V(t_1)V(t_2)r_{E-\bar{E}}(t_1 - t_2)\nabla\bar{W}(t_1)\nabla\bar{W}^T(t_2) dt_1 dt_2 = \\ &= \int_{t_1 \in \Omega} (V\nabla\bar{W})(t_1)(V\nabla\bar{W}^T * r_{E-\bar{E}})(t_1) dt_1 \ . \end{aligned}$$

The Hessian A is given by

$$\begin{aligned}
A &= \nabla^2 \int_{t \in \Omega} ((W - \bar{W}) + (E - \bar{E}))^2 dt = \\
&= \nabla \int_{t \in \Omega} ((W - \bar{W})^2 + 2(W - \bar{W})(E - \bar{E}) + (E - \bar{E})^2) dt = \\
&= 2 \int_{t \in \Omega} (\nabla \bar{W} (\nabla \bar{W})^T + (W - \bar{W})(-\nabla^2 \bar{W}) + \nabla \bar{W} (\nabla \bar{E})^T + \\
&\quad + (W - \bar{W})(-\nabla^2 \bar{E}) + \nabla \bar{E} (\nabla \bar{W})^T + (E - \bar{E})(-\nabla^2 \bar{W}) + \\
&\quad + \nabla \bar{E} (\nabla \bar{E})^T + (E - \bar{E})(-\nabla^2 \bar{E})) dt = \\
&= 2 \int_{t \in \Omega} (\nabla \bar{W} (\nabla \bar{W})^T + \nabla \bar{W} (\nabla \bar{E})^T + \nabla \bar{E} (\nabla \bar{W})^T + \\
&\quad + (E - \bar{E})(-\nabla^2 \bar{W}) + \nabla \bar{E} (\nabla \bar{E})^T + (E - \bar{E})(-\nabla^2 \bar{E})) dt .
\end{aligned}$$

The expectation value of the Hessian is

$$\begin{aligned}
\mathcal{E}[A] &= 2 \int_{t \in \Omega} (\nabla \bar{W} (\nabla \bar{W})^T + \nabla \bar{W} (\mathcal{E}[\nabla \bar{E}])^T + \mathcal{E}[\nabla \bar{E}] (\nabla \bar{W})^T + \\
&\quad + \mathcal{E}[(E - \bar{E})(-\nabla^2 \bar{W})] + \mathcal{E}[\nabla \bar{E} (\nabla \bar{E})^T] + \mathcal{E}[(E - \bar{E})(-\nabla^2 \bar{E})]) dt = \\
&= 2 \int_{t \in \Omega} (\nabla \bar{W} (\nabla \bar{W})^T) dt .
\end{aligned}$$

The simplifications are based on two facts. Firstly, the expectation value of the random fields and their derivatives are zero. Secondly, we have

$$\mathcal{E}[\bar{E} \nabla^2 \bar{E}] = -\mathcal{E}[\nabla \bar{E} (\nabla \bar{E})^T] = \begin{bmatrix} r''_{11} & r''_{12} \\ r''_{21} & r''_{22} \end{bmatrix} .$$

The method for doing sub-pixel correlation is thus unbiased:

$$\mathcal{E}[X] = 0 ,$$

and the covariance matrix of the residual is approximately given by

$$C = \mathcal{C}[X] \approx \mathcal{E}[A]^{-1} \mathcal{C}[b] \mathcal{E}[A]^{-T} , \quad (99)$$

where

$$\mathcal{E}[A] = 2 \int_{t_1 \in \Omega} (V \bar{W} \bar{W}^T)(t_1) dt_1 \quad (100)$$

and

$$\mathcal{C}[b] = \int_{t_1 \in \Omega} ((V \bar{W}) * r_{E-\bar{E}})(t_1) (V \bar{W})(t_1) dt_1 . \quad (101)$$

Experimental validation

Since many approximations were made to obtain the expression for the covariance (99) of the sub-pixel estimate, numerical simulations were made to validate the assumptions. The image in Figure 19 was used as a model of the ideal intensity W_{ideal} . The process of image formation was simulated with blur, discretization and added stochastic errors as described previously. In these simulations the true displacement between the image W and \bar{W} was $h = (0.3, 0.2)$. A simulation of 1000 correlation events gave the following empirical covariance matrix

$$C = 10^{-3} \begin{bmatrix} 0.1826 & -0.0061 \\ -0.0061 & 0.3983 \end{bmatrix} .$$

This corresponds to a standard deviation of 0.02 pixels in the estimated translation. Evaluating (99) we obtain the following estimate of the covariance matrix

$$C = 10^{-3} \begin{bmatrix} 0.1829 & -0.0074 \\ -0.0074 & 0.4022 \end{bmatrix} .$$

This agrees surprisingly well with the empirical result.

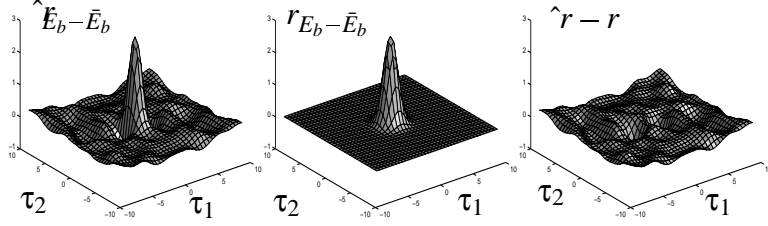


Figure 22: The figure shows the empirical estimate of the covariance function $\hat{r}_{E-\bar{E}}(\tau)$ and the theoretical covariance function $r_{E-\bar{E}}(\tau)$ given by (40) under the assumption that e_0 and \bar{e} are Gaussian white noise with standard deviation $\varepsilon = 1.67$. The difference between empirical and theoretical covariance functions is also shown.

Residual analysis

In order to test our model and to gain further empirical knowledge about the error random field, several images were taken of the same scene from almost the same view-point. The idea was that, at least locally, the ideal intensity W_{ideal} would be identical except for a translatory movement in the image plane. Thus we would have

$$W_b^1(t_1, t_2) = W_b^2(t_1 + \Delta t_1, t_2 + \Delta t_2)$$

for a fixed translation $\Delta t = (\Delta t_1, \Delta t_2)$. If our model describes the situation well, we should be able to estimate the translation Δt with sub-pixel accuracy by minimizing a least squares integral

$$F(\Delta t) = \int_{t \in \Omega} (W_b^1(t) - W_b^2(t + \Delta t))^2 dt$$

as in the previous section. The residual field

$$E_b^1(t) - E_b^2(t + \Delta t)$$

is then analyzed, to determine if it can be represented as a stationary random field. We should also be able to estimate its covariance function.

These investigations have been performed with several real images. One case is illustrated in Figure 22. The experiments indicate that the estimated covariance function is close to the theoretical covariance function using $\varepsilon = 1.67$.

8 Conclusions

In this paper we have modeled the image acquisition process, taking into account both the deterministic and stochastic aspects. In particular the discretization process is modeled in detail. This interplay between the continuous signal and its discretization is very fruitful and the increased knowledge sheds light on scale-space theory, feature detection and stochastic modeling of errors.

The relation between the continuous signal and its discretization 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 modeled 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 analyzed 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 modeled 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 analyze other feature detectors.

Acknowledgments

The authors thank Georg Lindgren for inspiration and valuable discussions.

References

- [1] A. Adler. *The Geometry of Random Fields*. Wiley, New York, 1985.
- [2] L. V. Ahlfors. *Complex Analysis*. McGraw-Hill Book Company, 1965.
- [3] 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.
- [4] N. K. Bose. *Digital Filters, Theory and Applications*. Elsevier Science Publishing Co, Inc, 1985.
- [5] F. Canny. A computational approach to edge detection. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 8(6):676–698, 1986.
- [6] H. Cramér and M. R. Leadbetter. *Stationary and Related Stochastic Processes*. Wiley, New York, 1967.
- [7] N. A. C. Cressie. *Statistics for Spatial Data*. Wiley, New York, 1991.
- [8] 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.
- [9] R. Deriche. Using Canny’s criteria to derive an optimal edge detector recursively implemented. *Int. Journal of Computer Vision*, 1:167–187, 1987.
- [10] O. Faugeras. *Three-Dimensional Computer Vision*. MIT Press, Cambridge, Mass, 1993.
- [11] G. H. Hardy, J. E. Littlewood, and G. Polya. *Inequalities*. Cambridge University Press, third edition, 1959.
- [12] E. Hecht. *Optics*. Addison-Wesley, Reading, Mass., 1987.
- [13] E. Hildreth and D. Marr. Theory of edge detection. *Proceedings of Royal Society of London*, 207:187–217, 1980.
- [14] T. Iijima. Basic theory on normalization of a pattern (in case of typical one-dimensional pattern). *Bulletin of Electrical Laboratory*, 26:368–388, 1980. in japanese.
- [15] J. J. Koenderink. The structure of images. *Biological Cybernetics*, 50:363–370, 1984.
- [16] V. A. Kotelnikov. On the transmission capacity of ‘ether’. In *Proc. First All-union Conference on Questions of Communications*, 1933.
- [17] Deepa Kundur and Dimitrios Hatzinakos. Blind image deconvolution. *IEEE signal processing magazine*, May 1996.
- [18] T. Lindeberg. *Scale-Space Theory in Computer Vision*. Kluwer Academic Publishers, 1994.
- [19] G. Lindgren and I. Rychlik. How reliable are contour curves - confidence sets for level contours. *Bernoulli*, 1(2), 1995.
- [20] D. Marr. *Vision*. W. H. Freeman and Company, 1982.
- [21] V. S. Nalwa and T. O. Binford. On detecting edges. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 8:699–714, 1986.
- [22] H. Nyquist. Certain topics in telegraph transmission theory. *AIEE Trans.*, 47, 1928.

- [23] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 12:629–639, 1990.
- [24] B. Pratt. *Digital Image Processing*. Wiley-Interscience, 1978.
- [25] G. Sapiro and A. Tannenbaum. An affine invariant scale-space. *Journal of Functional analysis*, 1993.
- [26] C. E. Shannon. Communication in the presence of noise. *Proc. IRE*, 37, 1949.
- [27] V. Torre and A. Poggio. On edge detection. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 8:147–163, 1986.
- [28] A. P. Witkin. Scale-space filtering. In *Proc. Eighth International Joint Conference on Artificial Intelligence, Karlsruhe, West Germany*, pages 1019–1022, 1983.