Stochastic Analysis of Scale-space Smoothing*

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

Abstract

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

1 Introduction

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

In this paper, image acquisition, interpolation and scale-space smoothing are modelled into some detail. *Image acquisition* is viewed as a composition of blurring, ideal sampling and added noise, similar to [10]. The discrete signal is analysed after *interpolation*. This makes it possible to detect features on a sub-pixel basis. Averaging or *scale-space smoothing* is often used to reduce the effects of noise. To understand feature detection in this framework, one has to analyse the effect of noise on interpolated and smoothed signals. In doing so a theory is obtained that connects the discrete and continuous scale-space theories.

*This work has been supported by the Swedish Research Council for Engineering Sciences (TFR), project 95-64-222

2 Image acquisition

To model the image acquisition process, the intensity distribution that would be caught by an ideal camera is first affected by aberrations in the optics of the real camera, e.g. blurring caused by spherical aberration, coma and astigmatism. Other aberrations deform the image, like Petzval field curvature and distorsion, see [7]. Such distorsion can typically be handled by geometric considerations in mid-level vision and will not be commented upon here. One way to model camera blur is to convolve the ideal intensity distribution with a kernel corresponding to the smoothing caused by the camera optics. This process also removes some amount of the high spatial frequencies.

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

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

$$W_{\mathrm{ideal}} \xrightarrow{\text{blur}} W \xrightarrow{\text{sampling}} w_0 \xrightarrow{\text{noise}} v_0 \ , \tag{1}$$

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

Here *blurring* is modelled as an abstract operator h, such that $W = h(W_{\mathrm{ideal}})$. We assume that no aliasing effects are present, when the function W is sampled at integer positions, i.e. $W \in \mathcal{B}(\mathbb{R}^n)$, where

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

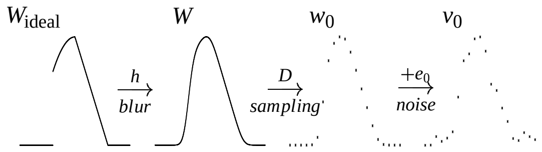

Figure 1. Illustration of image acquisition.

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

$$\mathcal{F}W(f) = \int_{\mathbb{R}^n} W(\tau)e^{-i2\pi f\cdot\tau}d\tau \ , \tag{3}$$

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

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

$$w(i,j) = (DW)(i,j) = W(i,j) \ . \tag{4}$$

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

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

These assumptions will serve as an initial model. Further improvements can be made by a more detailed camera acquisition model.

3 Interpolation and smoothing

Scale-space theory and its application to computer vision is discussed briefly in this section. A more thorough treatment is given in [9]. The idea is to associate to each signal a family of signals smoothed to different degrees. Each such signal captures the behaviour of the signal at one scale. The idea of smoothing is useful to attenuate high-frequency noise without disturbing the low-frequency components of the signal. There is a trade-off in choosing the smoothing parameter. The real strength in using the scale-space approach is the possibility to study the whole scale-space representation, This will, however, not be pursued in this paper. The emphasis will be made on the stochastic properties of each scale-space representation separately.

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

$$G_b(x) = \frac{1}{\sqrt{2\pi b^2}}e^{-|x|^2/2b^2} \tag{5}$$

is very natural. In fact, under some consistency conditions (symmetry, semi-group property, non-creation of local extrema), the Gaussian kernel is the only choice that gives a consistent scale-space theory, cf. [4, 8, 9, 11]. 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:

$$W_b = S_b(W) = G_b * W \ . \tag{6}$$

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

Scale-space theory in the discrete time case has been investigated in [9]. It turns out that just by sampling a continuous scale-space kernel, one obtains a discrete scale-space kernel. However, in doing so one does not obtain a scale-space theory with all the nice features of the continuous scale-space theory. There are difficulties with fine scales. In particular it is difficult to define higher order derivatives at fine scale levels. For the same reason it is difficult to define local extremum and zero crossings for fine scales. The semigroup property is lost. These questions are discussed in [9].

Interpolation and smoothing

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

$$W(s) = (I_F w)(s) = \sum_i F(s-i)w(i) \ . \tag{7}$$

We propose to use ideal low-pass interpolation

$$I = I_{\text{sinc}}$$

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

$$W \underset{D}{\overset{I}{\leftrightarrows}} w \ . \tag{8}$$

Note that if the camera induced blur cancels the high frequency components in W as in (2), the deterministic restoration $W_0 = I(w_0)$ is equal to W.

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

1. If the primary interest is the interpolated continuous signal, then *restore* the scale-space smoothed continuous signal W_b from the discrete signal w_0 first using ideal interpolation and then continuous scale-space smoothing.

2. If the primary interest is a discrete scale-space representation, then use the induced representation from the continuous scale-space, as defined in (8).

The procedure is illustrated by the diagram:

$$\begin{array}{ccc} W_0 & \xleftarrow{I} & w_0 \\ s_b \big\downarrow & & \big\downarrow s_b \\ W_b & \xrightarrow{D} & w_b \end{array} \tag{9}$$

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

$$w_b = s_b(w_0) \stackrel{\text{def}}{=} D(S_b(I(w_0))) \ , \tag{10}$$

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

$$g_b = D(G_b * \text{sinc}) \ . \tag{11}$$

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

$$||\,\text{sinc} * G_b - G_b||_2^2 \leq \frac{1}{b\sqrt{\pi}}\Phi(-\pi b\sqrt{2}) \ , \tag{12}$$

where Φ is the normal cumulative distribution function. Notice that the right hand side is small when b is large. The sampled Gaussian approach is also equivalent to using interpolation with the delta distribution followed by Gaussian smoothing. 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. Thererfore, there is no difficulty in defining higher order derivatives.

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

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

i.e. $s_b(w) = g_b * w$. In practice this scale-space theory is difficult to use for small scale parameters, because of the large tail of the sinc function. However, the function $\text{sinc} * G_b$ has a very small tail for larger scales. In practise one may use the approximation $\text{sinc} * G_b \approx G_b$ for large scales, according to (12). This simplifies implementation substantially.

4 The random field model

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

$$\begin{array}{ccc} W_0 + E_0 & \xleftarrow{I} & w_0 + e_0 \\ S_b \big\downarrow & & \big\downarrow s_b \\ W_b + E_b & \xrightarrow{D} & w_b + e_b \end{array} \tag{13}$$

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 the difference $\tau = t_1 - t_2$. ■

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*

$$R_X(f) = (\mathcal{F} r_X)(f) = \int r_X(\tau) e^{-i2\pi f \cdot \tau} d\tau \tag{14}$$

is also important. Again the same definition can be used for random fields with discrete parameters $s \in \mathbb{Z}^n$, but whereas the spectral density for random fields with continuous parameters is defined for all frequencies f, the spectral density of discrete random fields is only defined on an interval $f \in [-1/2, 1/2]^n$. Introductions to the theory of random processes and random fields are given in [1, 5, 6]. In these books you will find that convolution, discretisation and derivation preserves stationarity. The effect of these operations on the covariance function is also known:

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

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

Lemma 4.1.

$$\sum_i \operatorname{sinc}(s-i) \operatorname{sinc}(t-i) = \operatorname{sinc}(s-t) \quad . \tag{15}$$

Proof. See [2] ■

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

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

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

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

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

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

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

Proof. The proof that E is well defined is omitted due to lack of space, see [2]

It then follows that $m_E(s) = \mathcal{E}[\sum_i \operatorname{sinc}(s-i) e(i)] = \sum_i \operatorname{sinc}(s-i) \mathcal{E}[e(i)] = 0$. 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

$$\begin{aligned} r_E(s,t) = \mathcal{E}[E(s)E(t)] = \\ &= \sum_{i,j} \operatorname{sinc}(s-i) \operatorname{sinc}(t-j) r_e(i-j) = \\ &= \sum_{\substack{i \\ k=i-j}} \operatorname{sinc}(s-i) \operatorname{sinc}(t+k-i) r_e(k) = \\ &= \sum_k r_e(k) \operatorname{sinc}(s-t-k) = I(r_e)(s-t) \quad , \end{aligned} \tag{18}$$

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

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

Thus, all operations in the commutative diagram (13) preserve stationarity. This simplifies the modelling of errors in scale-space theory.

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

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

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

$$r_{E_b} = \varepsilon^2 \operatorname{sinc} * G_{b\sqrt{2}} \quad . \tag{19}$$

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

5 Sub-pixel correlation

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

A substantial improvement is obtained by using scale-space restoration of continuous images. This makes it possible to correlate regions in two images with sub-pixel translations with much higher precision than obtained by ordinary methods. Furthermore, a proper modelling of the residual field makes it possible to analyse the stochastic properties of the localisation error. The idea is that, at least locally,

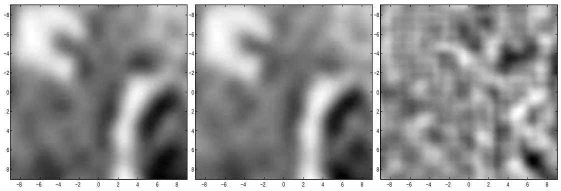

Figure 2. Regions in two images are correlated with sub-pixel translations using least squares of the residuals of the restored continuous scale-space representation at scale $b = 0.9$.

the images only differ by an unknown translation ρ. 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 $\rho_0 = (\rho_1, \rho_2)$, we thus have

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

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

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

The result of such a minimisation is shown in Figure 2.

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

The quality of the estimated sub-pixel translation,

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

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

$$C = \mathcal{C}[X] \approx A^{-1} B A^{-T} \ , \tag{20}$$

$$A = 2 \int_{t_1 \in \Omega} (V\bar{W}\bar{W}^T)(t_1) dt_1 \ , \tag{21}$$

$$B = \int_{t_1 \in \Omega} ((V\bar{W}) * r_{E - \bar{E}})(t_1)\, (V\bar{W})(t_1) dt_1 \ . \tag{22}$$

6 Conclusions

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

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

Furthermore, the stochastic errors in images are modelled and new results are given that show how these errors influence the continuous and discrete scale-space representations and their derivatives. This information is crucial in understanding the stochastic behaviour of scale-space representations as well as fundamental properties of feature detectors.

Acknowledgements

The authors thank Georg Lindgren for inspiration and valuable discussions.

References

[1] A. Adler. *The Geometry of Random Fields*. Wiley, New York, 1985.

[2] K. Åström. *Invariancy Methods for Points, Curves and Surfaces in Computational Vision*. PhD thesis, Dept of Mathematics, Lund University, Sweden, 1996.

[3] K. Åström and A. Heyden. Stochastic analysis of sub-pixel edge detection. In *Proc. International Conference on Pattern Recognition, Vienna, Austria*, 1996.

[4] 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.

[5] H. Cramér and M. R. Leadbetter. *Stationary and Related Stochastic Processes*. Wiley, New York, 1967.

[6] N. A. C. Cressie. *Statistics for Spatial Data*. Wiley, New York, 1991.

[7] E. Hecht. *Optics*. Addison-Wesley, Reading, Mass., 1987.

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

[9] T. Lindeberg. *Scale-Space Theory in Computer Vision*. Kluwer Academic Publishers, 1994.

[10] B. Pratt. *Digital Image Processing*. Wiley-Interscience, 1978.

[11] A. P. Witkin. Scale-space filtering. In *Proc. Eigth International Joint Conference on Artificial Intelligence, Karlsruhe, West Germany*, pages 1019–1022, 1983.