# On The Role of The Gaussian Kernel in Edge Detection and Scale-Space Methods

Ji-Young Lim

Academic advisor:
Prof. Dr.-Ing. H. Siegfried Stiehl

Universität Hamburg
Fachbereich Informatik
Arbeitsbereich Kognitive Systeme

**April 2001**

# Abstract

Edges in a digital image contain important information. At the lowest level of visual processing, multi-scale techniques based on the linear scale-space theory can be applied to edge detection for the purpose of generating a rich representation of image structure. We look into the fundamentals of edge detection as well as the linear scale-space theory and review some related approaches. In edge detection as well as in the scale-space theory, the Gaussian kernel plays a key role. In order to analyze theoretically the problems caused by discretization of the continuous Gaussian kernel, we investigate the difference between the continuous Gaussian and the sampled Gaussian which is commonly used in practice. The result of this investigation shows that the sampled Gaussian with a small scale is not appropriate for approximating the continuous Gaussian.



# Zusammenfassung

Kanten in einem digitalen Bild tragen wichtige Information. Bei der Extraktion solcher Bildstrukturen kommt insbesondere Multi-Skalen-Methoden eine zentrale Bedeutung zu. In diesem Bericht werden die Grundlagen der Kantendetektion, der linearen Skalenraum-Theorie sowie verwandter Ansätze aufgearbeitet und bewertet. Dabei wird herausgearbeitet, daß das Gaußfilter eine sehr wichtige Rolle sowohl bei der Kantendetektion als auch in der Skalenraum-Theorie spielt. Um die Problematik der Diskretisierung des Gaußfilters theoretisch zu analysieren, wird der Unterschied zwischen dem kontinuierlichen Gaußfilter und dem abgetasteten Gaußfilter, das normalerwiese in der Praxis angewendet wird, untersucht. Die Ergebnisse dieser Untersuchung zeigen, daß das abgetastete Gaußfilter bei kleinen Skalen nicht geeignet ist, das kontinuierliche Gaußfilter anzunähern.

# Contents

# 1 Introduction

Important physical properties of a scene taken by a CCD camera are provided by changes of intensities, which are generally called *edges* ([9], [12], [23], [31], [39]). Edges in the image are locations where the intensity significantly changes from one level to a different one ([23], [24]), i.e., transitions from a bright region to a dark region, or vice versa ([5]). Edges carry useful information about object boundaries, which can be used for image analysis or object recognition. Edge detection is to detect, localize, and measure local intensity discontinuities in terms of scales, contrast, orientation, and type. It denotes the process of identifying physical boundaries of three-dimensional objects represented as intensity changes in images. The correct recognition of objects in high-level computer vision also depends on the reliability of the applied edge detection method.

In spite of the plethora of publications on computational edge detection due to its importance in early visual processing, no precise and widely accepted mathematical definition of edge is available both in computer and biological vision systems ([23], [24]). One problem of edge detection is that the extracted edges do not necessarily correspond to boundaries of objects. With the exception of high quality images resulting from controlled image generation, edge detection typically results in spurious (noise-induced) edges and gaps, which may be caused by several reasons : First, even a very simple scene may contain a good deal of irrelevant detail, which may give rise to intensity changes in the image. Besides, images are corrupted by noise and discretization artifacts. Moreover, it is also very difficult to interpret effectively the characteristics of the scene given edges alone since the intensity changes stem from various physical phenomena, e.g., local differences in illumination, surface reflection, surface orientation, shadow, or texture.

The earliest edge detection schemes are based upon approximations of the intensity gradient ([3], [19]). Approximations of the derivative were computed using the Robert's cross operator, the Sobel operator, the Kirsch operator, or the Prewitt operator. However, those operators have the disadvantage that they assumed an underlying step edge and used a fixed operator size (or, in mathematical terms, support). Given a biological motivation (see e.g. [23], [24]), Marr and Hildreth [31] addressed the rotationally symmetric Laplacian operator and the special property of the Gaussian kernel as being the only real-valued func-

tion that minimizes the product of the variances of the filter in the spatial and frequency domain. Canny [13] considered the problem of determining an optimal smoothing filter of finite support for detecting step edges. His approach behind the definition of such a filter was to maximize a certain performance measure constituting a trade-off between detection and localization properties subject to avoiding multiple responses for a single edge.

An important step forward was the introduction of multi-scale approaches. From the viewpoint of multi-scale approaches related to edge detection, an object gives rise to different types of edge structure on different scales depending on the physical nature of the edge; only some of these stand out locally and seem to be more significant than others. Therefore, a natural requirement is to measure the locally stable scale for each edge structure. This is the main concept of multi-scale approaches to edge detection, which have been theoretically grounded by the scale-space theory. The scale-space theory, as a relatively new field, has established a well-founded and general multi-scale technique for image structure analysis, e.g., for 2-D, 3-D, and time series images. Given an image, its scale-space representation generated by convolution with the Gaussian kernel is particularly useful for analyzing image data at the lower levels in the chain of information processing of a visual system ([29], [30]). One scientifically attractive aspect of the linear scale-space theory, proposed first by Witkin [40] for a 1-D continuous signal and furthered by Koenderink [25] for image structure in general, is its solid mathematical framework (see [1], [25], or [40]). In particular, it is theoretically proved ([1], [41]) that the Gaussian kernel is the unique kernel to generate the scale-space for continuous signals. The essence of the results from the scale-space theory is, if one assumes that the first stages of signal processing should be as uncommitted as possible and should not have any *a priori* knowledge about the world from which the signal stems, that the convolution of the initial signal with the Gaussian kernel and its $n$th-order derivatives of different scale is singled out as a canonical class of computational low-level processes (see e.g., [25]).

In edge detection as well as in the scale-space theory, the Gaussian kernel plays a key role. However, when we deal with the Gaussian kernel in practice, we confront a problem: Whereas the Gaussian kernel is theoretically assumed to be continuous in the infinite spatial domain, we have to cope with discrete signals[1]. The problem lies in the

[1]Note that a few authors referred to "discrete signals" although, strictly speaking, "digital signals"

accuracy and validity of a sampled Gaussian which is commonly used for approximating the continuous Gaussian kernel defined in the infinite spatial domain.

In this report, after surveying edge detection methods and related multi-scale approaches based on the linear scale-space theory, we investigate how mathematically different a sampled Gaussian kernel is from the continuous Gaussian kernel. This report is organized as follows : In Section 2 we review fundamentals of edge detection such as the Gaussian kernel, differentiation, filtering, and edge models. Then we survey some milestone approaches to edge detection. In Section 3, after giving a brief address of the general concepts of the scale-space theory and its properties, we look over some of the key multi-scale approaches to edge detection. Then, we analyze the sampled Gaussian kernel in detail, from which we investigate the numerical difference between the continuous Gaussian and the sampled Gaussian in the case of a sampling period of one in 1-D as well as in 2-D and 3-D in Section 4. Finally, we summarize and present future work in Section 5.

would be the correct terminus technicus. For reason of convenience we here use the terminus, "discrete signals". Also, we here refer to "signals" as "image (or visual) signals".

# 2 Survey of Edge Detection

In this section, we review the fundamentals which are necessary to comprehend the basics of the method of edge detection. Then we survey some milestone edge detection approaches.

## 2.1 Fundamentals

Edge detection is a means of generating compact descriptions from digital images such that most of the structural information in an image is preserved. In general, the process of edge detection essentially consists of three steps, i.e., *filtering*, *differentiation*, and *description* ([24], [39]). That is to say, an image intensity function should be smoothed by filtering in order to reduce noise, and the presence of edges is detected through the differentiation step. Then simple features resulting from the differentiation step, e.g., peaks (positive and negative extrema) or zero-crossings, are used to describe edges. A Gaussian kernel essentially plays a very important role in the whole process of edge detection. In Section 2.1.2 and Section 2.1.3, we will explain in detail filtering and differentiation, respectively. In order to approach these tasks analytically, the early edge operators generally assume a specific edge model; they are designed to locate step edges which can be mathematically modeled by the Heaviside function for the 1-D case. Step intensity changes are important because they usually correspond to sharp contrasts. Based on the step edge model, other edge models can be analytically established, which will be addressed in Section 2.1.4.

### 2.1.1 The Gaussian Kernel

The Gaussian kernel has its origin in the normal distribution of random variables (measured values). The 1-D Gaussian probability density function is given as

$$G(x;\sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}},$$

where $\sigma$ is the standard deviation of $G$ and $\mu$ is the mean value as the most probable value. This probability density of the normal distribution taken as a convolution kernel (see below) is referred to as the so-called Gaussian filter, Gaussian kernel, or simply Gaussian. In terms of filtering (see Section 2.1.2) the Gaussian is applied to a signal $f(x)$ through convolution

given by

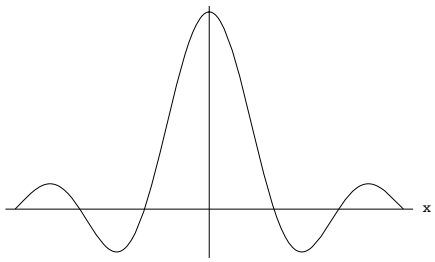
$$G(x;\sigma) * f(x) = \int_{-\infty}^{\infty} G(\xi;\sigma) f(x-\xi) d\xi.$$

A remarkable property of the Gaussian kernel is that the Gaussian is the only real-valued kernel which gives equality both in the spatial domain and in the frequency domain ([8], [10], [23], [24], [31]): The Fourier transform of the Gaussian is also a Gaussian. Let us consider the Fourier transforms of the *sinc* function and the $sinc^2$ function, and compare them with the Fourier transformed Gaussian function :

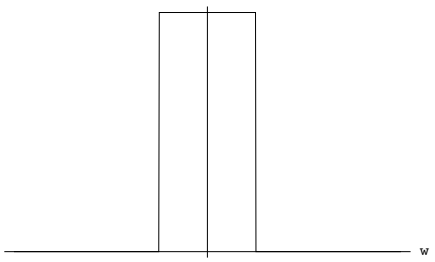
$$\begin{aligned} &\text{(a)} \qquad sinc(x) \circ\!\!-\!\!\bullet\, \Pi(\omega) \\ &\text{(b)} \qquad sinc^2(x) \circ\!\!-\!\!\bullet\, \Lambda(\omega) \\ &\text{(c)} \qquad e^{-\pi x^2} \circ\!\!-\!\!\bullet\, e^{-\pi \omega^2} \end{aligned}$$

The Fourier transform of the *sinc* function is the unit rectangle function (a) and the Fourier transformed $sinc^2$ function is the triangle function of unit height and area (b) (see e.g. [10], [15] for details), while the Fourier transform of the Gaussian function is a Gaussian again (c). These three transform pairs are illustrated in Fig. 1. In contrast to the other two functions, the response of the Gaussian kernel is as narrow as possible in both the spatial and the frequency domain, which minimizes the effects of aliasing introduced by a band-limited filter ([8], [10]). It is noticeable that the bandwidth-duration product of a signal cannot be less than a certain minimum value, which is the so-called uncertainty relation. It is well-known that the Gaussian kernel minimizes the uncertainty (see page 12).

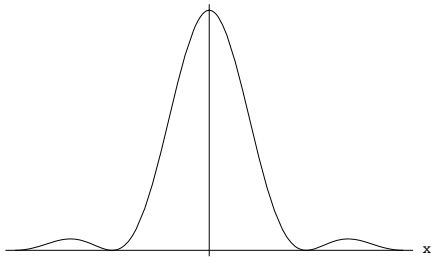
The higher dimensional Gaussian is the rotationally symmetric kernel that is separable in Cartesian coordinates. Moreover, the higher dimensional Gaussian kernel is also separable in convolution, which can be easily shown using the Fourier transform theorem ([10,



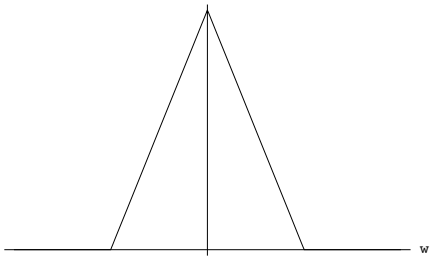

(a) $sinc(x)$



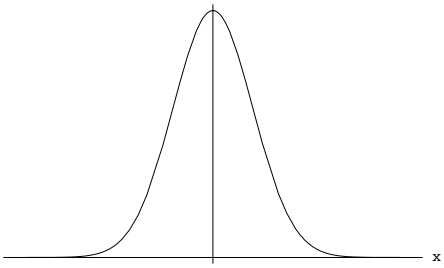

(b) $\Pi(\omega)$



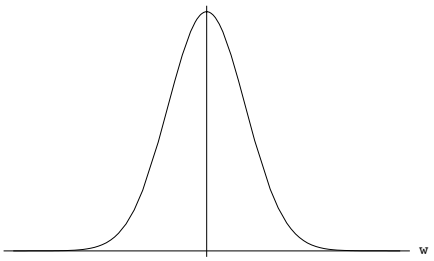

(c) $sinc^2(x)$




(d) $\Lambda(\omega)$




(e) $e^{-\pi x^2}$




(f) $e^{-\pi \omega^2}$

Figure 1: Some Fourier transforms : The Gaussian achieves equality both in the spatial domain and in the frequency domain.

pp. 245]). For example, in the case of the 2-D Gaussian kernel,

$$\begin{aligned}
G(x,y) * f(x,y) &= (G(x) \cdot G(y)) * f(x,y) \\
&= \mathcal{F}^{-1}\left\{ \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} G(x)G(y)e^{-j(ux+vy)}dxdy \right\} * f(x,y) \\
&= \mathcal{F}^{-1}\left\{ \int_{-\infty}^{+\infty} G(x)e^{-jux}dx \cdot \int_{-\infty}^{+\infty} G(y)e^{-jvy}dy \right\} * f(x,y) \\
&= \mathcal{F}^{-1}\left\{ G_{\mathcal{F}}(u) \cdot G_{\mathcal{F}}(v) \right\} * f(x,y) \\
&= (G(x) * G(y)) * f(x,y) \\
&= G(x) * (G(y) * f(x,y)) \\
&= G(y) * (G(x) * f(x,y)) \,.
\end{aligned}$$

### 2.1.2 Linear Filtering

**Convolution** Linear filtering is shift-invariant. Therefore, it is realized by a convolution. For two functions $f(x), g(x) : \mathbb{R} \to \mathbb{R}$, the convolution[2] denoted as $f(x) * g(x)$ is defined by the convolution integral as

$$f(x) * g(x) = (f * g)(x) = \int_{-\infty}^{+\infty} f(\xi)g(x-\xi)d\xi.$$

The following properties of the convolution hold for all $f, g, h : \mathbb{R} \to \mathbb{R}$ :

- $f * g = g * f$ (commutativity)
- $(f * g) * h = f * (g * h)$ (associativity)
- $f * (g + h) = (f * g) + (f * h)$ (distributivity)
- $(c \cdot f) * g = c \cdot (f * g)$ ($c$ is a constant)
- $\partial(f * g)/\partial x = \partial f/\partial x * g = f * \partial g/\partial x$

**Reducing Noise** In machine vision as well as in most numerical problems, data are noisy. Noise in the photon transduction process is ultimately unavoidable ([3], [19]). Unfortunately it is well known that differentiation is not robust against noise ([12], [39]).

[2]Sometimes $\otimes$ is also used for the notation of the convolution operator.

Hadamard first introduced the definition of ill-posedness in the field of partial differential equations. A problem is well-posed if i) its solution exists, ii) the solution is unique, and iii) it continuously depends on the initial data. Ill-posed problems fail to satisfy one or more of these criteria. The main idea behind solving ill-posed problems is to restrict the class of admissible solutions by introducing suitable *a priori* knowledge. This can be exploited through the formulation of the problem via variation calculus which imposes constraints on the possible solutions or as statistical properties of the solution space. In particular, the general term *regularization* is used for any method to make an ill-posed problem well-posed (see e.g. [7], [34], [35], or [39]).

The aim of filtering is to reduce noise, and it can be understood as regularization. Edge detection in case of noisy data realized by numerical differentiation is mildly ill-posed since the solution does not depend continuously on the noisy input data ([39]). For example, the Laplacian operator is sensitive to noise and thus, strictly speaking, not suitable for applying it to edge detection as it is. Alternatively the Laplacian of Gaussian (LoG) may be used to alleviate the problems caused by noise (see below).

Even a small amount of noise may perturb differentiation. Let us consider a function $\hat{f}(x)$ given as

$$\hat{f}(x) = f(x) + \epsilon \sin \omega x,$$

where the sin-term is meant to be simulating additive high-frequency noise. $f(x)$ may be close to $\hat{f}(x)$ according to standard norms $(L^2, L^\infty, \cdots)$, provided $\epsilon$ is sufficiently small. On the other hand, $df/dx$ may be much different from $d\hat{f}/dx$ if $\omega$ is large (viz, high-frequency noise),

$$\frac{d\hat{f}}{dx} = \frac{df}{dx} + \omega\epsilon \cos \omega x,$$

and it even becomes exponentially worse as the order of differentiation increases; the second-order derivative is derived as

$$\frac{d^2\hat{f}}{dx^2} = \frac{d^2 f}{dx^2} - \omega^2 \epsilon \sin \omega x,$$

and generally the $n$th-order derivative is

$$\frac{d^n\hat{f}}{dx^n} = \frac{d^n f}{dx^n} + \omega^n \epsilon \sin\left(\omega x + \frac{n\pi}{2}\right).$$

This noise-induced effect can be reduced by using a lowpass filter, e.g., a Gaussian filter. Let us consider a 1-D function $f(x)$ formed by the Heaviside function $\mathcal{H}(x)$ with additive high-frequency noise simulated by a sin-term such that

$$f(x) = \mathcal{H}(x) + \sin \omega_0 x.$$

As a lowpass filter, the Gaussian is applied through convolution

$$G(x;\sigma) * f(x) = G(x;\sigma) * (\mathcal{H}(x) + \sin \omega_0 x)\,.$$

Then, the first term is derived as

$$\begin{aligned} G(x;\sigma) * \mathcal{H}(x) &= \int_{-\infty}^{\infty} G(\xi;\sigma)\mathcal{H}(x-\xi)d\xi \\ &= \int_{-\infty}^{x} G(\xi;\sigma)d\xi \\ &= \Phi(x;\sigma), \end{aligned}$$

where $\Phi$ is the error integral function, and the second term is derived as

$$\begin{aligned} G(x;\sigma) * \sin \omega_0 x &= \mathcal{F}^{-1}\left\{\mathcal{F}\{G(x;\sigma)\} \cdot \mathcal{F}\{\sin \omega_0 x\}\right\} \\ &= \mathcal{F}^{-1}\left\{e^{-\frac{\omega^2\sigma^2}{2}} \cdot \frac{\pi}{j}\left(\delta(\omega-\omega_0) - \delta(\omega+\omega_0)\right)\right\} \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-\frac{\omega^2\sigma^2}{2}} \cdot \frac{\pi}{j}\left(\delta(\omega-\omega_0) - \delta(\omega+\omega_0)\right) \cdot e^{j\omega x} d\omega \\ &= e^{-\frac{\omega_0^2\sigma^2}{2}} \cdot \frac{j}{2}\left(e^{j\omega_0 x} - e^{-j\omega_0 x}\right) \\ &= e^{-\frac{\omega_0^2\sigma^2}{2}} \sin \omega_0 x. \end{aligned}$$

The noise convolved with the Gaussian filter exponentially decreases, whereas the Heaviside function is blurred to the error function (see Section 2.1.4 for the $\Phi$-function).

**Lowpass Filtering** Linear lowpass filters are usually used in digital image processing to suppress noise. So-called *lowpass* filters attenuate or, in the ideal case, even eliminate high-frequency components in the frequency domain while leaving low frequencies untouched, that is, only low frequencies "pass" the filter. Since high-frequency components characterize edges, noise, and other sharp contrast details in an image, the net effect of lowpass filtering is image blurring ([19]).

Torre and Poggio [39] addressed a few conditions the Fourier transform of a lowpass filter, $\mathcal{F}(\omega;\sigma)$ ($\sigma$ is the scale parameter), should satisfy :

1. $\mathcal{F}(\omega;\sigma)$ is bounded for $\sigma \geq 0$ and all $\omega$.
2. $\mathcal{F}(\omega;\sigma)$ is an even function with respect to $\omega$, and it belongs to $L_2(-\infty,+\infty)$.
3. $\mathcal{F}(\omega;\sigma)j\omega$ belongs to $L_2(-\infty,+\infty)$.
4. For every $\sigma > 0$, it holds $\lim_{|\omega|\to\infty} \mathcal{F}(\omega;\sigma) = 0$.
5. $\mathcal{F}(\omega;\sigma) \to 1$ as $\sigma \to 0$ and $\mathcal{F}(\omega;0) = 1$.

Moreover they classified the three kinds of lowpass filters used in edge detection: A band-limited filter for the frequency domain, a support-limited filter for the spatial domain, and a minimal uncertainty filter.

If a filter is band-limited, its Fourier transform vanishes off out of a band limit. The band-limited filter is an obvious choice for regularizing differentiation, since the simplest way to avoid harmful noise is to filter out high frequencies that are amplified by the differentiation effect as shown above. However, the band-limitedness in the frequency domain causes the infinite support in the spatial domain, which is unrealistic in practice (see Fig. 1, top row). Shanmugam et al. [37] used the band-limited filter for edge detection, which will be reviewed in Section 2.2.1.

In practice, filters have a finite extension and are, therefore, support-limited in spatial domain (compact support). Computational efficiency requires that the support of a filter is as compact as possible. As an example shown in Torre and Poggio [39], let us consider an ideal 1-D lowpass filter in the spatial domain which is the simplest even filter with a strictly limited support $I$ and unitary energy as

$$f(x) = \begin{cases} \frac{1}{\sqrt{2I}} & |x| \leq I \\ 0 & |x| > I \end{cases}.$$

Then, its Fourier transform $\mathcal{F}(\omega)$ is given by

$$\mathcal{F}(\omega) = \sqrt{\frac{2}{I}}\frac{\sin(I\omega)}{\omega}.$$

In this case, the filter fails to satisfy the third condition required for being a lowpass filter. The third condition means that the Fourier transform of the first derivative of a lowpass filter should belong to $L_2(-\infty, +\infty)$. In other words, the condition

$$\left| \int_{-\infty}^{\infty} (\mathcal{F}(\omega) j\omega)^2 \, d\omega \right| < \infty$$

must hold. The violation of this requirement can be simply shown :

$$\begin{aligned}
\left| \int_{-\infty}^{\infty} \left( \sqrt{\frac{2}{I}} \frac{\sin(I\omega)}{\omega} j\omega \right)^2 d\omega \right| &= \left| -\frac{2}{I} \int_{-\infty}^{\infty} \sin^2(I\omega) d\omega \right| \\
&= \frac{2}{I} \left[ \frac{\omega}{2} - \frac{1}{4I} \sin(2I\omega) \right]_{-\infty}^{\infty} \\
&= \lim_{\omega \to \infty} 2 \left[ \frac{\omega}{I} + \frac{1}{2I^2} \sin(2I\omega) \right] \\
&= \infty.
\end{aligned}$$

This is the reason why the class of support-limited filter introduces back in its differentiation high frequencies in the same amount as they are removed by this type of filtering ([19], [39]).

A band-limited filter has theoretically infinite spatial support since it is band-limited in the frequency domain. The drawback of support-limited filters is that they are regularizing only in a weak sense. Accordingly, it is natural to try to find an optimal compromise between these two types of filters, since the one in the spatial domain and the other in the frequency domain are conflicting. A measure of the spread of a function $f \in L_2(\mathbb{R})$ in the spatial domain and in the frequency domain is related by the uncertainty principle ([10], [39]), denoted as $\Delta U$, which is defined as $\Delta U = X\Omega$ :

$$\begin{aligned}
X^2 &= \frac{\int_{-\infty}^{+\infty} (x - \bar{x})^2 f^2(x) dx}{\int_{-\infty}^{+\infty} f^2(x) dx}, \\
\bar{x} &= \int_{-\infty}^{+\infty} x f^2(x) dx, \\
\Omega^2 &= \frac{\int_{-\infty}^{+\infty} (\omega - \bar{\omega})^2 \mid F(\omega) \mid^2 d\omega}{\int_{-\infty}^{+\infty} \mid F(\omega) \mid^2 d\omega}, \qquad \text{and} \\
\bar{\omega} &= \int_{-\infty}^{+\infty} \omega \mid F(\omega) \mid^2 d\omega,
\end{aligned}$$

where $X^2$ is the variance of $f^2(x)$, $\bar{x}$ is the mean value of $x$ being weighted according to the distribution of $f^2(x)$, and $\Omega^2$ and $\bar{\omega}$ are corresponding to those of the Fourier transform, respectively. The uncertainty principle states that $\Delta U = X\Omega \geq \frac{\pi}{4}$ (see [10, pp. 160-163] for its proof). The Gaussian kernel is the only real-valued function which gives the minimum uncertainty, i.e., $\frac{\pi}{4}$. The Gaussian also satisfies all five conditions required for a lowpass filter. On this ground Marr and Hildreth [31] have proposed the Gaussian as the optimal filter.

### 2.1.3 Differentiation

Most edge detection methods are based on a local derivative operator. The magnitude of the gradient of an image can be used to detect the presence of edges. Alternatively the sign change of the Laplacian can be used to determine where the edge is, since the second-order derivative of intensity change in images has a zero-crossing at the mid point of a gray-level transition.

**Gradient** In the case of a 2-D intensity function $f(x, y)$, the gradient of $f$ at a given point $(x, y)$ is defined as

$$\nabla f = \begin{pmatrix} f_x \\ f_y \end{pmatrix} = \begin{pmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{pmatrix},$$

where $\nabla$ is the nabla operator. The gradient vector points to the direction of the maximum rate of change of $f$ at $(x, y)$. The gradient magnitude (or, respectively, gradient vector norm) given by

$$\mid \nabla f \mid = \sqrt{f_x^2 + f_y^2}$$

is used as measure of the intensity change.

**Laplacian** The gradient (or, synonymously, nabla) operator employed twice to a scalar valued function $f$ in $\mathbb{R}^2$ or $\mathbb{R}^3$ results in the Laplacian operator[3] expressed by $\nabla^2 = \nabla^T \cdot \nabla$,

[3] It is sometimes denoted by $\triangle$.

and the Laplacian of $f(x, y)$ is defined by

$$\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}.$$

The Laplacian is a non-directional operator and has been extensively used due to its computational convenience. Being used for detecting edges, however, it has disadvantages ([3]) : First, the useful information of the edge direction is lost. Second, being an approximation to the second derivative, the Laplacian is more sensitive to noise than the first-order derivative operator. And third, the locus of the zero-crossing strongly depends on e.g. the local curvature of the object boundary to be detected (see e.g. [32]). For these reasons, the Laplacian usually plays only a secondary role in edge detection. A more general use of the Laplacian is in finding the location of edges using zero-crossings of the Laplacian of Gaussian (LoG) (see e.g. [19], [31] for details).

**Directional derivatives** The first-order directional derivative is the scalar product of the gradient of a function and its unit vector. It provides the derivative of the function with respect to the direction of the unit vector. Given a 2-D intensity function $f(x, y)$, its unit vector $n$ is given as

$$n = \frac{\nabla f}{|\nabla f|} = \left( \frac{f_x}{\sqrt{f_x^2 + f_y^2}} \quad \frac{f_y}{\sqrt{f_x^2 + f_y^2}} \right)^T,$$

and the first- and the second-order directional derivatives are derived as

$$\begin{aligned} \frac{\partial f}{\partial n} &= (\nabla f)^T \cdot n \\ &= \frac{f_x^2 + f_y^2}{\sqrt{f_x^2 + f_y^2}} = |\nabla f|, \\ \frac{\partial^2 f}{\partial n^2} &= \left\{ \nabla ((\nabla f)^T \cdot n) \right\}^T \cdot n \\ &= \frac{{f_x}^2 f_{xx} + 2 f_x f_y f_{xy} + {f_y}^2 f_{yy}}{{f_x}^2 + {f_y}^2}. \end{aligned}$$

Zero-crossings of the second-order directional derivative along the gradient of the intensity function may be considered as edges in the image. However, the second directional derivative can not be defined when $|\nabla f| = 0$.

If $f(x,y)$ can be represented as a function of only one variable, i.e., $f(x,y_0)$ where $y_0$ is a constant, the two operators $\nabla^2$ and $\frac{\partial^2}{\partial n^2}$ are equivalent. Also if $f_{yy} = f_{xy} = 0$ when $\frac{\partial^2 f}{\partial n^2} = 0$, the zeros of $\frac{\partial^2 f}{\partial n^2}$ coincide with those of $\nabla^2 f$ ([39]). However, for a circularly symmetric intensity function, the zeros of $\nabla^2$ are farther apart from edges than those of $\frac{\partial^2}{\partial n^2}$ (see [32]). This lack of accurate localization of $\nabla^2$ for circularly symmetric patterns can also be observed in the case of corners where zeros of $\nabla^2$ (not those of $\frac{\partial^2}{\partial n^2}$) swing wide of corners ([8], [36]).

### 2.1.4 Edge Models

If we intend to approach the detection of edge analytically, it is necessary to establish an edge model. That is, in order to assure accurate and reliable results from an edge detection scheme, an either implicitly assumed or explicitly formulated edge model are required ([39]). An edge model is usually designed to locate a step edge which can be mathematically modeled by the Heaviside function, and other simple edge models can be established based on the step edge model (see e.g., [2], [4], [8], [13], [18], [32], [33], [36], [37]).

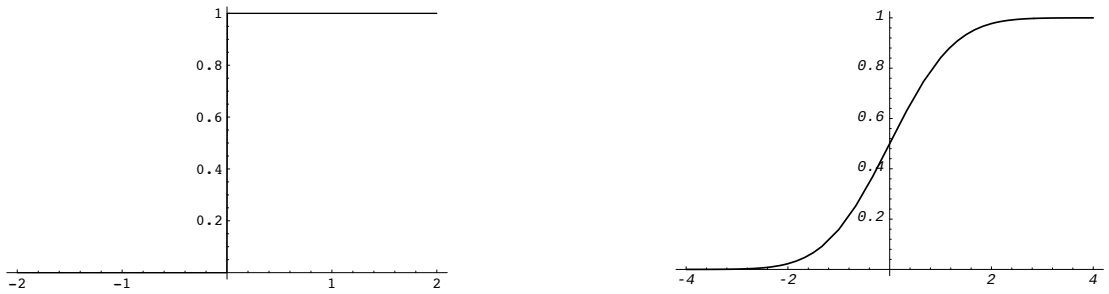
Here, we roughly classify edge models as a step edge, a double edge, a corner edge, and a blob edge model.

**Step Edge** The Heaviside function (so-called ideal step function), denoted by $\mathcal{H}(x)$ which jumps from 0 to 1 at $x = 0$, is defined as

$$\mathcal{H}(x) = \begin{cases} 0, & x < 0 \\ (\frac{1}{2}, & x = 0) \\ 1, & x > 0 \end{cases} .$$

The Heaviside function represents ideal step changes, but real edges show rather smooth intensity changes, which hence look sigmoid. A sigmoid step edge can be modeled by the normalized error integral curve given by

$$\Phi(x;\sigma) = G(x;\sigma) * \mathcal{H}(x),$$




(a) $\mathcal{H}(x)$ (b) $\Phi(x;\sigma)$

Figure 2: Step edge models

where $G(x;\sigma)$ denotes the 1-D Gaussian (the normal distribution with zero mean) and $\sigma$ is its standard deviation (i.e., the degree of blurring). Fig. 2 illustrates the Heaviside function and the $\Phi$-function (see, e.g., [2], [4], [6], [8], [12], [31], [42]).

The so-called $\Phi$-function is known for the primitive function of the Gaussian and is some kind of "basic function" in connection with blurring of the images [4]. Note that $\lim_{\sigma\to 0} G(x;\sigma) = \delta(x)$, and thus

$$\lim_{\sigma\to 0} \Phi(x;\sigma) = \delta(x) * \mathcal{H}(x) = \mathcal{H}(x)$$

holds. Explicit descriptions of blurring are relevant for explaining how locally non-curved gray level edge profiles change in scale-space (see Section 3).

**Other Edge Models** Bars, corners, or blobs should be distinguished from step edges, since they consist of more than one edge and are affected by one another as the scale parameter increases. A step edge model alone cannot significantly represent accurate information about positions or contrasts of bars, corners, or blobs.

As a matter of fact, each edge model can be formed in various ways. Here we consider only a few simple cases that are built upon the $\Phi$-function.

- Bar Edge (Double Edge)

Two parallel step edges lying only a small distance apart in the image are called bars or double edges ([31]). The unit 1-D bar edge undergoing Gaussian blurring can be defined by

$$\begin{aligned} f_{double} &= G(x;\sigma) * ((\mathcal{H}(x-a) - \mathcal{H}(x-b)) \\ &= \Phi(x-a;\sigma) - \Phi(x-b;\sigma), \end{aligned}$$

where the value of $|a-b|$ is the width between one edge and the other one (see e.g. [4], [31]).

- Corner (Junction) Edge

  A corner is formed by the meeting of at least two edges in the image. Among various types of corners (e.g., L-, T-, Y-, Arrow-corners), the simplest one of corner edges is a L corner. The unit L corner edge can be represented as

$$\begin{aligned} f_{corner} &= (G(x;\sigma) * \mathcal{H}(x-a)) \cdot (G(y;\sigma) * \mathcal{H}(y-b)) \\ &= \Phi(x-a;\sigma) \cdot \Phi(y-b;\sigma), \end{aligned}$$

  where $(a,b)$ represents the corner point (see e.g. [4], [8], [36]).

- Blob

  Closed contours formed by merging of two (or more) pairs of edges are called as blobs ([31]). An unit blob made of two pairs of parallel edges can be modeled as

$$\begin{aligned} f_{blob} &= (G(x;\sigma) * (\mathcal{H}(x-a) - \mathcal{H}(x-b))) \cdot (G(y;\sigma) * (\mathcal{H}(y-\alpha) - \mathcal{H}(y-\beta))) \\ &= (\Phi(x-a;\sigma) - \Phi(x-b;\sigma)) \cdot (\Phi(y-\alpha;\sigma) - \Phi(y-\beta;\sigma)), \end{aligned}$$

  where $(a,\alpha)$, $(a,\beta)$, $(b,\alpha)$ and $(b,\beta)$ are the corner points of the blob.

Each exemplified edge model shown above is the simplest case respectively. In each case, however, the amount of blurring with respect to the $\Phi$-function should be considered, and in higher dimensional cases the orientation of each edge model must be taken into account.

## 2.2 Related Work

### 2.2.1 Shanmugam, Dickey and Green

Shanmugam et al. [37] derived an optimal frequency domain filter for detecting intensity changes using the following criteria : The filter

1. yields maximum output in the vicinity of an edge for a given band width and resolution requirement in the image,
2. is a strictly band-limited filter (i.e., a filter whose Fourier transform has its support on an interval surrounding the origin),
3. yields a small output when the input is constant or slowly varying, and
4. is an even function in space.

Let $f(x)$ be an input function and $h(x)$ the impulse response of the filter whose transfer function we seek to derive. The filter output $g(x)$ and its Fourier transform are denoted as

$$g(x) = f(x) * h(x) \circ\!\!-\!\!\bullet\, F(\omega) \cdot H(\omega).$$

For given $I$ (spatial domain) and $\Omega$ (frequency domain), the authors derived the parameter $\gamma$ as

$$\gamma = \frac{\int_{-\frac{I}{2}}^{\frac{I}{2}} |g(\xi)|^2 d\xi}{\int_{-\infty}^{\infty} |g(\xi)|^2 d\xi}.$$

The maximum value of the measure of $\gamma$ represents the portion of the output signal energy within the resolution interval $I$ in the vicinity of the edge.

In this context, they derived the optimal filter transfer function $H(\omega)$ which maximizes $\gamma$ :

$$H(\omega) = \begin{cases} K\omega\psi_1\left(c, \frac{\omega I}{2\Omega}\right), & |\omega| \leq \Omega \\ 0, & |\omega| > \Omega \end{cases},$$

where $K$ is a real-valued constant, $H(\omega)$ is band-limited to $\Omega$, and $\psi_1(c, x)$ is a linear prolate spheroidal wave function of band-limited functions which possesses the property of being

orthogonal over both a given finite interval and $(-\infty, \infty)$. Through the approximation of $\psi_1(c, x)$, the transfer function of the optimum filter is expressed as

$$H_{opt}(\omega) = \begin{cases} \frac{(K\omega)\exp\left[-(cI^2\omega^2)/4\Omega^2\right]}{jF(\omega)}, & |\omega| \leq \Omega \\ 0, & \text{elsewhere} \end{cases}.$$

Assuming $f(x)$ to be a step edge represented by $\mathcal{H}(x)$,

$$f(x) \circ\!\!-\!\!\bullet\, F(\omega) = \frac{1}{2}\delta(\omega) - j\frac{1}{2\pi\omega}$$

holds. Then the transfer function of the optimal filter for a step edge is reduced to

$$H_{step}(\omega) = \begin{cases} (K_1\omega^2)\exp\left[-(cI^2\omega^2)/4\Omega^2\right], & |\omega| \leq \Omega \\ 0, & \text{elsewhere} \end{cases}.$$

According to the theorem of differentiation in the frequency domain given by

$$\frac{\partial f(x)}{\partial x} \circ\!\!-\!\!\bullet\, j\omega \cdot \mathcal{F}(\omega),$$

we can derive the Fourier transform which corresponds to the second-order derivative of the Gaussian :

$$\begin{aligned} \frac{\partial^2 G(x)}{\partial x^2} &\circ\!\!-\!\!\bullet\, -\omega^2 \cdot \mathcal{F}\{G(x)\} \\ \nabla^2 G(x) &\circ\!\!-\!\!\bullet\, -\omega^2 G(\omega) \approx H_{step}(\omega) \end{aligned}$$

This means that, in the case of detecting step intensity changes, the optimal frequency domain filter corresponds to a spatial operator that approximates the second-order derivative of the Gaussian (for a given bandwidth) [24].

### 2.2.2 Marr and Hildreth

Marr and Hildreth [31] proposed an edge detection scheme based on filtering using a 2-D symmetric Gaussian followed by the localization of zero-crossings of $\nabla^2 g(x, y)$, where $g(x, y)$ is the Gaussian filtered image. They addressed two physical considerations that should be combined to determine the appropriate filter: First, the spectrum of the filter should be roughly band-limited in the frequency domain. Second, the filter should be also

smooth and localized in the spatial domain, and in particular its spatial variance should be small. Considering the localization of such a filter both in the spatial and frequency domain, they inferred the requirements that the filter should be both band- and support-limited. Furthermore, they remarked that the conflicting requirements of support- and band-limitedness are optimally reconciled by minimizing the support-bandwidth product. Since the Gaussian achieves the minimum with respect to the uncertainty principle (see Section 2.1.1), they chose a Gaussian.

Marr and Hildreth were interested in finding points at the maximum of the directional derivative as edge locations, and chose to locate these maxima as zero-crossings of the second-order derivative. Considering the cost of computation, they used an isotropic derivative operator, $\nabla^2$ (as the second-order derivative), and computed $\nabla^2(G * f)$, where $G$ is the Gaussian and $f$ is the input image. Since

$$\nabla^2(G * f) \quad = \quad (\nabla^2 G) * f,$$

they convolved $f$ with $\nabla^2 G$, which they approximated by a difference of Gaussians (DoG).

Although their approach has inspired a good deal of research on edge detection, their ideas are relatively intuitive and not based on a well-founded theoretical scheme. Marr and Hildreth assumed that coincident zero-crossings from a set of independent $\nabla^2 G$ channels over a contiguous range (by varying the scale parameter $\sigma$ of the Gaussian) imply the existence of an edge and conversely. The problem is that this so-called *spatial coincidence assumption* is not well-supported by any formal argument. The only situation for which this assumption really makes sense is that of e.g. a straight-line-like, very sharp edge between fairly large regions with constant intensity. In a given image, however, only structures over a certain range of scales can be observed, while fine details disappear as scale parameter increases.

### 2.2.3 Canny

Canny [12], [13] investigated the desirable properties of an optimal edge detector based on the efficiency of detection and reliability in localization. Canny used a step edge model (see Section 2.1.4) corrupted by additive white Gaussian noise and precisely formulated the criteria for effective edge detection. It was assumed that detection is performed by

convolving the noisy edge with a spatial function $h(x)$ which is the impulse response of the filter to be sought. The edges can be marked at the maxima from the output of convolution.

The input signal $f(x)$ can be represented as

$$f(x) = A\mathcal{H}(x) + n(x),$$

where $A$ is the amplitude of the step edge, $\mathcal{H}(x)$ is the Heaviside function, and $n(x)$ is noise. The filter output $g(x)$ is given by the convolution integral as

$$\begin{aligned} g(x) &= \int_{-\infty}^{+\infty} h(\xi) f(x-\xi) d\xi \\ &= A\int_{-\infty}^{+\infty} h(\xi)\mathcal{H}(x-\xi)d\xi + \int_{-\infty}^{+\infty} h(\xi) n(x-\xi) d\xi \\ &= A\int_{-\infty}^{x} h(\xi)d\xi + n_0\sqrt{\int_{-\infty}^{+\infty} h^2(\xi)d\xi}. \end{aligned}$$

Let $g_{edge}$ and $g_n$ denote respectively the responses of the filter to an edge and to noise. Then the signal-to-noise ratio (SNR) at the edge (i.e., at $x = 0$) can be derived as

$$\begin{aligned} \text{SNR} &= \frac{g_{edge}(x)}{g_n(x)} \\ &= \frac{A\int_{-\infty}^{0} h(\xi)d\xi}{n_0\sqrt{\int_{-\infty}^{+\infty} h^2(\xi)d\xi}}, \end{aligned}$$

where ${n_0}^2$ is the variance of the input noise. Assuming there is a local maximum in the total response at the point $x = x_0$, then we have

$$g'_{edge}(x_0) + g'_n(x_0) = 0.$$

The Taylor expansion of $g'_{edge}(x_0)$ at $x = 0$ gives

$$g'_{edge}(x_0) = g'_{edge}(0) + g''_{edge}(0)x_0 + O({x_0}^2).$$

Since the response of the filter in the absence of noise has a local maximum at $x = 0$, i.e.,

$g'_{edge}(0) = 0$, the above equations give

$$\begin{aligned} g''_{edge}(0)x_0 &\approx -g'_n(x_0) \\ {x_0}^2 &\approx \frac{{g'_n}^2(x_0)}{{g''_{edge}}^2(0)}, \\ [h(x) * \mathcal{H}(x)]'' &= h'(x) * \mathcal{H}'(x) \\ \overline{x_0} &= \frac{\sqrt{{n_0}^2 \int_{-\infty}^{+\infty} h'^2(\xi)d\xi}}{\left|A \int_{-\infty}^{+\infty} h'(\xi)\delta(x-\xi)d\xi\right|} \\ &= \frac{n_0\sqrt{\int_{-\infty}^{+\infty} h'^2(\xi)d\xi}}{A|h'(0)|}. \end{aligned}$$

$\overline{x_0}$ is an approximation to the standard deviation of $x_0$. The localization is derived from the reciprocal of $\overline{x_0}$ as

$$\begin{aligned} \text{Localization} &= \frac{1}{\overline{x_0}} \\ &= \frac{A|h'(0)|}{n_0\sqrt{\int_{-\infty}^{+\infty} h'^2(\xi)d\xi}}. \end{aligned}$$

Let $h_\sigma$ denote the spatially scaled filter derived from $h$, where $h_\sigma(x) = h(\frac{x}{\sigma})$. Substituting $h_\sigma$ into the formulae for SNR and for Localization, one can obtain the performance of the scaled filter with respect to SNR

$$\sum h_\sigma = \sqrt{\sigma} \sum h$$

and to Localization

$$\bigwedge h'_\sigma = \frac{1}{\sqrt{\sigma}} \bigwedge h'.$$

These equations imply that the signal-to-noise ratio of an ideal step edge is favored by a broad filter, whereas a narrow filter gives better localization than a broad one. The product of SNR and Localization is invariant under the changes in scale parameter, i.e.,

$$\left(\sum h_\sigma\right) \cdot \left(\bigwedge h'_\sigma\right) = \left(\sum h\right) \cdot \left(\bigwedge h'\right).$$

Through the scaling of $h$ one can balance the detection performance against localization, but both cannot be improved simultaneously.

The natural choice for the composite criterion from above is to maximize this product. Canny [12], [13] used three criteria to derive an optimal operator:

1. good detection ability, i.e., there should be a low probability of failing to detect real edges and of falsely detecting edges that do not exist,

2. good localization ability, i.e., the position of the detected edge should be as close as possible to the true position of the edge, and

3. uniqueness of detection, i.e., a given edge should be detected only once.

The first two criteria are related by *an uncertainty principle*; as the detection ability increases by using filters with broad local support, the localization ability decreases, and vice versa. The impulse response of the optimal step edge filter was shown to approximate the first derivative of a Gaussian ([12], [13]).

# 3 Survey of Scale-Space Method

Computer vision deals with the problem of deriving meaningful and useful information from the visual data ([3], [28]). Koenderink [25] pointed out that in every image analysis task the problem of *spatial scale* must be faced, since a given image has a limited window (the "outer scale") as well as a limited resolution (the "inner scale"). In particular, the first stages of visual processing that perform directly on raw image data are collectively termed the *visual front-end* by him. If a vision task is approached without strict presumption (or *a priori*) with respect to the visual front-end and its tasks, then a fundamental question arises: What information should be extracted, and for that, which operators to use, where to apply, and how they should be related to scale. The task of extracting information from image data is severely influenced by the inherent *measurement* problem meaning that real-world structures appear in different ways depending on the scale of observation ([25], [28], [30]). As a consequence, if one attempts to describe fully the structure of objects in the world projected onto 2-D images, a multi-scale representation is of importance. The main idea of creating a multi-scale representation is to generate a one-parameter family of signals derived from the original signal such that details are successively suppressed with increasing scale parameter.

Witkin [40] proposed a new way of describing zero-crossings across scales. Since the scale may be considered as a continuous parameter, a 1-D signal is first smoothed by convolution with a Gaussian filter of varying scale parameter $\sigma$, and then zeros of the second derivative are localized and followed through scales as the size of the filter increases (decreases). Koenderink [25] soon furthered the approach, which has been developed into the scale-space theory providing a well-founded mathematical framework for dealing with image structure in general (see e.g. [16]). The scale-space theory focuses on the basic fact that image structures, like objects in the world, exist as meaningful entities over certain ranges of spatial scale, and that one neither can expect to know in advance what scales are appropriate for detecting them nor where in the image which scale appears. For example, a differential operator is often used to detect edges (see Section 2.1.3). The obtained edges depend not only on the image data but also on the scale of the operator, i.e., the size of the kernel of local operator used.

From a given signal one can generate a family of derived signals by successively removing fine-detail structures when moving from finer to coarser scales, according to the scale-space theory. The behavior of structure as scale changes can be analytically described based on a precise mathematical definition. The essence of the results from scale-space theory is that if one assumes that the first stages of visual processing have to be as uncommitted as possible and have no *a priori* knowledge about the world from which the image stems, then convolution with Gaussian kernels and their $n$-th order derivatives of different scale is singled out as a canonical class of low-level operators (see e.g. [25]).

## 3.1 The Principles of Scale-Space

A principled approach proposed by Witkin [40] and Koenderink [25] for obtaining such a multi-scale representation of a measured signal is to embed the signal into a one-parameter family of derived signals, the linear scale-space, where the scale parameter, $t \in \mathbb{R}_+$[4], describes the current scale. The attractive aspect of the scale-space theory lies in its solid mathematical framework. For this reason quite some emphasis has been put on the development of the theoretical foundations. Applications of the scale-space theory can be found by now in any field of analysis of sampled data in the spatial, temporal, or whatever domain ([29]).

The linear scale-space concept and the associated theory provides a conceptually robust model for early visual computations. On the other hand, there are certain limitations in basing a vision system on the Gaussian kernel only ([30]); e.g., smoothing across object boundaries may effect both the shape and the localization of edges in edge detection. Similarly, surface orientation estimates computed by shape from texture algorithms are affected since the anisotropy of surface pattern may decrease when smoothed using a rotationally symmetric Gaussian. These are some basic motivations for considering non-linear extensions from the linear scale-space theory. In this report, however, we consider only the linear scale-space theory.

[4] $\mathbb{R}_+$ denotes the set of real non-negative numbers, and $\mathbb{R}_+ \setminus \{0\}$ the corresponding set excluding zero.

### 3.1.1 Gaussian Smoothing and Linear Diffusion

Given a continuous 1-D signal $f : \mathbb{R} \to \mathbb{R}$, the scale-space representation $L : \mathbb{R} \times \mathbb{R}_+ \to \mathbb{R}$ is defined such that the representation at "zero scale ($t = 0$)" is equal to the original signal

$$L(x; 0) = f(x),$$

and the representation at coarser scales are given by convolution of the signal with the Gaussian kernel of successively increasing scale parameter

$$L(x; t) = G(x; t) * f(x),$$

where $G : \mathbb{R} \times \mathbb{R}_+ \setminus \{0\} \to \mathbb{R}$ is the 1-D Gaussian kernel. Then the scale-space representation at the scale $t$[5] is given by

$$L(x; t) = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi t}} e^{\frac{-\xi^2}{2t}} f(x - \xi) d\xi.$$

In terms of differential equations, the evolution of 1-D image structure over scales of the scale-space family $L$ can be described by the 1-D linear diffusion equation given by

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

This notation allows for an analogy and gives a direct physical interpretation of the Gaussian smoothing filtering: The scale-space representation $L$ of a signal $f$ can be understood as the result of letting an initial heat distribution $f$ evolve over time $t$ in a homogeneous spatially infinite medium. Hence, it can be expected that fine-scale details will disappear and images become more diffuse as the scale parameter increases ([17], [29], [30]).

Koenderink [25] introduced *causality* (see Section 3.1.2) in the 2-D scale-space theory, meaning that new level surfaces must not be created when the scale parameter increases. By combining *causality* with the notions of *homogeneity* and *isotropy*, which essentially means that all spatial points and all scale levels must be treated in a same manner, he showed that the scale-space representation for a 2-D signal by necessity satisfies the diffusion equation,

$$\frac{\partial L}{\partial t} = \frac{1}{2} \nabla^2 L = \frac{1}{2} \left( \frac{\partial^2 L}{\partial x^2} + \frac{\partial^2 L}{\partial y^2} \right).$$

[5]The scale parameter $t$ is equal to $\sigma^2$ used in the previous sections.

Since convolution with the Gaussian kernel describes the solution of the diffusion equation for an infinite domain, the Gaussian kernel is the unique one for generating a scale-space, which has been theoretically proved (see e.g. [1], [41]). It can be regarded as well-established that the scale-space formulation in terms of the diffusion equation, within the class of linear transformations, describes the canonical way for constructing a multi-scale representation.

### 3.1.2 Linear Scale-Space Properties

The family of the scale-space representation with increasing scale parameter $t$ possesses some attractive properties ([29], [38]) :

- **Non-creation of local extrema/zero-crossings**

  As the scale parameter is increased, additional local extrema or additional zero-crossings cannot appear; note that an extremum in $L$ is equivalent to a zero-crossing in the first derivative of $L$. The non-creation of the local extrema means that zero-crossings in any derivative of $L$ form a closed curve across scales. This property shows that Gaussian convolution satisfies certain sufficiency requirements for being a smoothing operation.

- **Causality**

  The necessity of a Gaussian smoothing for the scale-space representation is proved by introducing the concept of causality; new level surfaces must not be created in the scale-space representation when the scale parameter is increased. That is to say, the property of causality implies that $L(\cdot; t_2)$ depends exclusively on $L(\cdot; t_1)$ if $t_2 > t_1$ $(t_1, t_2 \geq 0)$.

- **Shift-Invariance**

  The Gaussian blurring as a linear filtering is shift-invariant.

- **Semi-group Property**

  Convolving a Gaussian kernel with a Gaussian kernel results in another Gaussian kernel of which scale value corresponds to the sum of the scale values of each Gaussian

kernel, i.e.,

$$G(\cdot; t_1) * G(\cdot; t_2) = G(\cdot; t_1 + t_2).$$

In terms of the scale-space representation, this means that a representation at a coarse scale $L(\cdot; t_2)$ can be computed from the representation at a finer scale $L(\cdot; t_1)$ by convolution with a Gaussian kernel with parameter value $t_2 - t_1 > 0$,

$$\begin{aligned} L(\cdot; t_2) &= G(\cdot; t_2) * f \\ &= (G(\cdot; t_2 - t_1) * G(\cdot; t_1)) * f \\ &= G(\cdot; t_2 - t_1) * (G(\cdot; t_1) * f) \\ &= G(\cdot; t_2 - t_1) * L(\cdot; t_1). \end{aligned}$$

## 3.2 Multi-Scale Approaches to Edge Detection

The idea of scale is critical for a symbolic description of the significant intensity changes in images or other classes of signals. Such changes must be detected at different levels of scale, since in general different physical causes may be associated with a characteristic behavior across different scales. In an image, intensity changes take place at many spatial scales depending on their physical origin, e.g., abrupt changes at occluding boundaries versus smooth changes at shadows. A multi-scale analysis, i.e., tracking the behavior of some feature of the original signal across scales, can reveal precious information about the nature of the underlying physical process which gives rise to the feature in the image.

However, there is a trade-off problem, which is common for all edge detection schemes concerning the amount of smoothing, i.e., a good detection versus a poor localization, and vice versa (see Section 2.2.3), In other words, a large amount of smoothing generally on one hand has the desirable effect of increasingly suppressing noise and other interfering fine-scale structures. This, on the other hand, may lead to shape distortions at edges by smoothing across object boundaries. By contrast, a small amount of smoothing with a smaller scale operator can improve the localization at the cost of a lower signal-to-noise ratio. To cope with this kind of problem, several approaches have been advocated. In the field of computer vision, Rosenfeld (1971) was one of the first who explicitly proposed an edge detection scheme based on a multiple resolutions. Marr (1976) strongly advocated the use of the second-order derivatives of Gaussian-shaped filters of different sizes with the goal of detecting intensity changes at different scales. Moreover, Young (1985) showed that there are receptive fields in the mammalian retina and visual cortex, the measured response profiles of which can be modeled by Gaussian derivatives. It is interesting to note that receptive fields similar to the Laplacian of the Gaussian (center-surround receptive fields) have been reported to be dominant in the retina (see e.g. [23]).

In the following sections, we will review some of the key approaches with respect to the scale-space method used in edge-detection schemes.

### 3.2.1 Korn's Method

Korn [26] suggested a multi-scale edge detection scheme based upon the gradient of a

normalized Gaussian. He showed that special gradient operators with different scales can be defined by normalization, which are suitable for the comparison of differently filtered images. He assumed that edge detection should be based on the measurement of intensity changes across regions in an image, which is determined by the convolution of $f(x,y)$ with operators XGG and YGG (the normalized gradient of the Gaussian in the direction of the $x-$ and of the $y-$axis, respectively) which are defined by

$$\begin{aligned}
\text{XGG} &= k(\sigma)\frac{\partial}{\partial x}G(x,y;\sigma) = -k(\sigma)\cdot\frac{x}{\sigma^2}G(x,y;\sigma) = \hat{G}_x,\\
\text{YGG} &= k(\sigma)\frac{\partial}{\partial y}G(x,y;\sigma) = -k(\sigma)\cdot\frac{y}{\sigma^2}G(x,y;\sigma) = \hat{G}_y,
\end{aligned}$$

where $G(x,y;\sigma)$ is the 2-D Gaussian with standard deviation $\sigma$.

The normalization requirement states that positive and negative parts of the first-order derivatives of the 2-D Gaussian should be separately normalized to +1 and -1, respectively ([26]);

$$\begin{aligned}
k(\sigma)\int_{-\infty}^{+\infty}\int_{0}^{+\infty}\left|\hat{G}_x\right|dxdy &= k(\sigma)\int_{-\infty}^{+\infty}\int_{0}^{+\infty}\left|-\frac{x}{\sigma^2}G(x,y;\sigma)\right|dxdy\\
&= \frac{k(\sigma)}{\sigma^2}\int_{-\infty}^{+\infty}\int_{0}^{+\infty}xG(x,y;\sigma)dxdy\\
&= -\sigma^2\frac{k(\sigma)}{\sigma^2}\int_{-\infty}^{+\infty}\int_{0}^{+\infty}\frac{\partial G(x,y;\sigma)}{\partial x}dxdy\\
&= -k(\sigma)\int_{-\infty}^{+\infty}G(y;\sigma)dy\int_{0}^{+\infty}\frac{\partial G(x;\sigma)}{\partial x}dx\\
&= -k(\sigma)\cdot 1\cdot\int_{0}^{+\infty}-\frac{x}{\sigma^2}\cdot\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{x^2}{2\sigma^2}}dx\\
&= k(\sigma)\cdot\frac{1}{\sqrt{2\pi}\sigma^3}\cdot\int_{0}^{+\infty}xe^{-\frac{x^2}{2\sigma^2}}dx\\
&= k(\sigma)\cdot\frac{1}{\sqrt{2\pi}\sigma^3}\cdot\sigma^2\\
&= k(\sigma)\cdot\frac{1}{\sqrt{2\pi}\sigma}.
\end{aligned}$$

That is to say, according to the normalization requirement,

$$k(\sigma)\int_{-\infty}^{+\infty}\int_{0}^{+\infty}\left|\hat{G}_x\right|dxdy = k(\sigma)\int_{0}^{+\infty}\int_{-\infty}^{+\infty}\left|\hat{G}_y\right|dxdy = 1,$$

$k(\sigma) = \sigma\sqrt{2\pi}$ is obtained.

XGG and YGG represent both components of the gradient of the Gaussian, leaving aside the normalization factor $k(\sigma)$. Then, convolution of the image intensity function $f(x, y)$ with $\nabla G(x, y; \sigma)$ can be written

$$-k(\sigma)(\nabla G(x, y; \sigma) * f) = -k(\sigma)\nabla(G(x, y; \sigma) * f) = (n_1, n_2)^T = n.$$

Edge detection is now reduced to the problem of finding appropriate features in the vector field $n$. The responses of the convolution can be analyzed by the magnitude $A(x, y; \sigma)$ and the direction $\alpha$ of the vector $n$, which are defined by

$$|n| = \sqrt{n_1^2 + n_2^2} = A(x, y; \sigma), \qquad \alpha(x, y) = \arctan\left(\frac{n_2}{n_1}\right).$$

In this way, the attributes $A$ and $\alpha$ can be assigned to every point in the image. Edges are defined by the maxima of the magnitude of $A(x, y; \sigma)$ in the direction of maximal intensity changes, i.e., in the direction $\alpha$. In the context of multi-scale analysis, on the other hand, the scale parameter $\sigma_i$ of the magnitude of $A(x, y; \sigma)$ is varied

$$|A(x, y; \sigma_1)|, \cdots, |A(x, y; \sigma_n)|.$$

Then, the maximal magnitudes of $A(x, y; \sigma)$ with different scales are compared, searching for a maximum as a function of $\sigma$. Such a scale $\sigma = \sigma_m$ giving a maximum corresponds to the suitable (or, optimal) scale value with respect to the edge to be detected.

### 3.2.2 Edge Focusing and Signature Approach

Edge detection in a gray-value image at a fine resolution is typically subject to noise and unnecessary detail, whereas edge detection at a coarse resolution distorts edge contours through a large degree of blurring causing poor localization, particularly for high-curvature boundaries. To approach this problem, Bergholm [4] suggested the *edge focusing* method, i.e., a coarse-to-fine tracking in a continuous manner.

The main idea of edge focusing is to detect edges at a coarse scale, where the detection problem can be expected to be easier, and then track the edges to a finer scale, to improve the localization, which can be very poor at coarse scales. It was shown that if the focusing procedure is performed in such a way that the scale step $\triangle\sigma$ is less than $\frac{1}{2}$, then for the

most common edge configurations the edges are guaranteed to move not more than one pixel from one scale level to the next.

The edge focusing scheme, however, does neither address the problem of determining the correct scale in the detection step, nor at which scales the edges should be localized. Therefore, it serves mainly as a selection procedure, which among all the edges at the finer (localization) scale selects those that can be traced to corresponding edges at the coarser (detection) scale.

Marr advocated the use of varying the scale parameter for classifying edges and suggested the *signature* approach for 1-D images. Signature approaches aim at combining the information from an image in different ways in order to obtain a better representation of scene content ([42]). Bergholm and Zhang [6], [42] showed that signature can be obtained by using edge focusing and approximate linking of edge elements between the scale levels in the stack of edge images generated by edge focusing (see, e.g., [4], [5], [6], [21], [42] for further reading).

### 3.2.3 BNS Approach

Back et al. [2][6] proposed a 1-D multi-scale edge detection and description scheme following the approaches of Marr and Korn.

The normalized error integral curve

$$\Phi(x;\tau) = G(x;\tau) * \mathcal{H}(x) = \frac{1}{\sqrt{2\pi}\tau} \int_{-\infty}^{x} \exp\left(-\frac{\xi^2}{2\tau^2}\right) d\xi$$

serves as an explicit model of a smoothed step edge (see Section 2.1.4), which is generalized with arbitrary contrast $c$ to

$$\Phi_g(x;\tau) = a_0 + c\,(G(x;\tau) * \mathcal{H}(x))\,,$$

where $c = |a_1 - a_0|$ and $a_0$ and $a_1$ correspond to the intensity values of local plateaus neighboring the contrast edge. The standard deviation $\tau$ of the normal distribution corresponds to the width of the scaled step edge.

According to Marr's *signature*, the operator response in 1-D can be represented as

$$M_g(x;\tau,\sigma) = \Phi_g(x;\tau) * \hat{G}_x(x;\sigma),$$

[6]The acronym **BNS** stems from **B**ack, **N**eumann, and **S**tiehl.

where $\hat{G_x}(x;\sigma)$ is Korn's operator given by

$$\hat{G}_x(x;\sigma) = k(\sigma)G_x(x;\sigma),$$

whereas the response of the operator applied to $\Phi_g$ is expressed as

$$\begin{aligned} M_g(x,\tau,\sigma) &= \Phi_g(x;\tau) * \hat{G}_x(x;\sigma) \\ &= (a_0 + c\,(G(x;\tau) * \mathcal{H}(x))) * \hat{G}_x(x;\sigma) \\ &= a_0 * \hat{G}_x(x;\sigma) + (c\,(G(x;\tau) * \mathcal{H}(x))) * \hat{G}_x(x;\sigma). \end{aligned}$$

Convolution of the constant $a_0$ with $\hat{G}_x(x;\sigma)$ is derived as

$$\begin{aligned} a_0 * \hat{G}_x(x;\sigma) &= a_0 * \frac{\partial \hat{G}(x;\sigma)}{\partial x} \\ &= \frac{\partial a_0}{\partial x} * \hat{G}(x;\sigma) \\ &= 0 * \hat{G}(x;\sigma) \\ &= 0, \end{aligned}$$

while the second term as

$$\begin{aligned} (c\,(G(x;\tau) * \mathcal{H}(x))) * \hat{G}_x(x;\sigma) &= c\left(\frac{\partial}{\partial x}\left(G(x;\tau) * \hat{G}(x;\sigma)\right) * \mathcal{H}(x)\right) \\ &= c\left(\left(G(x;\tau) * \hat{G}(x;\sigma)\right) * \frac{\partial \mathcal{H}(x)}{\partial x}\right) \\ &= c\left(G(x;\tau) * \hat{G}(x;\sigma)\right), \end{aligned}$$

which is derived using the convolution theorem

$$\begin{aligned} G(x;\tau) * \hat{G}(x;\sigma) &= \mathcal{F}^{-1}\left\{\mathcal{F}\{G(x;\tau)\} \cdot \mathcal{F}\{\hat{G}(x;\sigma)\}\right\} \\ &= \mathcal{F}^{-1}\left\{\int_{-\infty}^{\infty} G(x;\tau)\exp(-j\omega x)\,dx \cdot \int_{-\infty}^{\infty} \hat{G}(x;\sigma)\exp(-j\omega x)\,dx\right\} \\ &= \mathcal{F}^{-1}\left\{\exp\left(-\frac{\omega^2\tau^2}{2}\right) \cdot \sqrt{2\pi}\sigma\exp\left(-\frac{\omega^2\sigma^2}{2}\right)\right\} \\ &= \mathcal{F}^{-1}\left\{\sqrt{2\pi}\sigma\exp\left(-\frac{\omega^2(\tau^2+\sigma^2)}{2}\right)\right\} \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\sqrt{2\pi}\sigma\exp\left(-\frac{\omega^2(\tau^2+\sigma^2)}{2}\right)\exp(j\omega x)\,d\omega \\ &= \frac{\sigma}{\sqrt{\tau^2+\sigma^2}}\exp\left(-\frac{x^2}{2(\tau^2+\sigma^2)}\right). \end{aligned}$$

Through the derivation above, the response of the operator is expressed as

$$M_g(x;\tau,\sigma) = \frac{c\sigma}{\sqrt{\tau^2+\sigma^2}} \exp\left(-\frac{x^2}{2(\tau^2+\sigma^2)}\right).$$

Consequently, the unknown 1-D discontinuity at $x_{max} = 0$ (i.e., the position of the edge) has parameters

$$c = \sqrt{2} M_g(x_{max} = 0, \tau, \sigma_i = \tau)$$

and the scale $\sigma_\Phi = \sigma_i = \tau$.

Since local discontinuity profiles at arbitrary loci in 2-D images can be characterized as possessing prior unknown scales along the gradient direction, the convolution of the intensity function with kernels of fixed spatial operator support results in only locally suboptimal responses and thus it is not reliable. Consequently, Back et al. [2] proposed a scale-space integration scheme for the 2-D case to track the convolution results, i.e., the operator responses, through successive scales until termination of the scheme in an optimum scale, where the operator support optimally fits the transition width of the underlying unknown discontinuity (see [18], [21], [32], [33] for the further investigation of BNS approach).

# 4 Analysis of The Sampled Gaussian Kernel

From the careful review in the previous sections, it becomes clear that the Gaussian kernel plays an key role in edge detection and is the unique kernel for generating a linear scale-space. However, it is necessary to remark here that the Gaussian kernel mentioned in the previous sections is defined continuously in the infinite domain. In general, a continuous function refers to a function defined on the entirety of $\mathbb{R}^n$ in the spatial domain ([20]). Given a continuous signal $f : \mathbb{R} \to \mathbb{R}$, for example, the scale-space representation $L : \mathbb{R} \times \mathbb{R}_+ \to \mathbb{R}$ is defined by $L(x; 0) = f(x)$ and convolution of $f$ with the continuous Gaussian kernel $G : \mathbb{R} \times \mathbb{R}_+ \setminus \{0\} \to \mathbb{R}$ is given by

$$L(x;\sigma) = \int_{-\infty}^{\infty} G(\xi;\sigma) f(x-\xi) d\xi. \tag{1}$$

On the other hand, a discrete function is defined on a set of discrete points where the number of points is either finite or infinite. The spatial functions or signals with which computer vision practically deals are always discrete since their spatial domain in practice is discrete, not continuous ([20]). That is to say, the signal convolved with the Gaussian is not actually continuous but discrete, e.g. intensity values of a digital image. Therefore, it cannot be avoided to discretize[7] the continuous Gaussian kernel. A commonly adapted simple technique for discrete signals is to sample the continuous Gaussian kernel using the rectangle rule of integration. For example, with respect to the linear scale-space theory, (1) is approximated using a sampled Gaussian given by

$$\hat{L}(k;\sigma) = \sum_{n=-\infty}^{\infty} \hat{G}(n;\sigma) \hat{f}(k-n), \tag{2}$$

where $k, n \in \mathbb{Z}$, $\hat{f} : \mathbb{Z} \to \mathbb{R}$, and $\hat{G} : \mathbb{Z} \times \mathbb{R}_+ \setminus \{0\} \to \mathbb{R}$ is a sampled Gaussian kernel.

In the context of the scale-space theory, Lindeberg ([28], [29]) showed that a sampled Gaussian can lead to undesirable effects (see [28, Proposition 12]), which exemplifies the fact that properties derived for the continuous case might be violated after discretization.

[7] Discretization of a continuous function can be considered both spatially and in amplitude; discretization of the spatial coordinates is called a sampling, and amplitude discretization is called a quantization ([19]). In case of the continuous Gaussian kernel, the spatial discretization (i.e., a sampling) is considered only.

In the following sections, we analyze in detail a sampled Gaussian kernel with respect to the sampling period as well as the scale parameter. Based on the sampling theorem we show that the continuous Gaussian kernel cannot be fully reconstructed by sampling, i.e. from a finite number of samples, and it does not properly approximate the continuous Gaussian kernel as the scale parameter decreases.

## 4.1 The Sampling Theorem

When a continuous function $f$ is sampled to $\hat{f}$, the sampling process should be ideally invertible, i.e., it should be possible to reconstruct $f$ from $\hat{f}$ with no error. This cannot be achieved in general since $f$ may be assumed to have arbitrary values between the sampled points. Using the sampling theorem, however, it is desirable to state sufficient conditions in order to make the sampling procedure one-to-one (refer to e.g., [10], [15], [19] and [20] for further details).

If the Fourier transform $F(\omega)$ of a continuous function $f(x)$ is zero for all frequencies $|\omega|$ higher than $\frac{\omega_0}{2}$, that is

$$F(\omega) = 0 \qquad \text{for} \qquad |\omega| > \frac{\omega_0}{2},$$

then $f(x)$ can be uniquely determined from its uniformly sampled values $\{\hat{f}(k)\}$ $(k \in \mathbb{Z})$. In other words, if a function is band-limited then its sampling period $T$ can be selected to satisfy

$$T \leq \frac{2\pi}{\omega_0}.$$

This condition implies that the selection of the sampling period $T$ is dependent on the highest frequency content of the signal to be sampled (the sampling process of the Gaussian is in detail exemplified in Section 4.2).

The sampling theorem states that it is in fact possible to recover the full range of original values with full accuracy given the condition that the function is "band-limited" ([10]). However, if we directly apply the sampling theorem to a real signal which is not bounded in the frequency domain, $T = 0$ must hold, which is clearly unrealistic. That is because a strictly band-limited real signal does not exist in general.

## 4.2 Sampling The 1-D Gaussian Kernel

Let us consider the sampling process of the 1-D Gaussian kernel. The Gaussian $f_G(x)$ and its Fourier transform $F_G(\omega)$ are given by

$$f_G(x) = G(x;\sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}},$$
$$F_G(\omega) = \int_{-\infty}^{\infty} G(x;\sigma) e^{-j\omega x} dx = e^{-\frac{\sigma^2\omega^2}{2}}.$$

For sampling, we introduce the *shah* symbol, III(x), which represents an infinite sequence of unit impulses spaced at unit interval defined ([10, p. 214]) by

$$\mathrm{III}(x) = \sum_{n=-\infty}^{\infty} \delta(x-n).$$

$\mathrm{III}(x)$ is periodic with unit period, and it holds

$$\mathrm{III}\left(\frac{x}{T}\right) = T \sum_{n=-\infty}^{\infty} \delta(x-nT),$$

where $T$ is the sampling period ($T \in \mathbb{Z}_+$).

By multiplying $\frac{1}{T}\mathrm{III}(\frac{x}{T})$ with $f_G(x)$, we obtain a sampled Gaussian $f_G^{\#}(x)$ with the sampling period $T$ in the spatial domain

$$\begin{aligned} f_G^{\#}(x) &= f_G(x) \cdot \frac{1}{T}\mathrm{III}\left(\frac{x}{T}\right) \\ &= \sum_{n=-\infty}^{\infty} f_G(n) \cdot \delta(x-nT). \end{aligned}$$

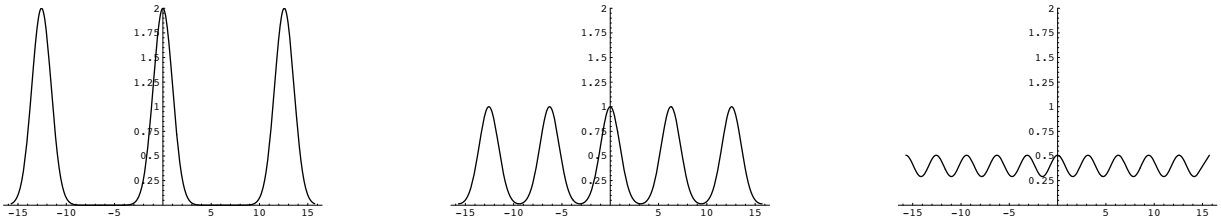
The values of $f_G(x)$ between integers where $x \neq nT$ are not conserved in the product by multiplication, whereas the value of $f_G(x)$ is conserved where $x = nT$. In other words, information about $f_G(x)$ is retained only at the sampling points where $x$ is an integral multiple of the sampling period $T$, while the intermediate values are lost.

Using the property of $\mathrm{III}(x)$ in the Fourier transform ([15, p. 189])

$$\frac{1}{T}\mathrm{III}\left(\frac{x}{T}\right) \circ\!\!-\!\!\bullet\, \mathrm{III}\left(\frac{\omega}{\omega_0}\right) \qquad \left(T = \frac{2\pi}{\omega_0}\right),$$

and the convolution property in the frequency domain ([15, p. 71])

$$f_1(x) \cdot f_2(x) \circ\!\!-\!\!\bullet\, \frac{1}{2\pi}\left(F_1(\omega) * F_2(\omega)\right),$$



(a) $T = \frac{1}{2}$ (b) $T = 1$ (c) $T = 2$

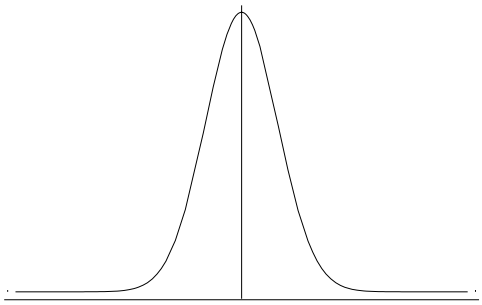
Figure 3: $F_G^{\#}(\omega)$

the corresponding Fourier transform $F_G^{\#}(\omega)$ is derived as

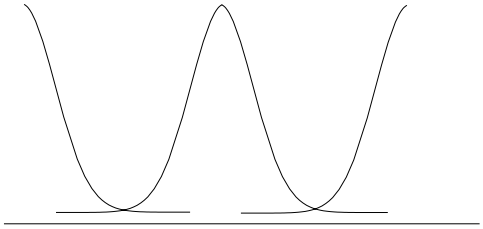
$$\begin{aligned} F_G^{\#}(\omega) &= \frac{1}{2\pi}\left(F_G(\omega) * \mathrm{III}(\frac{\omega}{\omega_0})\right) \\ &= \frac{1}{2\pi}F_G(\omega) * \left(\omega_0 \sum_{n=-\infty}^{\infty} \delta(\omega - n\omega_0)\right) \\ &= \frac{\omega_0}{2\pi} \sum_{n=-\infty}^{\infty} F_G(\omega - n\omega_0) \\ &= \frac{1}{T} \sum_{n=-\infty}^{\infty} F_G(\omega - n\omega_0). \end{aligned}$$

It is noticeable that a *replicating* property appears in the frequency domain when $\mathrm{III}(x)$ enters into convolution ([10]), as seen in $F_G^{\#}(\omega)$ derived above. That is, the multiplication of $f_G(x)$ by $\mathrm{III}(\frac{x}{T})$ has the effect of replicating the spectrum $F_G^{\#}(\omega)$ scaled by the factor $\frac{1}{T}$ at intervals $\omega_0$ in the frequency domain. Fig. 3 shows $F_G^{\#}(\omega)$ [8] when using $T = \frac{1}{2}, 1$, and 2. Since the sampling period $T$ is inversely proportional to $\omega_0$, as $T$ increases the scaling factor decreases and the periodicity becomes smaller, and vice versa. According to the sampling theorem, we can reconstruct $f_G(x)$ from the sampled Gaussian $f_G^{\#}(x)$ when we can recover $F_G^{\#}(\omega)$ satisfying the band-limitedness; provided that we are able to cut off one spectrum $\hat{F}_G(\omega)$ from $F_G^{\#}(\omega)$ for $|\omega| \leq \frac{\pi}{T}$ which brings about no overlap in the replicated peaks, the recovery is possible.

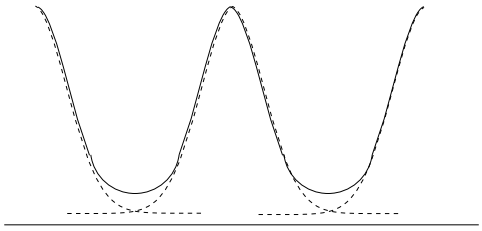
[8] For simplicity, the scale parameter $\sigma$ is set to one.



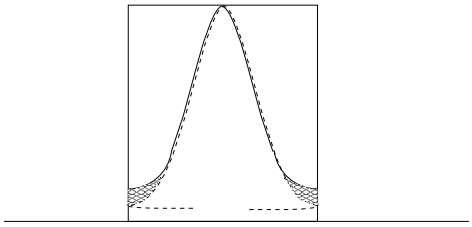
(a) $F_G(\omega)$



(b) $F_G^{\#}(\omega)$



(c) Aliasing effect



(d) $\hat{F}_G(\omega)$

Figure 4: Aliasing; (a) The Fourier Transform of the continuous Gaussian; (b) The replicating effect occurs in the frequency domain by the sampling in the spatial domain; (c) High-frequency components impersonate low frequencies; (d) The cutoff spectrum contains the high-frequency tail caused by the aliasing.

Then, let us cut off one aperiodic spectrum; Fig. 4 illustrates this. The Fourier transform of the continuous Gaussian kernel (Fig. 4-(a)) becomes replicated by the sampling in the spatial domain, so that $F_G^{\#}(\omega)$ (Fig. 4-(b)) has the repetitive spectrum. The Gaussian kernel looks band-limited, however, strictly speaking it is not band-limited. Therefore, when the spectrum is repeated high-frequency components are overlapped; this effect is referred to as an "*aliasing*". The contribution of high-frequency components is reflected to low-frequency components; the solid line depicts the reflection of aliasing effect on the dashed line (Fig. 4-(c)). Using the rectangle function function we cut off one spectrum; we define here the filled area of both sides (Fig. 4-(d)) as the high-frequency tail.

Now we can numerically derive $\hat{F}_G(\omega)$[9]. By multiplying $F_G^{\#}(\omega)$ with the rectangle function $\Pi(\frac{\omega}{\omega_0})$ defined by

$$\Pi\left(\frac{\omega}{\omega_0}\right) = \begin{cases} 1, & |\frac{\omega}{\omega_0}| \leq \frac{1}{2} \\ 0, & |\frac{\omega}{\omega_0}| > \frac{1}{2}, \end{cases}$$

where $\omega_0$ is replaced by $2\pi$ ($T = \frac{2\pi}{\omega_0} = 1$), we obtain $\hat{F}_G(\omega)$;

$$\begin{aligned} \hat{F}_G(\omega) &= \Pi\left(\frac{\omega}{2\pi}\right) \cdot F_G^{\#}(\omega) \\ &= \Pi\left(\frac{\omega}{2\pi}\right) \sum_{n=-\infty}^{\infty} F_G(\omega - 2\pi n) \\ &= \Pi\left(\frac{\omega}{2\pi}\right) \sum_{n=-\infty}^{\infty} e^{-\frac{\sigma^2(\omega-2\pi n)^2}{2}}. \end{aligned}$$

The integral of $\hat{F}_G(\omega)$ is derived as

$$\begin{aligned} \int_{-\infty}^{\infty} \hat{F}_G(\omega)d\omega &= \int_{-\infty}^{\infty} \Pi\left(\frac{\omega}{2\pi}\right) \sum_{n=-\infty}^{\infty} e^{-\frac{\sigma^2(\omega-2\pi n)^2}{2}} dw \\ &= \int_{-\pi}^{\pi} \sum_{n=-\infty}^{\infty} e^{-\frac{\sigma^2(\omega-2\pi n)^2}{2}} dw \qquad (\text{let } \xi = \omega - 2\pi n) \\ &= \sum_{n=-\infty}^{\infty} \int_{-\pi-2\pi n}^{\pi-2\pi n} e^{-\frac{\sigma^2\xi^2}{2}} d\xi \\ &= \underbrace{\int_{-\pi}^{\pi} e^{-\frac{\sigma^2\xi^2}{2}} d\xi}_{n=0} + \underbrace{2\sum_{n=1}^{\infty} \int_{-\pi-2\pi n}^{\pi-2\pi n} e^{-\frac{\sigma^2\xi^2}{2}} d\xi}_{n=\cdots,-2,-1,1,2,\cdots} \qquad (3) \\ &= \int_{-\pi}^{\pi} e^{-\frac{\sigma^2\xi^2}{2}} d\xi + 2\left[\int_{-3\pi}^{-\pi} + \int_{-5\pi}^{-3\pi} + \cdots + \int_{-\pi-\infty}^{\pi-\infty} e^{-\frac{\sigma^2\xi^2}{2}} d\xi\right] \\ &= \int_{-\infty}^{\infty} e^{-\frac{\sigma^2\omega^2}{2}} d\omega \\ &= \int_{-\infty}^{\infty} F_G(\omega)d\omega. \end{aligned}$$

[9]We choose here the sampling period $T$ to be fixed, $T = 1$ ($\omega_0 = 2\pi$), for the reason that first the input signal with which the Gaussian kernel is convolved is in general the intensity function of digital images whose pixel distance is one, and second $F_G^{\#}(\omega)$ is not influenced by the scaling factor when $T = 1$ (see Fig. 3).

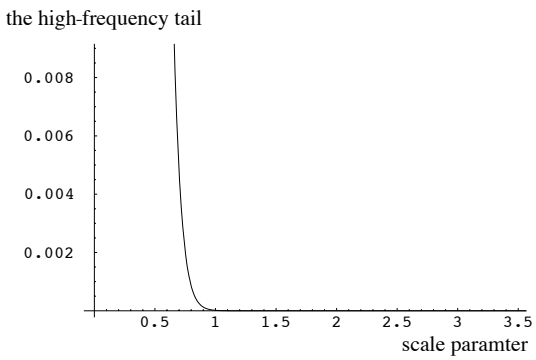



Figure 5: The high-frequency tail of a sampled Gaussian ($\omega_0 = 2\pi$)

It is clear from (3) that the amount of the high-frequency tail is equivalent to that of the aliasing. This implies that we can numerically measure the aliasing by calculating the high-frequency tail. When the amount of the high-frequency tail is null (i.e., no aliasing exists), we can fully reconstruct the continuous Gaussian from the sampled Gaussian. In other words, the smaller amount of aliasing (i.e., the high frequency tail) gives the better sampling result.

As a next step, let us measure the high-frequency tail in case that $T = 1$ (i.e., $\omega_0 = 2\pi$). Since the high-frequency tail of $F_G(\omega)$ belongs to the area $|\omega| > \pi$, the difference between the integral value of $F_G(\omega)$ for $(-\infty, \infty)$ and that of $F_G(\omega)$ for $(-\pi, \pi)$ corresponds to the high-frequency tail $F_{G_{HFT}}$;

$$\begin{aligned} F_{G_{HFT}} &= \int_{-\infty}^{\infty} |F_G(\omega)d\omega - \Pi(\frac{\omega}{2\pi})F_G(\omega)|^2 d\omega \\ &= 2\left[\int_0^{\infty} F_G^2(\omega)d\omega - \int_0^{\pi} F_G^2(\omega)d\omega\right] \\ &= \frac{\sqrt{\pi}}{\sigma} - \frac{\sqrt{\pi}}{\sigma}\mathrm{erf}(\pi\sigma), \end{aligned}$$

where the error integral $\mathrm{erf}(x)$ is defined as

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{\xi^2} d\xi.$$

$F_{G_{HFT}}$ is expressed as the equation with respect to the scale parameter $\sigma$ (see Fig. 5). It is clear from Fig. 5 that as the scale parameter decreases (in case $T = 1$) the high-frequency tail increases, which consequently implies that a small scale parameter[10] gives a deficient

[10]Since $F_{G_{HFT}}$ is expressed by $\mathrm{erf}(x)$ of which value depends on $x$, a roughly estimated value can be

sampling result.

On the other hand, turning back to the spatial domain, using the Parseval theorem ([15, p. 73]), which states that

$$\int_{-\infty}^{\infty} |f(x)|^2 dx = \frac{1}{2\pi} \int_{-\infty}^{\infty} |F(\omega)|^2 d\omega,$$

we can derive the high-frequency tail corresponding to the inverse Fourier transform of $F_{G_{HFT}}$. Since

$$F_G(\omega) - \Pi(\frac{\omega}{2\pi}) F_G(\omega) \bullet\!\!-\!\!\circ f_G(x) - \mathrm{sinc}(x) * f_G(x)$$

holds, it follows

$$\begin{aligned} f_{G_{HFT}} &= \int_{-\infty}^{\infty} |f_G(x) - \mathrm{sinc}(x) * f_G(x)|^2 dx \\ &= \frac{1}{2\pi} \left( \frac{\sqrt{\pi}}{\sigma} - \frac{\sqrt{\pi}}{\sigma} \mathrm{erf}(\pi\sigma) \right). \end{aligned}$$

$f_{G_{HFT}}$ is also inversely proportional to $\sigma$ such as $F_{G_{HFT}}$. This result consequently means that the sampled Gaussian kernel is not appropriate for approximating the continuous Gaussian kernel when the scale parameter is small.

## 4.3 Sampling The Higher Dimensional Gaussian Kernel

The sampling process of the 1-D Gaussian was closely investigated in the previous section, where we examined based on the sampling theorem how properly the sampled Gaussian kernel approximates the continuous Gaussian kernel. Given a sampling period fixed to one, we could derive the measure of difference between the continuous Gaussian and the sampled Gaussian as a formula of the scale parameter.

The sampling of the higher dimensional Gaussian kernel is analogous to that of the 1-D Gaussian based on its nice property "separability". In this section, we show in detail the sampling process both of the 2-D and of the 3-D Gaussian kernel, and as done in the 1-D case, derive the measure of difference between the continuous Gaussian and the sampled one in 2-D and in 3-D with a fixed sampling period (e.g., fixed to one).

given only.

**Sampling The 2-D Gaussian Kernel** The 2-D Gaussian[11] is given by

$$^2f_G(x,y;\sigma) = \frac{1}{2\pi\sigma^2}e^{-\frac{x^2+y^2}{2\sigma^2}},$$

which due to its separability may be written as

$$\begin{aligned} ^2f_G(x,y;\sigma) &= \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{x^2}{2\sigma^2}} \cdot \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{y^2}{2\sigma^2}} \\ &= f_G(x;\sigma) \cdot f_G(y;\sigma). \end{aligned}$$

Its Fourier transform is due to the iterated integration law expressed as

$$\begin{aligned} ^2F_G(u,v;\sigma) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} {}^2f_G(x,y;\sigma)e^{-j(ux+vy)}dxdy \\ &= \int_{-\infty}^{\infty} f_G(x;\sigma)e^{-jux}dx \cdot \int_{-\infty}^{\infty} f_G(y;\sigma)e^{-jvy}dy \\ &= e^{-\frac{\sigma^2u^2}{2}} \cdot e^{-\frac{\sigma^2v^2}{2}} \\ &= e^{-\frac{\sigma^2(u^2+v^2)}{2}}. \end{aligned}$$

Analogously to the 1-D case, we use here the 2-D *shah* symbol[12], denoted as $^2\text{Ш}(x,y)$, defined by ([10, Chap. 5])

$$^2\text{Ш}(x,y) = \sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty} {}^2\delta(x-m,y-n),$$

where $^2\delta(x,y)$ is the 2-D impulse symbol as a natural generalization of $\delta(x)$ ([10, Chap. 5]) defined as

$$^2\delta(x,y) = \begin{cases} 0, & x^2+y^2 \neq 0 \\ \infty, & x^2+y^2 = 0 \end{cases}.$$

Based on the separability of $^2\delta(x,y)$ given by

$$^2\delta(x,y) = \delta(x) \cdot \delta(y),$$

[11] We discriminate 2-D and 3-D functions from 1-D ones by adding the superscripts $^2$ and $^3$, respectively, before the notations of functions.

[12] We follow the suggestion by Bracewell for this terminology. See [10, footnote on p. 74].

${}^2\mathrm{III}(x,y)$ is also separable

$$\begin{aligned}{}^2\mathrm{III}(x,y) &= \sum_{m=-\infty}^{\infty}\delta(x-m)\sum_{n=-\infty}^{\infty}\delta(y-n)\\ &= \mathrm{III}(x)\cdot\mathrm{III}(y).\end{aligned}$$

Besides, ${}^2\mathrm{III}(x,y)$ is doubly periodic given as

$${}^2\mathrm{III}(x+m,y+n) = {}^2\mathrm{III}(x,y) \qquad m,n\in\mathbb{Z},$$

and

$$\frac{1}{|XY|}{}^2\mathrm{III}\left(\frac{x}{X},\frac{y}{Y}\right)$$

represents a doubly periodic array of 2-D unit impulses with period $X$ in the $x$ direction and $Y$ in the $y$ direction ([10, p. 90]). Also, as with $\mathrm{III}(\cdot)$ in one dimension, ${}^2\mathrm{III}(\cdot,\cdot)$ has the distinguished property of being its own 2-D Fourier transform, which is easily derived from

$$\begin{aligned}\mathcal{F}\left\{\frac{1}{|XY|}{}^2\mathrm{III}\left(\frac{x}{X},\frac{y}{Y}\right)\right\} &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{1}{|XY|}{}^2\mathrm{III}\left(\frac{x}{X},\frac{y}{Y}\right)e^{-j(ux+vy)}dxdy\\ &= \int_{-\infty}^{\infty}\frac{1}{|X|}\mathrm{III}\left(\frac{x}{X}\right)e^{-jux}dx\int_{-\infty}^{\infty}\frac{1}{|Y|}\mathrm{III}\left(\frac{y}{Y}\right)e^{-jvy}dy\\ &= \mathrm{III}\left(\frac{u}{u_0}\right)\cdot\mathrm{III}\left(\frac{v}{v_0}\right) \qquad \left(X=\frac{2\pi}{u_0},Y=\frac{2\pi}{v_0}\right)\\ &= {}^2\mathrm{III}\left(\frac{u}{u_0},\frac{v}{v_0}\right).\end{aligned}$$

By multiplying $\frac{1}{T^2}{}^2\mathrm{III}(\frac{x}{T},\frac{y}{T})$ with ${}^2f_G(x,y;\sigma)$, then, we obtain a 2-D sampled Gaussian, denoted as ${}^2f_G^{\#}(x,y;\sigma)$, with the sampling period $T$ in the 2-D spatial domain;

$$\begin{aligned}{}^2f_G^{\#}(x,y;\sigma) &= {}^2f_G(x,y;\sigma)\cdot\frac{1}{T^2}{}^2\mathrm{III}\left(\frac{x}{T},\frac{y}{T}\right)\\ &= \sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty}{}^2f_G(x,y;\sigma){}^2\delta\left(\frac{x}{T},\frac{y}{T}\right) \qquad (m,n\in\mathbb{Z})\\ &= \sum_{m=-\infty}^{\infty}f_G(x;\sigma)\delta(x-mT)\sum_{n=-\infty}^{\infty}f_G(y;\sigma)\delta(y-nT)\\ &= \sum_{m=-\infty}^{\infty}f_G(m;\sigma)\sum_{n=-\infty}^{\infty}f_G(n;\sigma)\\ &= \sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty}{}^2f_G(m,n;\sigma).\end{aligned} \tag{4}$$

Now let us derive the Fourier transform of ${}^2f_G^{\#}(x,y;\sigma)$ of the form (4). The convolution integral of given 2-D functions $f(x,y)$ and $g(x,y)$ is defined ([10, p. 331]) by

$$f * g = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(\xi,\nu)g(x-\xi,y-\nu)d\xi d\nu,$$

and this convolution integral is valid in the frequency domain ([15, p. 71]). Therefore,

$$F(u,v) = F_1(u,v) * F_2(u,v) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F_1(\xi,\nu)F_2(u-\xi,v-\nu)d\xi d\nu$$

holds. On the other hand, since the inverse Fourier Transform is extended to 2-D such that

$$f(x,y) = \frac{1}{4\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(u,v)e^{j(ux+vy)}dudv,$$

the inverse Fourier transform of $F(u,v)$ is derived as

$$\begin{aligned}
&F(u,v) \bullet\!\!-\!\!\circ f(x,y)\\
&= \frac{1}{4\pi^2}\int_{u=-\infty}^{\infty}\int_{v=-\infty}^{\infty}\left[\int_{\xi=-\infty}^{\infty}\int_{\nu=-\infty}^{\infty} F_1(\xi,\nu)F_2(u-\xi,v-\nu)d\xi d\nu\right]e^{j(ux+vy)}dudv\\
&= \int_{\xi=-\infty}^{\infty}\int_{\nu=-\infty}^{\infty} F_1(\xi,\nu)\underbrace{\frac{1}{4\pi^2}\int_{u=-\infty}^{\infty}\int_{v=-\infty}^{\infty} F_2(u-\xi,v-\nu)e^{j(ux+vy)}dudv}_{f_2(x,y)\cdot e^{j(\xi x+\nu y)}}d\xi d\nu\\
&= \underbrace{\int_{\xi=-\infty}^{\infty}\int_{\nu=-\infty}^{\infty} F_1(\xi,\nu)e^{j(\xi x+\nu y)}d\xi d\nu}_{4\pi^2\cdot f_1(x,y)}\cdot f_2(x,y)\\
&= 4\pi^2 f_1(x,y)\cdot f_2(x,y).
\end{aligned}$$

Consequently,

$$f_1(x,y)\cdot f_2(x,y) \circ\!\!-\!\!\bullet \frac{1}{4\pi^2}\left(F_1(u,v) * F_2(u,v)\right)$$

holds. Using this 2-D convolution theorem shown above, returning back to (4), the Fourier

transformed sampled Gaussian, denoted as ${}^2F_G^{\#}(u,v;\sigma)$, is derived;

$$
\begin{aligned}
{}^2F_G^{\#}(u,v;\sigma) &= \frac{1}{4\pi^2}\left({}^2F_G(u,v;\sigma) * {}^2\text{III}\left(\frac{u}{\omega_0},\frac{v}{\omega_0}\right)\right) \qquad (T=\frac{2\pi}{\omega_0}) \\
&= \frac{1}{4\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} {}^2F_G(\xi,\nu;\sigma)\,{}^2\text{III}\left(\frac{u-\xi}{\omega_0},\frac{v-\nu}{\omega_0}\right)d\xi d\nu \\
&= \frac{1}{4\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F_G(\xi;\sigma)F_G(\nu;\sigma)\text{III}\left(\frac{u-\xi}{\omega_0}\right)\text{III}\left(\frac{v-\nu}{\omega_0}\right)d\xi d\nu \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} F_G(\xi;\sigma)\text{III}\left(\frac{u-\xi}{\omega_0}\right)d\xi \cdot \frac{1}{2\pi}\int_{-\infty}^{\infty} F_G(\nu;\sigma)\text{III}\left(\frac{v-\nu}{\omega_0}\right)d\nu \\
&= \underbrace{\frac{1}{2\pi}\left(F_G(u;\sigma) * \text{III}\left(\frac{u}{\omega_0}\right)\right)}_{\frac{1}{T}\sum_{m=-\infty}^{\infty}F_G(u-m\omega_0;\sigma)} \cdot \underbrace{\frac{1}{2\pi}\left(F_G(v;\sigma) * \text{III}\left(\frac{v}{\omega_0}\right)\right)}_{\frac{1}{T}\sum_{n=-\infty}^{\infty}F_G(v-n\omega_0;\sigma)} \\
&= \frac{1}{T^2}\sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty} {}^2F_G(u-m\omega_0, v-n\omega_0;\sigma).
\end{aligned} \tag{5}
$$

It is clear from (5) that the Fourier transform of the 2-D sampled Gaussian is separable; that is, ${}^2F_G^{\#}(u,v;\sigma)$ can be derived from multiplying $F_G^{\#}(u;\sigma)$ by $F_G^{\#}(v;\sigma)$. This separability makes the rest of the derivation much easier, since the 1-D result can be directly used for the 2-D (and thus also for the 3-D) calculation.

Meanwhile, convolution with ${}^2\text{III}(\cdot,\cdot)$ in (5) which corresponds to the multiplication in the spatial domain (see Eq. (4)) describes replication in the frequency domain. According to the sampling theorem, as shown for the 1-D case in the previous section, this replication property does not satisfy the condition of the band-limitedness for the reconstruction of the original signal. Therefore, in order to isolate one aperiodic spectrum from the replicated spectra of ${}^2F_G^{\#}(u,v;\sigma)$, we introduce a 2-D rectangular function ${}^2\Pi(\frac{x}{X},\frac{y}{Y})$ $(X,Y \in \mathbb{Z}_+)$ defined by ([10, p.331])

$${}^2\Pi\left(\frac{x}{X},\frac{y}{Y}\right) = \begin{cases} 1, & \frac{x}{X} \quad \text{and} \quad \frac{y}{Y} < \frac{1}{2} \\ 0, & \text{elsewhere} \end{cases},$$

which is expressed as a product

$${}^2\Pi\left(\frac{x}{X},\frac{y}{Y}\right) = \Pi\left(\frac{x}{X}\right)\cdot\Pi\left(\frac{y}{Y}\right).$$

Given a fixed sampling period $T = \frac{2\pi}{\omega_0} = 1$, multiplying ${}^2F_G^{\#}(u, v; \sigma)$ of the form (5) by ${}^2\Pi(\frac{u}{\omega_0}, \frac{v}{\omega_0})$ leads to the band-limited Fourier Transform denoted as ${}^2\hat{F}_G(u, v; \sigma)$,

$$\begin{aligned}
{}^2\hat{F}_G(u, v; \sigma) &= {}^2\Pi\left(\frac{u}{\omega_0}, \frac{v}{\omega_0}\right) \cdot {}^2F_G^{\#}(u, v; \sigma) \\
&= \Pi\left(\frac{u}{2\pi}\right)\Pi\left(\frac{v}{2\pi}\right)\sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty} F_G(u - 2\pi m)F_G(v - 2\pi n) \\
&= \Pi(\frac{u}{2\pi})\sum_{m=-\infty}^{\infty} e^{-\frac{\sigma^2(u-2\pi m)^2}{2}} \cdot \Pi(\frac{v}{2\pi})\sum_{n=-\infty}^{\infty} e^{-\frac{\sigma^2(v-2\pi n)^2}{2}} \\
&= \hat{F}_G(u; \sigma) \cdot \hat{F}_G(v; \sigma).
\end{aligned}$$

As a consequence, ${}^2\hat{F}_G(u, v; \sigma)$ is not replicated and band-limited, though it contains the contribution of the high-frequency components caused by the aliasing effect. In turn, ${}^2\hat{F}_G(u, v; \sigma)$ is separable into $\hat{F}_G(u; \sigma)$ and $\hat{F}_G(v; \sigma)$, which means that the further derivation is reducible to the 1-D calculations.

Here it is noticeable to recollect the fact that the integral of $\hat{F}_G(\cdot, \sigma)$ (i.e., the integral of the (band-limited) Fourier transform of a sampled Gaussian) is equivalent to that of $F_G(\cdot; \sigma)$ (i.e., the integral of the Fourier transform of the continuous Gaussian). This implies that the amount of the overlapped area from the replicated spectra (i.e., from the aliasing effect) is equivalent to the amount of the high-frequency component. As the sampling period $T$ decreases, the overlapped area does also; a limit case is given for $T = 0$, which results in a non-overlapped area and thus brings about a non-sampling effect. Consequently, similarly to the 1-D result, it can be said that the amount of the high-frequency component corresponds to the measure of the error from the sampling. Given a fixed sampling period, this measure can be numerically derived as a formula of the scale parameter.

The measure of difference between ${}^2F_G(u, v; \sigma)$ and ${}^2\hat{F}_G(u, v; \sigma)$ given a sampling period implies the amount of the high-frequency tail of the 2-D Gaussian kernel, which corresponds to the measure of the error of its sampling. From this point of view, the amount of the high-frequency tail of the 2-D Gaussian kernel given a sampling period $T = 1$, denoted as

${}^2F_{G_{HFT}}$, can be calculated analogously to the case of 1-D;

$$\begin{aligned}
{}^2F_{G_{HFT}} &= \int_{-\infty}^{\infty} |F_G(u;\sigma)du - \Pi(\frac{u}{2\pi})F_G(u;\sigma)|^2 du \int_{-\infty}^{\infty} |F_G(v;\sigma)du - \Pi(\frac{v}{2\pi})F_G(v;\sigma)|^2 dv \\
&= 2\left[\int_0^{\infty} F_G^2(u;\sigma)du - \int_0^{\pi} F_G^2(u;\sigma)du\right] \cdot 2\left[\int_0^{\infty} F_G^2(v;\sigma)dv - \int_0^{\pi} F_G^2(v;\sigma)dv\right] \\
&= \left(\frac{\sqrt{\pi}}{\sigma} - \frac{\sqrt{\pi}}{\sigma}\mathrm{erf}(\pi\sigma)\right)^2,
\end{aligned}$$

where the error integral $\mathrm{erf}(x)$ is defined as

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{\xi^2} d\xi.$$

**Sampling The 3-D Gaussian Kernel** As a natural extension of the 2-D Gaussian, the 3-D Gaussian kernel can be sampled in the same way except for adding the independent variable $z$. This is plausible since the Gaussian kernel is separable, and other functions to be used for the sampling are also separable (see [10, p.89-92]).

The 3-D Gaussian kernel is given by

$$\begin{aligned}
{}^3f_G(x,y,z;\sigma) &= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}} \cdot \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{y^2}{2\sigma^2}} \cdot \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{z^2}{2\sigma^2}} \\
&= f_G(x;\sigma) \cdot f_G(y;\sigma) \cdot f_G(z;\sigma),
\end{aligned}$$

and according to the 3-D Fourier transform ([10, p.340-343]), its Fourier transform is

$$\begin{aligned}
{}^3F_G(u,v,w;\sigma) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} {}^3f_G(x,y,z;\sigma) e^{-j(ux+vy+wz)} dxdydz \\
&= e^{-\frac{\sigma^2(u^2+v^2+w^2)}{2}}.
\end{aligned}$$

One can easily induce the derivation from ${}^3F_G(u,v,w;\sigma)$ to ${}^3\hat{F}_G(u,v,w;\sigma)$, analogously to the 2-D case, using ${}^3\delta(\cdot,\cdot,\cdot)$, ${}^3\mathrm{III}(\cdot,\cdot,\cdot)$ and ${}^3\Pi(\cdot,\cdot,\cdot)$ (here we will not describe the sampling process of the 3-D Gaussian in detail since it is analogous to the 2-D case). Consequently, we derive the amount of the high-frequency tail of the 3-D Gaussian kernel which we denote

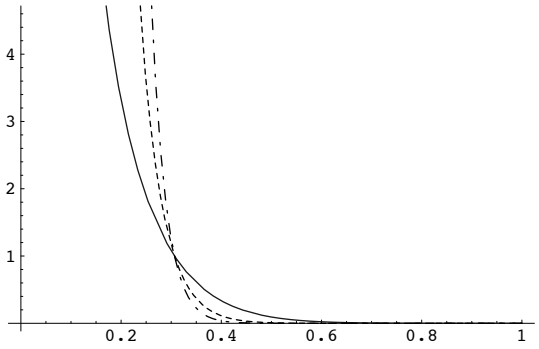



Figure 6: The high-frequency tail of the Gaussian kernel; the horizontal axis is the scale value and the vertical one is the amount of the high frequency tail. The gray curve represents $F_{G_{HFT}}$, the dashed one ${}^2F_{G_{HFT}}$, and the dot-dashed one ${}^3F_{G_{HFT}}$.

as ${}^3F_{G_{HFT}}$. With the sampling period $T = 1$, ${}^3F_{G_{HFT}}$ [13] is

$$ {}^3F_{G_{HFT}} = \left( \frac{\sqrt{\pi}}{\sigma} - \frac{\sqrt{\pi}}{\sigma} \mathrm{erf}(\pi\sigma) \right)^3 . $$

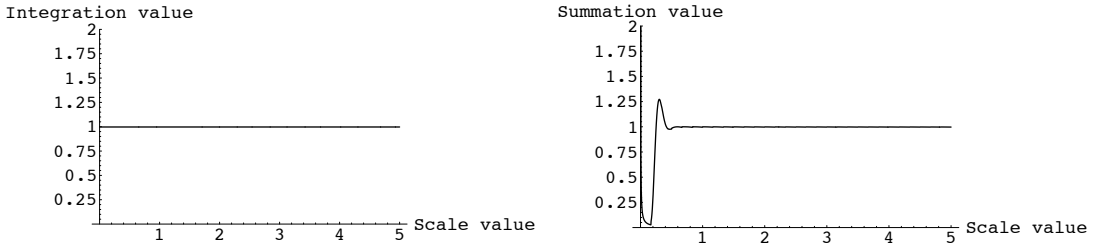
Fig. 6 shows the measure of difference between the continuous Gaussian and the sampled Gaussian with the sampling period of one in 1-D, 2-D, and 3-D; we can easily recognize from Fig. 6 that as the scale parameter decreases the difference increases in each dimension. As a consequent, it can be generally said that a sampled Gaussian with a small scale value is not appropriate for approximating the continuous Gaussian.

## 4.4 Truncating The Gaussian Kernel

Regardless of whether a signal is continuous or discrete, there is no physical device that can generate or store a signal of infinite extension ([20]). Hence, signals in the context of

[13] This derivation can be further generalized to the higher dimensional Gaussian kernel. That is,

$$ {}^N F_{G_{HFT}} = \left( \frac{\sqrt{\pi}}{\sigma} - \frac{\sqrt{\pi}}{\sigma} \mathrm{erf}(\pi\sigma) \right)^N . $$




(a) $\int_{-3\sigma}^{3\sigma} G(x;\sigma)dx$ (b) $\sum_{-3\sigma}^{3\sigma} G(x;\sigma)$

Figure 7: Truncations of the Gaussian

computer vision are not only discrete but also truncated. Therefore, when we are to deal with a signal in practice, we have to consider its truncation, which can be considered both in the continuous case and in the discrete case.

When the Gaussian kernel is used as the convolution kernel (e.g. for generating the scale-space representation), it should be normalized such that

$$\int_{-\infty}^{\infty} G(x;\sigma)dx = 1.$$

However, it is unrealistic to evaluate the infinite convolution integral. From the practical point of view, as mentioned above, it is reasonable to truncate the infinite range of integration with some sufficiently large $N \in \mathbb{Z}$, i.e., by cutting the integral range from $(-\infty$ $\infty)$ short to $[-N, N]$. For doing this, we should determine the value $N$. Hereby, we can

observe an interesting relationship with respect to the Gaussian integral;

$$\begin{aligned} N = \sigma &\rightarrow \int_{-N}^{N} G(x;\sigma)dx = \int_{-\sigma}^{\sigma} G(x;\sigma)dx \approx 0.683 \\ N = 2\sigma &\rightarrow \int_{-2\sigma}^{2\sigma} G(x;\sigma)dx \approx 0.954 \\ N = 3\sigma &\rightarrow \int_{-3\sigma}^{3\sigma} G(x;\sigma)dx \approx 0.997. \end{aligned}$$

This relationship implies that the absolute error caused by the truncation (i.e. from ($-\infty$ $\infty$) to $[-N, N]$), can be kept consistent regardless of the scale parameter value, provided $N = k\sigma$, ($k \in \mathbb{Z}_+$). Fig. 7-(a) depicts that the integration of the Gaussian bounded within $[-3\sigma,\ 3\sigma]$ consistently approximates about 99.7% of the infinite integration regardless of the size of the scale parameter.

In case of a sampled Gaussian kernel, the discrete truncation must be considered. In other words, the infinite range of summation operation (instead of integral operation in the continuous case) has to be truncated with some sufficiently large $N \in \mathbb{Z}$, i.e., by cutting the integral range from ($-\infty$ $\infty$) short to $[-N, N]$. In a similar manner, we apply the relationship between the value $N$ and the scale parameter $\sigma$ to the summation operation. Fig. 7-(b), illustrates that the truncation of the summation operation does not keep the absolute error consistent as the integration operation does. Even worse, as the scale parameter decreases the variation increases.

# 5 Conclusion

Edges in a digital image contain important information for higher level visual processing. At the lowest level of visual processing (i.e., without any *a priori* knowledge of the image to be processed) multi-scale techniques based on the linear scale-space theory should be applied to edge detection as a pre-requisite for generating a rich representation of image structure.

In the first two sections, we looked into the basic fundamentals related to edge detection and reviewed some related approaches to edge detection. Then, we described the principles of the scale-space theory and surveyed some multi-scale approaches to edge detection. Using the property of the scale-space theory, one can obtain a multi-scale representation of the measured signal through Gaussian smoothing, where the Gaussian smoothing can be directly interpreted in terms of the linear heat diffusion equation. Both in edge detection and in the linear scale-space method the Gaussian kernel plays an essential role.

The Gaussian is the unique kernel for generating a linear scale-space of continuous signals, which is theoretically proved ([1], [41]). In practice, however, it is not avoidable to discretize inevitably the continuous Gaussian kernel. Given this problem, a commonly adapted technique is to sample the Gaussian kernel. In the last section, we closely investigated the sampling process of Gaussian kernels (in 1-D as well as in 2-D and in 3-D) in detail, and derived the measure of difference between the continuous Gaussian and the sampled Gaussian expressed as a formula of the scale parameter: When the Gaussian kernel is practically sampled for convolution with discrete signals (e.g. image data), the sampling period is fixed to one. Our question was, in this case, how we can measure the degree of approximation of the sampled Gaussian to the continuous Gaussian. Observing carefully the sampling process of the Gaussian kernel, we could numerically measure the difference between the sampled Gaussian with the sample period of one and the continuous Gaussian, which can be regarded as the error measurement caused by the sampling. From this investigation, as a consequence, it becomes clear that a sampled Gaussian with a small scale value is not appropriate for approximating the continuous Gaussian.

Then, one important question follows: "How can we correctly deal with a Gaussian, whose concept is only valid for the case of continuous signals, for the practical case of

computation in the discrete and finite domain?" In the context of the scale-space theory, one starts from a robust theory, but the gap between a discretization and the continuous theory breaks its robustness. From this point of view, further studies are indispensable for approaching the problems caused by discretization. It is noticeable that Lindeberg [28] remarked that the commonly used techniques with a sampled Gaussian can lead to undesirable effects since violations of the continuous scale-space properties might occur in the corresponding representation. To approach this problem, he presented the discrete scale-space formulation for discrete signals in which the discrete nature of discrete signals is taken into consideration; Lindeberg's discrete scale-space formulation for higher dimensional signals was not fully derived, which is the motivation of our further research on an improved discrete scale-space formulation for higher dimensional signals (especially for 2-D and 3-D discrete signals).

# References


[1] J. Babaud, A. P. Witkin, M. Baudin, and R. O. Duda. Uniqueness of the Gaussian kernel for scale-space filtering. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 8(1):26–33, 1986.

[2] S. Back, H. Neumann, and H. S. Stiehl. On scale-space edge detection in computed tomograms. In *Proc. 11. Mustererkennung DAGM-Symposium, 1989, Springer-Verlag, Berlin*, pages 216–223, 1989.

[3] D. H. Ballard and C. M. Brown. *Computer Vision.* Prentice-Hall, Inc., 1982.

[4] F. Bergholm. Edge focusing. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 9(6):726–741, 1987.

[5] F. Bergholm and K. Rohr. A comparison between two approaches applied for estimating diffuseness and height of step edges. Technical Report CVAP–83, Computational Vision and Active Perception Laboratory, 1991.

[6] F. Bergholm and W. Zhang. On the usage of signatures and recent developments of edge focusing. In *Proc. The 8th Scandinavian Conference on Image Analysis*, pages 1061–1069, 1993.

[7] M. Bertero, T. Poggio, and V. Torre. Ill-posed problems in early vision. *Proceedings of the IEEE*, 76(8):869–889, 1988.

[8] V. Berzins. Accuracy of Laplacian edge detectors. *Computer Vision, Graphics, and Image Processing*, 27:195–210, 1984.

[9] A. P. Blicher. Edge detection and geometric methods in computer vision. Technical Report STAN–CS–85–1041, Stanford University, Dept. Computer Science, 1985.

[10] R. N. Bracewell. *The Fourier Transform and Its Applicationss;third edition.* McGraw-Hill, 2000.

[11] I. N. Bronstein and K. A. Semendjajew. *Teubner-Taschenbuch der Mathematik.* B. G. Teubner, Stuttgart, 1996.

[12] J. F. Canny. Finding edges and lines in images. Technical Report 720, Massachusetts Institute of Technology, Artificial Intelligence Laboratory, 1983.

[13] J. F. Canny. A computational approach to edge detection. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 8(6):679–698, 1986.

[14] A. Diller. *LaTeX Line by Line*. John-Wiley & Sons, 1999.

[15] N. Fliege. *Systemtheorie*. Informationstechnik. Teubner, Stuttgart, 1991.

[16] L. M. J. Florack. *Image Structure*. Kluwer Academic Publishers, 1997.

[17] L. M. J. Florack, B. M. ter Haar Romeny, J. J. Koenderink, and M. A. Viergever. Scale and the differential structure of images. *Image and Vision Computing*, 10(6): 376–388, 1992.

[18] G. Gabrielides, H. Neumann, and H. S. Stiehl. Estimating partial volume induced local blurring of organ contours in computed tomograms. In Lemke et al. [27], pages 556–562.

[19] R. C. Gonzalez and R. E. Woods. *Digital Image Processing*. Addison-Wesley, Reading, MA, 1993.

[20] G. H. Granlund and H. Knutsson. *Signal Processing for Computer Vision*. Kluwer Academic Publishers, Dordrecht, Netherlands, 1995.

[21] A. Hagemann. *Theoretische und experimentelle Untersuchungen zur Attributschätzung von Grauwertkanten*. Diplomarbeit, Fachbereich Informatik, Universität Hamburg, Germany, 1996.

[22] J. W. Harris and H. Stocker. *Handbook of Mathematics and Computational Science*. Springer, 1998.

[23] E. C. Hildreth. The detection of intensity changes by computer and biological vision systems. *Computer Vision, Graphics, and Image Processing*, 22:1–27, 1983.

[24] E. C. Hildreth. Edge detection. Technical Report 858, Massachusetts Institute of Technology, Artificial Intelligence Laboratory, 1985.

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

[26] A. F. Korn. Toward a symbolic representation of intensity changes in images. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 1093(5):610–625, 1988.

[27] H.U. Lemke, M.L. Rhodes, C.C. Jaffee, and R. Felix, editors. *Proc. Computer Assisted Radiology and Surgery 1991 (CAR'91)*. Springer-Verlag, Berlin, 1991.

[28] T. Lindeberg. Scale-space for discrete signals. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 12(3):234–264, 1990.

[29] T. Lindeberg. *Scale-Space Theory in Computer Vision.* Kluwer Academic Publisher, Boston, MA, 1994.

[30] T. Lindeberg and B. M. ter Haar Romeny. *Linear scale-space* (to appear *in Geometry-Driven Diffusion in Computer Vision* ). Series in Mathematical Imaging and Vision. Kluwer Academic Publishers, Dordrecht, Netherlands, 1994.

[31] D. Marr and E. Hildreth. Theory of edge detection. In *Proc. Royal Society of London B 207*, pages 187–217, 1980.

[32] H. Neumann, K. Ottenberg, and H. S. Stiehl. Accuracy of Regularized Differential Operators for Discontinuity Localization in 1-D and 2-D Intensity Functions. In *Proc. 1. Int. IEEE Workshop on Robust Computer Vision*, pages 214–260. R. M. Haralic, W. Förstner (Eds.) Academic Press, 1992.

[33] H. Neumann, K. Ottenberg, and H. S. Stiehl. Finding and Describing Local Structure in Discrete Two-Dimensional Computed Tomograms. In *Proc. IAPR Int. Conf. on Pattern Recognition (ICPR-92)*, pages (III)408–412. IEEE Computer Society Press, 1992.

[34] T. Poggio and C. Koch. Ill-posed problems in early vision: from computational theory to analogue networks. In *Proc. Royal Society of London B 226*, pages 303–323, 1985.

[35] T. Poggio, V. Torre, and C. Koch. Computational vision and regularization theory. *Nature*, 317(26):314–319, 1985.

[36] K. Rohr. Recognizing corners by fitting parametric models. *Internat. Journal of Computer Vision*, 9(3):213–230, 1992.

[37] K. S. Shanmugam, F. M. Dickey, and J. A. Green. An optimal frequency domain filter for edge detection in digital pictures. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 1:37–49, 1979.

[38] J. Sporring, M. Nielsen, L. M. J. Florack, and P. Johansen. *Gaussian Scale-Space Theory.* Kluwer Academic Publishers, 1997.

[39] V. Torre and T. Poggio. On edge detection. Technical Report 768, Massachusetts Institute of Technology, Artificial Intelligence Laboratory, 1984.

[40] A. P. Witkin. Scale-space filtering. In *Proc. of 8th Int. Joint Conf. Artificial Intelligence, Karlsruhe*, pages 1019–1022, 1983.

[41] A. L. Yuille and T. A. Poggio. Scaling theorems for zero-crossings. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 8(1):15–25, 1986.

[42] W. Zhang. *Understanding Intensity and Illumination Transitions.* Dissertation, Department of computer science, University of Stockholm, Sweden, 1995.