CHAPTER 2

Gaussian Scale Space

> Physics is finished, young man. It's a dead-end street.
>
> PHILIPP VON JOLLY,
> MAX PLANCK'S TEACHER

This chapter shortly reviews Gaussian scale space, its axioms and known properties. The interested reader may also take a look at some of the present scale space literature.

- Obviously, one can take Koenderink's first seminal paper [139], as well as some of his tutorials [141] or his book [145]. He generally takes the physical and geometrical point of view.
- The first "scale space book" is by Lindeberg [174]. It may nowadays already be a bit dated, lacking the research of the last ten years, but it still gives a lot of information on and insight in the basic ideas, the transfer of continuous concepts to discrete algorithms, some mathematical properties and applications.
- A more mathematical point of view is taken by Florack [65], showing nicely how "heavy" mathematical equipment is a powerful tool proving that Gaussian scale space is the mathematical way to deal with images.
- A forthcoming book by Ter Haar Romeny [104] gives a tutorial introduction, using the interactive software package Mathematica [250], enabling the user to play around with scale space concepts. A strong emphasis is put on the relation between scale and human vision.
- Weickert [245] discusses the Gaussian scale space in the context of the axiomatics leading to use partial differential equations in image processing, see also section 2.2.

- More detailed information can be found in the proceedings of a scale space workshop in 1996 [125], also published as a book [233], and the proceedings of the subsequent scale space conferences in 1997 [106], 1999 [197], and 2001 [136], although these proceedings also contain a lot of Gaussian-scale-space-related papers, a direct consequence of the results we will discuss in section 2.2.
- Finally, the papers of Salden [221, 222] contain detailed information on the axiomatic view and a lot of citations.

In the next sections we will use the line of reasoning of some of these authors, although most of the following arguments is taken from the works by Florack and Ter Haar Romeny. However, it is presented in a strongly reduced way. For details the reader is referred to the literature mentioned in this chapter.

2.1 Scale Space Basics

In order to understand why at all one should use a (Gaussian) scale space, the underlying concepts of discrete images, some physics and mathematics are combined yielding unavoidable evidence that a scale space is a necessary concept to be used when dealing with images.

2.1.1 The Beginning...

Scale space in the Western literature started in 1983 by a paper of Witkin [249], discussing the blurring properties of one dimensional signals. The extension to more dimensional images was made in 1984 by Koenderink [139]. We will summarise this thought in the following. When blurring an image, each blurred version of the image is caused by the initial image. It is physically not possible that new structures are created (for instance, dark circles or spots in a white area).

This notion of *causality* implies the non-enhancement of local extrema: everything is flattened. So the intensity of maxima decrease and those of minima increase during the blurring process. At these points, all eigenvalues are negative and positive, respectively. The sum of the eigenvalues equals the trace of the Hessian, the matrix with all second order derivatives. Taking for simplicity a 2D image $L(x, y)$, the trace of the Hessian becomes $L_{xx} + L_{yy}$, shortly denoted as ΔL. Then the causality principle states that at maxima $\Delta L < 0$ (being the sum of the eigenvalues) and $L_t < 0$ (decreasing intensity for increasing scale). At minima the opposite holds: $\Delta L > 0$ and $L_t > 0$. And thus, in all cases, $\Delta L \cdot L_t > 0$.

Obviously, this holds in any dimension. One thus obtains an $(n + 1)$-dimensional image $L(\mathbf{x}, t)$, with $\mathbf{x} \in \mathbb{R}^n$. Imposing linearity between ΔL and L_t yields $\Delta L = \alpha L_t$, with $\alpha > 0$ as a possible ("simplest") solution. We may take $\alpha = 1$ without loss of generality[1]. This results in the differential equation

$$\begin{cases} L_t(\mathbf{x}, t) &= \Delta L(\mathbf{x}, t) \\ \lim_{t \downarrow 0} L(\mathbf{x}, t) &= L_0(\mathbf{x}) \, , \end{cases} \tag{2.1}$$

where $L_0(\mathbf{x})$ denotes the original image. This equation has the general solution

$$L(\mathbf{x}, t) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{4\pi t}^{\,n}} \exp\left(-\frac{(\mathbf{x} - \mathbf{x}')^2}{4t}\right) L(\mathbf{x}') d\,\mathbf{x}'. \tag{2.2}$$

[1] Although one may also encounter $\alpha = 1/2$.

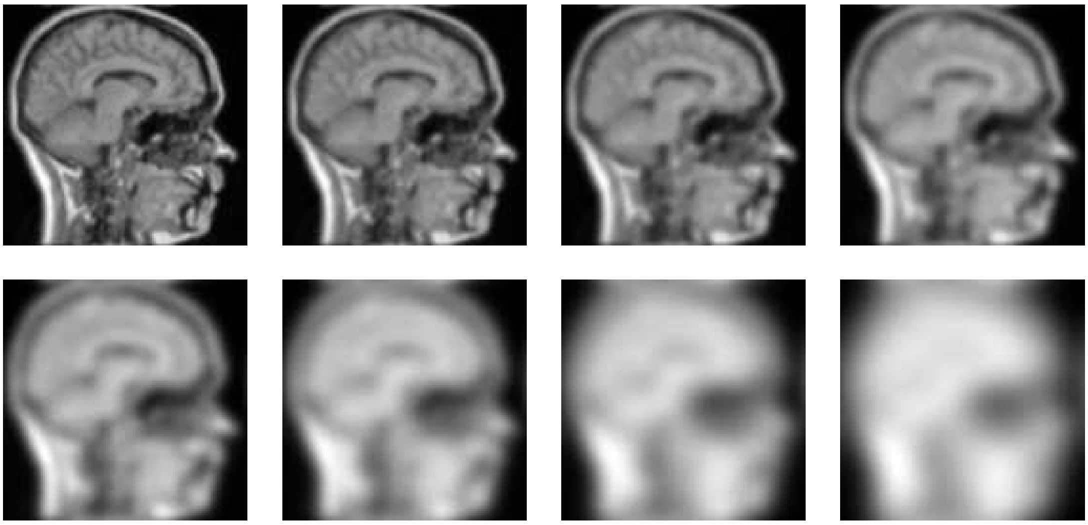

Figure 2.1: An MR image at successive scales $t = \frac{1}{2}e^i$, for $i = 0, \ldots, 4.2$ in steps of 0.6.

So the blurred image has to be taken as the convolution of the original image with a Gaussian filter. An example of a series of blurred images in given in Figure 2.1, showing an MR image at increasing scales. The set of *all* blurred images is the Gaussian scale space (image).

... and Before

Although the papers of Witkin and Koenderink are the start of scale space in the Western literature, the idea was already twenty years old, as Weickert et *al.* describe [245, 247, 248]. The Japanese Iijima [119] wrote a paper deriving the Gaussian as unique filter. Unfortunately, the paper was in Japanese, as some more interesting literature [208, 257, 258, 259].

But there is more to say than that a Gaussian is to be used for *blurring*. This is based on mathematics and physical properties of the image.

2.1.2 Physical Properties

Before applying any algorithm or whatsoever on images, it is firstly necessary to look at the properties of the objects to be described themselves. The observation is that physical units have *dimensions*, like "meters" or "Candela". In any equation describing the objects, the dimensions need to be correct.

The *Law of Scale Invariance* states that the physical laws must be independent of the choice of the parameters. The *Pi Theorem* [15] relates the number of independent dimensionless combinations of the variables and the dimensions of the variables.

As an example, take the flow of some fluid through a pipe. The behaviour of the flow depends on some of the parameters, like density ρ (in kilogram per cubic meter), the velocity v (in meters per second), the diameter of the pipe d (in meters) and the viscosity of the fluid μ (in kilogram per meter per second). In this case the Pi theorem "returns" the combination $Re = \frac{\rho v d}{\mu}$, which is the only parameter in the so-called Navier-Stokes equation describing the flow through a pipe [15].

One can verify that indeed Re is dimensionless. This number, the Reynolds number, is an indicator whether the flow will be turbulent or stay laminar. So one can double the velocity and to be assured that the properties of the flow remain equal, for instance bisect the diameter of the pipe, or double the viscosity. In both cases Re remains equal. The strength of this property lies in the fact that it is possible to build and test scaled versions of object with – more or less – the same properties as the real sized object, *e.g.* ships or planes.

What happens when we obtain an image? Obviously, we look at some object by virtue of some light source, through a certain aperture. Then there are the illuminance of the outside world L_o and the processed image L, both in Candela per squared meter, as well as the sizes of the aperture σ and the outside word $\mathbf{x}$, both in meters. The Pi theorem returns the scale invariances $\frac{L}{L_o}$ and $\frac{\mathbf{x}}{\sigma}$. Clearly, it is *meaningless* to say anything about $\mathbf{x}$ without specifying σ. That is, ignoring σ boils down to the implicit meaningless choice $\sigma = 1$. And consequently, both fractions are related:

$$\frac{L}{L_o} = F(\frac{\mathbf{x}}{\sigma}). \tag{2.3}$$

To determine this relation more precisely, several axioms are imposed. Firstly a mathematical intermezzo is needed, since we deal with discrete data.

2.1.3 Theory of Distributions

The mathematical pre-route to Gaussian scale space follows from Schwartz' "Theory of Distributions" [224]. Its relevance becomes clear from the following example: Disturb a function $f(x)$ by a small perturbation $\epsilon \sin(\delta x), |\epsilon| \ll 1$ and δ arbitrary, *e.g.* $1/\epsilon^2$. It doesn't alter $f(x)$ much. The derivative, however, does: $f'(x) + \epsilon\delta \cos(\delta x)$ shows large variations, compared to $f'(x)$.

This indicates that differentiation is *ill-posed* (in the sense of Hadamard). An operation is well-posed if the solution exists, is uniquely determined and depends *continuously* on the the initial or boundary data. That means: there is one solution and it is stable. So the problem is not the function, but the differentiation performed on it. This gets even worse, when dealing with discontinous functions: they even cannot be differentiated. And in fact, all computer data *is* discontinuous.

To overcome ill-posedness Schartz introduced the (large) *Schwartz space*, containing smooth test functions. These functions are infinitly differentiable and decrease sufficiently fast at the boundaries. A *regular tempered distribution* is the correlation of a smooth test function and some function (or: discrete image). The outcome is that this regular tempered distribution can be regarded as a *probing* of the image with some mathematically nice filter. Derivatives are obtained as probing with the derivative of the filter and indeed depend continuously on the input image.

Having solved the problem of discontinuity, one now needs to find the proper smooth test function in combination with physical laws and state a number of axioms.

2.1.4 Uncommittedness

Concerning the size of the aperture σ, some axioms about *uncommittedness* are desirable. The following axioms state that "we know nothing and have no preference". We remark that that immediately poses a restriction cancelling out cases in which we *do* know something and want to use that.

- *Spatial homogeneity* means that all locations in the field of view are a priori equivalent. So there is no preferred location that should be measured in a different fashion, *i.e.* there is shift invariance.
- *Spatial isotropy* indicates that there is no a priori preferred orientation in a point (or collection of them). Horizontal and diagonal structures are equally measured.
- *Spatial scale invariance* does not discriminate between large, small, and intermediate objects. There is no reason to emphasize details or large areas.
- *Linearity* is imposed, since there is no preferred way to combine observations. Non-linearity in a system implies "feedback", *i.e.* memory, or knowledge.

The axioms of linear shift invariance lead to the observation that the image must be a convolution of the original image by the aparture function. Since in the Fourier domain a convolution of functions becomes the product of them, we will turn to it. Stated in Fourier space Eq. (2.3) becomes

$$\mathcal{L}(\omega;\sigma) = \mathcal{L}_0(\omega)\mathcal{F}(\omega;\sigma). \tag{2.4}$$

The Pi theorem states that $(\omega;\sigma)$ is a function of $(\omega\,\sigma)$. Obviously, there must also be some "hidden" scale ϵ in $\mathcal{L}_0$ such that its argument is likewise dimensionless, say $\omega\epsilon$; this could be something like "pixel scale". The spatial isotropy implies that $\mathcal{F}$ only depends on the magnitude $\Omega = ||\omega\,\sigma||^p$:

$$\mathcal{F}(\omega;\sigma) = \mathcal{F}(\Omega) \tag{2.5}$$

Scale invariance and linearity turn out [65] to require that observing an observed item equals observing it by an aperture size that is a linear combination of the two aperture sizes:

$$\mathcal{F}(\Omega_1)\mathcal{F}(\Omega_2) = \mathcal{F}(\Omega_1 + \Omega_2)\,. \tag{2.6}$$

Eq. (2.6) has the solution $\mathcal{F}(\Omega) = \exp(\alpha\Omega)$, so Eq. (2.5) becoms

$$\mathcal{F}(\omega;\sigma) = \exp(\alpha||\omega\,\sigma||^p). \tag{2.7}$$

For a fixed ω, the limit for σ to zero has to leave the image un-scaled, which is true. For the limit σ to infinity it has to fully scale the image, *i.e.* averaging it completely. So necessarily $\alpha < 0$. For notational purposes (in the diffusion equation as will be shown in section 2.2), we take $\alpha = -\frac{1}{2}$, although one might also encounter $\alpha = -\frac{1}{4}$, the choice of Lindeberg, see *e.g.* [174] and "followers".

This leaves only freedom of choosing p. The additional requirement of *separability* into the spatial dimensions yields $p = 2$, although other values for p still bring up linear scale spaces, albeit non-Gaussian [52]. So we find in the Fourier domain for Eq. (2.7)

$$\mathcal{F}(\omega;\sigma) = \exp(-\frac{1}{2}\omega^2\,\sigma^2) \tag{2.8}$$

and in the spatial domain the inverse Fouriertransform of Eq. 2.8 gives the kernel

$$F(\mathbf{x};\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}^{\,n}}\exp(-\frac{\mathbf{x}^2}{2\sigma^2}). \tag{2.9}$$

Note that Eq. (2.9) is identical to the convolution kernel of Eq. (2.2) when we set $t = \frac{1}{2}\sigma^2$. The name "Gaussian" scale space is obvious. The Gaussian kernel is an element of the Schwartz space, as being a smooth test function: it is infinitely differentiable and it decreases sufficiently fast at the boundaries, just as its derivatives.

2.1.5 Regularisation

Another route to Gaussian scale space is due to regularisation. For a full treatment the reader is referred to Nielsen et *al.* [65, 104, 194, 195, 196]. Here we outline this approach. The task is to find a solution f that is close in the L_2-norm to some signal g, given the constraints that all derivatives of f are also bounded in the L_2-norm. Using so-called Euler-Lagrange multipliers λ_i, this can be combined to minimisation of the "energy"-functional

$$E[f] = \frac{1}{2}\int_{-\infty}^{\infty}\left((f-g)^2 + \sum_{i=1}^{\infty}\lambda_i(\frac{\partial^i}{\partial x^i}f)^2\right)\,d\,x\,. \tag{2.10}$$

In the Fourier-domain Eq. (2.10) is simplified to

$$E[\hat{f}] = \frac{1}{2}\int_{-\infty}^{\infty}\left((\hat{f}-\hat{g})^2 + \sum_{i=1}^{\infty}\lambda_i\omega^{2i}\hat{f}^2\right)\,d\,\omega, \tag{2.11}$$

since the Fourier transform $(\frac{\partial}{\partial x}f(x))^2$ equals $(-i\omega\hat{f}(\omega))(i\omega\hat{f}(\omega))$, the product of the complex function with its complex conjugate. Also Parsevals theorem is used, stating that the Fourier transform of $\int_{-\infty}^{\infty} f^2 dx$ equals $\int_{-\infty}^{\infty}\hat{f}^2 d\omega$, with $\hat{f}(\omega)$ the Fourier transform of $f(x)$. The solution of Eq. (2.11) is found by so-called calculus of variations: $\frac{\delta E}{\delta \hat{f}} = 0$, yielding

$$\hat{f} - \hat{g} + \sum_{i=1}^{\infty}\lambda_i\omega^{2i}\hat{f} = 0\,. \tag{2.12}$$

Consequently, Eq. (2.12) gives the linear system $\hat{g} = \hat{h}^{-1}\hat{f}$. The optimal $\hat{f}$ is thus the linear filtering of $\hat{g}$ by $\hat{h}$. Taking $\lambda_0 = 1$, we find the filter

$$\hat{h}^{-1} = \sum_{i=0}^{\infty}\lambda_i\omega^{2i}. \tag{2.13}$$

The Pi Theorem implies that $\lambda_i \propto \omega^{-2i}$, since $\hat{h}$ is dimensionless. Assuming the semi-group property on this filter, such that filters can be added linearly, one obtains $\lambda_i = t^i/i!$ and thus Eq. (2.13) becomes

$$\hat{h}^{-1} = \sum_{i=0}^{\infty}\frac{t^i}{i!}\omega^{2i} = e^{\omega^2 t}, \tag{2.14}$$

and the Gaussian filter is again obtained, *cf.* Eq. (2.8) with $t = \frac{1}{2}\sigma^2$. In this case separability is included. In fact, a series of regularisation filters $e^{(\omega^2 t)^p}$ can be obtained for $p \in \mathbb{N}$. Results hold for multi-dimensional rotationally invariant regularisation.

2.1.6 Entropy

Nielsen also gave an alternative route, based on the statistics of the aperture function [104, 194]. A statistical measure for the disorder of this aperture function $g(x)$ is given by the entropy, defined as

$$H(g) = \int_{-\infty}^{\infty} g(x)\log[g(x)]d\,x\,, \tag{2.15}$$

using the natural logarithm. This measurement states something like "there is nothing ordered, ranked", if it takes its maximum. However, there are some constraints: Firstly, measuring a complete image shouldn't cause a global enhancement or amplification; the function must be normalised. Secondly, measuring at some point x_0, we do expect the mean of the measurement to be at x_0, which we can take equal to zero, since all points are regarded the same. Thirdly, the function has some size, say σ. So the standard deviation of $g(x)$ is related to this size. Finally, the aperture, as a real object, is positive. These constraints yield

$$\begin{cases} \int_{-\infty}^{\infty} g(x)d\,x &= 1\ , \\ \int_{-\infty}^{\infty} xg(x)d\,x &= 0\ , \\ \int_{-\infty}^{\infty} x^2 g(x)d\,x &= \sigma^2\ , \\ g(x) &> 0 \end{cases} \tag{2.16}$$

Note that the Pi Theorem requires that one replaces g by g/g_0 for some dimensionally compatible constant unit g_0, but this only yields irrelevant constants in Eq. (2.15), given the first constraint. So just as in the previous section, it is the task to maximise $H(g)$, Eq. (2.15), given these constraints. This yields, just as in the previous section an Euler-Lagrange equation:

$$E[g] = \int_{-\infty}^{\infty} \Big(g(x)\log[g(x)] + \lambda_0\, g(x) + \lambda_1 x\, g(x) + \lambda_2\, x^2 g(x) \Big)\ d\,x$$

Again solving $\frac{\delta E}{\delta g} = 0$ gives

$$1 + \log[g(x)] + \lambda_0 + \lambda_1\, x + \lambda_2\, x^2 = 0\ , \tag{2.17}$$

so obviously

$$g(x) = e^{-(1+\lambda_0+\lambda_1 x+\lambda_2\, x^2)}\ . \tag{2.18}$$

Checking the constraints, Eq. (2.16), results in $\lambda_1 = 0$, $\lambda_2 = \frac{1}{2\sigma^2}$, $\lambda_0 = -1 + \frac{1}{4}\log[4\pi^2\sigma^4]$, yielding

$$g(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}\ . \tag{2.19}$$

And again we have the Gaussian, Eq. (2.2), which is, from a statistical point of view, not very strange.

2.2 Differential Equations

The previous sections reveals the Gaussian kernel as a non-spurious detail generating filter, as a smooth test-function, as an uncommitted resultant, as regularisation filter, and as an orderlessness operator. But one can also investigate it from the point of view of differential equations.

2.2.1 Heat Equation

As shown by Koenderink, convolution of the original image with the Gaussian filter is the general solution of the partial differential equation $L_t = \Delta L$. Such a solution is called the *Greens function*, or fundamental solution. It is well-known in the field of physical transport processes. For instance in heat and thermodynamics, where this equation describes the evolution of the temperature when *e.g.* a plate is locally heated [2]. From this field of physics, the equation has become known as the *Heat* or *Diffusion Equation*.

As a consequence, much effort has been put in investigation of this equation, both theoretically (*e.g.* with respect to remaining spatial maxima, the so-called hot spots [13, 14, 30, 34, 123, 182]) and applied-numerically [2, 28, 254]. Note that "scale" has been replaced by "time", a minor *conceptual* change.

One even might put the point of view on this side and take the equation as an *axiom* for scale space. Then a scale space image is the result of an initial image under action of time. In general the "converged" image is the only desired image, the scale space is just the way to reach it.

2.2.2 Partial Differential Equations

A comprehensive investigation of using several different types of partial differential equations (PDEs) has been made by Weickert [245], where also details can be found. We will only shortly mention some of them, to explain the role of scale space (the Heat Equation) in relation to PDEs.

From the physical background, the heat equation originates from *Ficks law* $j = -D \cdot \nabla L$, describing that a flux is caused compensating some concentration gradient by some tensor D, together with the continuity equation $L_t = -\nabla \cdot j$, stating that the diffusion only transports heat. The combination yields $L_t = \nabla \cdot (D \cdot \nabla L)$. If D does not depend on the evolving image, the diffusion is called *linear*. Obviously, Gaussian scale space is obtained by taking $D = \mathbf{1}_n$. *Non-linear diffusion*, depending on the image and its geometry, is also known as *geometry driven diffusion*. Investigation of these models becomes harder, since, in general, Greens functions are not known. However, this is not a problem, since usually only the final, converged, image is of interest. This image is supposed to reveal the best solution to the task the equation is setup for, for example segmentation and / or denoising [35].

Perona Malik

One of the most well-known and relatively simple (but problematic) non-linear scale spaces is obtained by the Perona Malik filter [211], where $D = g(|\nabla L|^2)$, for instance $g(s^2) = 1/(1 + s^2/\lambda^2),\ \lambda > 0$ and $g(s^2) = exp(-s^2/\lambda^2),\ \lambda > 0$. The basic idea is that edges should be preserved, while the rest of the image should be smoothened.

Reaction-Diffusion

These types of PDEs minimise some energy functional under some constraints. Examples leading to Gaussian scale space were given in sections 2.1.5 and 2.1.6. Reaction-Diffusion equations include an extra function describing the (desired) behaviour on edges [245].

Total Variation

Related to the previous PDEs are Total Variation methods. They minimise (some function of) the absolute value of the gradient of the image under certain conditions of the noise (zero mean and given standard deviation). The converged image is smoothened, while edges are preserved [20].

Curvature Based

In the image there are isophotes and, perpendicular to them, flow lines. Instead of smoothing both, one may want to smooth only along the isophotes, ending up with *mean curvature motion*. The motion of the curve is known as *Euclidean shortening flow*, or *geometric heat equation* [138]. An adapted variant of it is (among other names) the *affine shortening flow*. Applications appear in context of active contour models ("snakes") [209].

Morphology

Morphology in its oldest form yields probing an image with a binary structuring element, an $n \times n$ window. This is a discrete model applicable to discrete images, as can be found in any elementary book on image processing, *e.g.* [115, 231]. Applying morphological elements in a "clever" way, one can obtain a multi-scale system [1, 192] Using a parabolic structuring element, yields a morphological scale space equivalent to Gaussian scale space [25, 26, 27, 121, 122]. It has been shown that this equivalence can be expressed by a combining PDE. Both cases appear to be the limiting cases of this PDE [68].

2.3 Biological Inspiration

The human system is capable of looking around and identifying object of different sizes simultaneously. We can see a building with windows and bricks at the same time. All these objects have different sizes. So the eye and the system behind it is capable of working multi-scale [63, 66, 70, 104, 105, 116, 120, 256]. Besides, not only the eye, also our haptic system is a multi-scale system [183].

Models for describing the so-called *receptive field* in the retina can use Gaussian scale space, as argued by Lindeberg and Florack [174, 178, 179], and Koenderink [140, 142, 146, 147, 148, 149], *cf.* ter Haar Romeny [104].

The Laplacean of the Gaussian, known as the Mexican hat [185], can be used to model the sensitivity profile of a so-called *centre surround receptive field*. As Ter Haar Romeny states "we observe the Laplacian of the world" [104].

These observations, and many more that can be found in the literature mentioned, motivate the investigation of Gaussian scale space from the field of biology. But not only Gaussian scale space. At some visual stage in the brain a large amount of feedback to the eye is found. This implies the use of image structure, or memory, or non-linearity. This argues for the use of geometry driven (non-linear) models [102], as a stage next to linear models.

2.4 Hierarchies

An important notion in image analysis is that of a hierarchy: There is some *nesting* of several objects within the image. One can think of a road, containing cars, containing licence-plates, containing numbers and letters. More generally stated: regions within regions. This nesting of regions can be obtained if the regions are known, for instance due to edge detection or segmentations.

One way is to focus solely on the image and to try to build a *graph* [210, 228, 229]. One would like to end up with a tree, *i.e.* a graph without the possibility to walk around and visit parts multiple times.

A *tree structure* is straightforward and simplifiers the structure. However, this is not always possible when describing objects within the image, but difficulties can be reduced by using the special technique of Reeb graphs [18, 19, 81, 137, 226, 227].

Early approaches in image analysis used pyramids, where stacks of images (or image primitives) of decreasing size are generated, *e.g.* by averaging four pixels into one in the next level, thus ending up with pyramidal structures. An example of such a structure based on the Laplacean is due to Burt and Adelson [31]. The idea is that global structures will live long in the pyramid, and the successive disappearance of structure returns a hierarchy. In fact, all hierarchy approaches need some stack of images, or image primitives (like the magnitude of the gradient), that simplifies in some sense going up in the tree. Therefore scale space methods and hierarchical approaches are strongly related, see *e.g.* Nacken [192] and Lester and Arridge [166].

2.4.1 Multi-Scale Watershed

One example of an image primitive is the magnitude of the gradient, yielding a non-linear approach. It is used in watershed segmentations, a very old principle going back at least one and a half century [32, 187]. The idea is that while flooding a landscape, water will flow into pools. At some specific heights, pools will merge. In Mathematical Morphology watersheds are commonly used [122].

Olsen [203, 204, 205, 206, 207] generated a multi-scale watershed algorithm based on Gaussian derivatives: on each scale the watershed is calculated, yielding a *watershed space*. Interactively a user can select regions and refine (or coarsen) them [44]. Multi-scale watershed algorithms yield usable hierarchical (segmentation driven) algorithms [82, 246].

2.5 Two Decades of Linear Scale Space

Since the Gaussian filter has been used for decades in signal processing, *i.e.* the one dimensional case, much effort has been made to investigate its properties. The Laplacean gives information with respect to edges, a reason for investigating its zero crossings (locations where it changes sign) [3, 50, 118, 181, 219, 225, 244, 252]. They can reconstruct the original signal – under certain conditions, see also the series of papers by Johansen (et *al.*) [124, 125, 126, 127, 128, 129].

Emerging since the papers of Koenderink and Witkin, two dimensional investigations started from the image algorithmic point of view [12, 38, 39, 253], including topics like robustness [111, 112] and implementations [21, 22, 23, 24, 213].

In the last decade of the previous century, papers appeared based on the considerations and axioms described in this chapter, *e.g.* by Florack (et *al.*) [67, 69, 75, 76, 77, 78, 79, 80, 107], exploiting the differential structure of scale space, and relations between linear and non-linear scale spaces [68, 73], by Lindeberg (et *al.*), building a hierarchical structure and paying attention to the discrete implementation, [169, 170, 171, 172, 173, 174], and the detection of image entities [175, 176, 177, 180], and by Griffin (et *al.*) [95, 98, 99, 100].

Scale space ideas can also be extended to the uncertainty in grey value detection [92, 96, 97, 150, 151], to optic flow [74], and orientation analysis [133]. Also time can be scaled in a similar fashion (taking into account that time uses a half-axis: the future cannot be modelled!) [65, 108, 143]. A possible tracking of surfaces with the same intensity over scale has been described by Fidrich [60, 61]. From the mathematical point of view, especially the singularities (special points, to be encountered later on) are interesting [17, 45, 46, 47, 48, 71, 72, 154, 155].

Applications [103] can be found in various fields, *e.g.* stochastics [11], statistics [33], clustering [167, 193], recognition [37], segmentation [113, 132], and image enhancement by deblurring [65, 117, 214]. Medical applications of Gaussian scale space can be found in [42, 43, 110, 168, 180, 199, 200, 201, 202, 218, 230, 237] and in the comprehensive overview by Duncan and Ayache [53]. Already in the early nineties the linear scale space segmentation tool called hyperstack was used for segmentation in medical context, see *e.g.* [94, 152, 153, 201, 202, 240, 241, 242, 243]. It should be noted that it also contains heuristic models in order to provide a segmentation.

Recently, Geusebroek et *al.* applied Gaussian scale space theory to colour images, yielding good segmentations [84, 85, 86, 87, 88, 89, 90, 109].

2.5.1 Sub-Structures

Edge detection has been one purpose of image analysis from the beginning. However, edge detection methods do not always yield desired results [58, 59, 62]. Much research has therefore been done on the properties of sub-structures, *i.e.* structure within (scale space) images [134, 234], like curve-linear ones [236], and relative critical sets, like ridges [36, 54, 55, 56, 57, 83] and its extension cores [49], their generic properties and their relations to the medial axis [48, 49, 135], obtained by convolving the original image with $-\sigma^2 \Delta G$, with G the normalised Gaussian.

2.5.2 Deep Structure

Most research has been done on *using* scale space, *e.g.* for selecting some proper scale to derive some nice result like edge-detection or segmentation. The emphasis is then put on the *scale* part of scale space. The scale *parameter* gives extra information or an extra degree of freedom to "play around" with the image.

There is, obviously, also the *space* part of scale space. Then the emphasis is on the extra *dimension* that is available due to the scale parameter. The investigation of this extra dimension is the subject of this thesis. In his original paper, Koenderink called this *deep structure*: the image at all scales simultaneously [139].

Related and relevant research that has been done by others on the field of this deep structure, will be discussed in each of the following chapters.

2.5.3 Poisson Scale Space

Although linear scale space is often the synonym for Gaussian scale space (and vice versa), *this is not true*. The Gaussian scale space is an instance of a linear scale space. Recently, Duits et *al.* [52] showed that given the axioms of section 2.1.4, also another linear kernel can be used if separability[2] is not required, recall Eq. (2.7). Then there are infinite linear scale spaces, spanned by

$$\mathcal{F}(\omega;\sigma) = \exp(\alpha ||\omega\ \sigma||^p). \tag{2.20}$$

[2]Note that separability is a coordinate-dependent notion and therefore not a fundamental requirement.

The specific choice $p = 1$ in n spatial dimensions yields the kernel

$$F(\mathbf{x};\sigma) = \frac{\Gamma(\frac{n+1}{2})\;\sigma}{\pi^{(n+1)/2}\;(||\mathbf{x}||^2+\sigma^2)^{\frac{n+1}{2})}}\;\;, \tag{2.21}$$

where $\Gamma(i)$ is the Euler gamma function, given by $\int_0^\infty t^{i-1}e^{-t}d\,t$. Although F is a smooth function, it is not an element of the Schwartz space. The kernels and the filtered images are harmonic functions, since Eq. (2.21) is the Greens function of another famous physical equation, given by

$$\Delta L + L_{\sigma\sigma} = 0. \tag{2.22}$$

In mathematics Eq. (2.22) is called “the Poisson equation in half space”, and its kernel “half-space Poisson kernel” [101] and one consequently obtains a Poisson scale space. Taking all derivatives within the Δ-operator results in $\Delta_{\mathbf{x},\sigma}L = 0$, the “famous” Laplace-equation with Dirichlet boundary condition, *viz.* $\lim_{\sigma\downarrow 0} L(\mathbf{x},\sigma) = L(\mathbf{x})$, is obtained.

Using proper boundary conditions it can be shown that Eq. (2.22) is equivalent to the evolution equation

$$\frac{\partial L}{\partial\sigma} = -\sqrt{-\Delta}L\;, \tag{2.23}$$

using a fractional power of a derivative operator. In fact, instead of a squareroot, any power between zero and one can be used, see Duits et *al.* [52].

This recent discovery of the infinite set of linear scale spaces may give rise to a revival of investigation of linear scale space properties, both theorethical and practical. Linear scale space isn't finished. It isn't a dead-end street[3].

[3]In 1874, at the age of 16, Max Planck entered the University of Munich. Before he began his studies he discussed the prospects of research in physics with Philipp von Jolly, the professor of physics there, and was told that physics was essentially a complete science with little prospect of further developments.