Scale and the differential structure of images

Luc M J Florack, Bart M ter Haar Romeny, Jan J Koenderink and Max A Viergever

Why and how one should study a scale-space is prescribed by the universal physical law of scale invariance, expressed by the so-called Pi-theorem. The fact that any image is a physical observable with an inner and outer scale bound, necessarily gives rise to a 'scale-space representation', in which a given image is represented by a one-dimensional family of images representing that image on various levels of inner spatial scale. An early vision system is completely ignorant of the geometry of its input. Its primary task is to establish this geometry at any available scale. The absence of geometrical knowledge poses additional constraints on the construction of a scale-space, notably linearity, spatial shift invariance and isotropy, thereby defining a complete hierarchical family of scaled partial differential operators: the Gaussian kernel (the lowest order, rescaling operator) and its linear partial derivatives. They enable local image analysis through the detection of local differential structure in a robust way, while at the same time capturing global features through the extra scale degree of freedom. In this paper we show why the operations of scaling and differentiation cannot be separated. This framework permits us to construct in a systematic way multiscale, cartesian differential invariants, i.e. true image descriptors that exhibit manifest invariance with respect to a change of cartesian coordinates. The scale-space operators closely resemble the receptive field profiles found in mammalian front-end visual systems.

Keywords: scale-space, Gaussian kernel, Gaussian derivatives, differential invariants

Over the last few years there has been an increasing tendency in the image analysis literature towards a multiscale approach. A historical contribution to such an approach was the introduction of the *pyramid*[1]. Though being based on a rather *ad hoc* method of averaging neighbouring pixels, this first model did capture the crucial observation of the inherently multi-scale character of image structure.

For some time there has been discussion on the question of how to generate a *scale-space*, the continuous analogue of the pyramid, in a unique way, as there seemed to exist no clear way to choose among the many possible scale-space filters[2–6]. One obviously needed a set of natural, *a priori* scale-space constraints.

A fundamental approach was adopted by Koenderink[7], Witkin[8] and Yuille and Poggio[9], who formulated an *a priori* constraint in the form of a causality requirement: no 'spurious detail' should be generated upon increasing scale. This, together with some symmetry constraints, unambiguously established the Gaussian kernel (i.e. the *Green's function* of the *isotropic diffusion equation*) as the unique scale-space filter. Its width σ can be identified with spatial scale.

One can model an image as a scalar field on a finite-dimensional manifold and apply fundamental mathematical operations, like differentiations, to reveal local image structure. There exist many useful and rather well-established mathematical disciplines, notably differential geometry, tensor calculus, invariants theory, all of which have an increasing impact on nowadays image structure analysis.

In this paper we discuss the fundamental concept of scaling as well as some natural constraints of a front-end visual system, and show that a complete hierarchical set of *scaled differential operators* follows from these considerations. The lowest order kernel is the isotropic Gaussian. The higher order kernels are the scaled Gaussian derivatives, which constitute the natural differential operators on a given scale.

With this set we can study local image geometry to any desired order. To this end we will introduce the concept of a *local jet* of order N, $J^N[L(P)]$, also called N-jet[10], defined as the equivalence class of functions L which share the same N-truncated Taylor expansion at a given point P. In other words, all images in a given N-jet are locally indistinguishable modulo higher order differences. Such a local N-jet can be represented with respect to a cartesian coordinate system by the set of partial derivatives up to Nth order, evaluated at the point P, so:

$$J^N[L(P)] \triangleq \{L_{i1\cdots i_n}(P)\}_{n=0}^{N} \quad (1)$$

Computer Vision Research Group, Utrecht University hospital, Heidelberglaan 100, 3584 CX Utrecht, The Netherlands

Paper received: 7 February 1992

0262–8856/92/006376–13

The lower spatial indices attached to L all have values within the range $1 \ldots D$, where D is the dimension of the image domain, and denote differentiation with respect to the associated spatial variable.

Derivatives of arbitrary order are generally well-defined and robust provided they can be calculated on a sufficiently high scale (relative to pixel scale and noise correlation width), and provided we have a sufficient resolution of intensity values (dynamic resolution, noise). We will not present a detailed discussion on these trade-offs here, but refer to Blom *et al.*[11]. In this paper we will restrict ourselves to $N \leq 3$.

The approach is valid in D dimensions, whereas much of the literature is limited to 1 or 2 dimensions[1,2,4,8].

THEORY

Physical versus mathematical operators

The only way to obtain structural information about a physical scene is to extract *observables* (i.e. images) with the help of some measuring apparatus. We inevitably have to face the problem of fixing the proper scale, because observables are always characterized by an intrinsic, finite scale range. Its lower bound is determined by the sampling characteristics of the device, whereas the upper bound is limited by the scope of the field of view.

The very fact that an image is a physical observable makes it subject to an extra constraint imposed by the *universal law of scale invariance*, which governs all laws of physics. There is no such scaling constraint on a mathematical, i.e. a dimensionless scalar field, defined on a dimensionless manifold, but it is instructive to observe how mathematicians alternatively constrain it by imposing convenient regularity conditions: a mathematical function is typically assumed to be 'sufficiently smooth', say a $C^N(\Omega)$-function on a D-dimensional domain Ω, with N sufficiently large to justify the operations performed on it. For a physical observable we cannot pose such smoothness constraints.

Clearly, it makes no sense to define a derivative of a sampled image in the strict mathematical sense (this would require the existence of an infinitesimal neighbourhood as well as a smoothness constraint on neighbouring image values). One usually circumvents this problem by considering neighbouring pixels instead of infinitesimal neighbourhoods in the definition of a derivative. A well-known example of this is the 5-point Laplacean kernel[12]. This is, however, a non-robust and rather *ad hoc* solution that crucially relies on imaging conditions, like grid size and pixel shape. Using this operational Laplacean amounts to the implicit assumption that the structures of interest have a spatial extent close to pixel scale. Moreover, it assumes that the structures of this scale are meaningful, which is generally not the case (think of pixel-correlated noise or dithered images).

Disregarding the intrinsic dimensionality of an image or, in other words, the scaling degree of freedom, is the cause of the failure of naively applying differential methods in image analysis. Dimensional analysis is a well recognized concept in physics, and its precise mathematical formulation will be used to argue for the necessity of a multiscale approach and to derive the unique scale-space operators for arbitrary dimensions $D > 1$.

Basic front-end vision constraints

Many interpretations of a front-end vision system are possible. We assume that its sole task is to establish a *representation* of a given observable in a *convenient format*. The *interpretation* is left to dedicated postprocessing routines, which read out the formatted data represented by the front-end (cf. the 'sensorium' in Koenderink[13]). By definition, a front-end vision system is assumed to be completely ignorant of any *a priori* geometry of its input. This lack of *a priori* geometrical knowledge argues for an *a priori* symmetric sampling and preprocessing of its input. Hence it is quite natural to define a front-end vision system by formulating a set of plausible symmetries. We propose the following set*:

- *linearity*: allowing for superposition of input stimuli.
- *spatial shift invariance*: implied by the absence of a perferred location.
- *isotropy*: implied by the absence of a preferred direction.
- *scale invariance*: implied by the absence of a preferred scale.

These basic symmetry requirements are rather weak, because we do not want the front-end system to commit itself to any specific task beyond representation. Note that none of these symmetry constraints are strictly necessary for the sole purpose of data representation, but they do significantly decrease the burden on interpreting routines that address the front-end, since these will now be refrained from the overhead of having to reconcile the data with the symmetries of the environment that are known in advance anyway: the front-end system will make this *a priori* knowledge of the environment *manifest*. In this precise sense, the front-end postulates will make up for a convenient format.

Scale invariance

Let $F(x_1, \ldots, x_D)$ be some physical observable, e.g. the image luminance as a function of spatial coordinates, time, etc. From a pure mathematical point of view there is no restriction whatsoever on the form of the function F. But because we are dealing with a *physical* entity, the requirement of *scale invariance* imposes a restriction on the form of F: only those functions are allowed that 'scale properly'. The precise meaning of this statement is expressed by the following law[14,15]:

*There may be asymmetries in the external environment the system has to operate in, like gravity, etc., which might argue for a less symmetric front-end vision model. On the other hand, this does not really limit the usefulness of these symmetries, since they can always be broken in a postprocessing stage. In addition, more restrictive (usually task-specific) symmetries may be imposed *a posteriori* as well.

Universal law of scale invariance:

Physical laws must be independent of the choice of fundamental parameters.

This is equivalent to:

Dimensional analysis:

A function relating physical observables must be independent of the choice of dimensional units.

Important are those quantities that do not change under the given scalings. These are called *dimensionless*. It is a necessary requirement to be able to express a physical relation in a unit free form. Hence, simple dimensional analysis will reveal scale invariance. Remember, though, that it is the very notion of scale, in relation to the law of scale invariance, that justifies the method of dimensional analysis. The rigorous way of formulating dimensional analysis is through the *Pi-theorem* (for a detailed discussion see Olver[16]).

Inner and outer scale

An image is just another physical observable, with inner scales limited to a finite range determined by the resolution of the sampling device (grid size) and by the field of view.

In image analysis, there is a widespread concern with discretization effects. Strictly speaking, when we are interested in local image structure on the sampling device's inner scale, we are facing an apparent undersampling problem, from which there is only one escape: zooming in on the scene or resorting to a higher resolution acquisition.

Once having fixed the inner scale, all smaller scale image geometry has been destroyed and can by no means be reconstructed. This 'catastrophical' destruction is of an intrinsically irreversible nature. One cannot expect things to be geometrically correct at the limiting lower scale boundary, where we will have 'spurious detail'. The scales of interest should therefore be relatively large compared to the imaging device's sampling width. Local image analysis on these scales will then be of a continuous rather than a discrete nature (for discretization issues, see also Lindeberg[4]).

It pays to study the human front-end visual system[17], being an astonishingly well performing device, having evolved over many millions of years. The front-end focuses on scales which are considerably larger than the eye's true inner scale: the scale of a typical rod or cone. It is not the output of individual rods and cones that is transferred, but only a *weighted* sum over typically several hundreds of them, making up a *receptive field* (RF). The profile of such a RF takes care of the small-scale 'spurious detail' generated by the individual rods and cones by scaling up to a larger inner scale in a very specific way. Only these larger scales are subject to further analysis. Indeed, numerous physiological measurements[18,19] support the theory that RF profiles can be modelled by Gaussian filters of various widths or their partial derivatives[20], which, as we will prove, precisely turns out to provide a complete solution to our front-end requirements.

We also often encounter problems related to the device's limited field of view. Finiteness of the image domain poses restrictions on the largest inner scales that are presented in each point of the image, depending on its position relative to the boundary. The further away from the nearest boundary, the larger the largest scale that is locally represented. On the boundary itself there is no scale information left. Reliable local geometry on a given spatial scale can only be found a certain minimal distance proportional to that scale away from the boundaries. The boundary problem as such, however, is clearly a *scale-independent* problem, similar on all scales, and consists of formulating the trade-off between the (scale-independent) proportionality constant and accuracy together with a rigorous accuracy quantification.

We will not give a rigorous solution for this notorious boundary problem, but merely give a qualitative indication of how to deal with it. To this end, we may resort to the physiology of our own front-end. Here, multiple scaling is essentially achieved already at the acquisition stage (due to the many RF sizes), rather than by a postprocessing of a fixed-scale sampled image (the output of individual rods and cones is ignored as such). Since RF's never overlap with the 'boundary', the boundary problem simply *does not arise* in our visual system. In this operational sense the boundary is non-physical.

These simple though important scale observations should suffice to support the claim that *a multiscale description of image structure is an indisputable necessity in front-end image analysis.*

We will introduce a continuous scale parameter σ to account for the spatial scaling freedom. It has the dimension of a length and is used to define the notion of an 'immediate neighbourhood' of a point P on scale σ as the 'fuzzy' set of points within a sphere of radius $r(\sigma) \propto \sigma$ centred at P, i.e. the smallest spatial 'volume' (a length in 1D, an area in 2D) within which the image structure at that scale varies 'neither too much nor too little'.

Usually, of course, if σ is larger than the pixel size, the image structure *does* vary significantly over a distance σ, because of irrelevant small scale details. So then σ cannot denote inner scale. To reveal the 'pure' σ-scale structure of the image, we have to suppress those irrelevant details. This is most easily done in the Fourier domain by suppressing 'high' spatial frequences: when interested in an inner scale of order σ we need a cut off frequency of order $\omega(\sigma) \approx 1/\sigma$. The question then arises of how to do the cut off. Danielsson and Seger[21] use 'constant plateau' filters, but they use the presumption that the sampled signal is *bandlimited* and ignore the scale degree of freedom. From our operational point of view, bandlimitedness is irrelevant and without this presumption, as Witkin[8] and Koenderink[7] showed, there is essentially only one sensible way to do it. Their derivation relies on an assumption that can be phrased as 'prohibition of spurious detail', the interpretation of which has led to some confusion in the literature. We will not use this argument, but show that the simple front-end symmetries we proposed have exactly the same consequences, and we will emphasize on how they may set the stage for local image analysis in D dimensions.

Natural scaling operator in scale-space

In this section we will derive the unique scaling strategy

for D-dimensional images ($D>1$), using its semigroup nature in combination with the front-end vision symmetries.

Linear shift invariance implies that a rescaled image must be a convolution of the original image by some kernel $G(\boldsymbol{x};\sigma)$*, so:

$$L(\boldsymbol{x};\sigma)=\{L_0 * G(.;\sigma)\}(\boldsymbol{x};\sigma) \tag{2}$$

It is especially attractive to consider this property in the Fourier domain, in which the kernel becomes diagonal: (2) then becomes an algebraic relation:

$$\mathcal{L}(\boldsymbol{\omega};\sigma)=\mathcal{L}_0(\boldsymbol{\omega})\mathcal{G}(\boldsymbol{\omega};\sigma) \tag{3}$$

The Pi theorem states that because of conventional scale invariance there are only two independent dimensionless variables in this case. We may take these to be $\mathcal{G}=\mathcal{L}/\mathcal{L}_0$ and $\boldsymbol{\Omega}\stackrel{\text{def}}{=}\sigma\,\boldsymbol{\omega}$. Let us therefore define:

Natural frequency (spatial) coordinates:

Natural frequency (spatial) coordinates are defined as the dimensionless numbers $\boldsymbol{\Omega}(\boldsymbol{X})$ associated with the frequency (spatial) coordinates $\omega(\boldsymbol{x})$ at scale-space level $\sigma>0$ through:

$$\boldsymbol{\Omega}=\sigma\,\boldsymbol{\omega}\quad\left(\boldsymbol{X}=\frac{\boldsymbol{x}}{\sigma}\quad\text{respectively}\right) \tag{4}$$

According to the Pi theorem we may thus write the kernel $\mathcal{G}(\boldsymbol{\omega};\sigma)$ as a function of $\boldsymbol{\Omega}$:

$$\mathcal{G}(\boldsymbol{\omega};\sigma)=\mathcal{L}/\mathcal{L}\stackrel{\text{def}}{=}\mathcal{G}(\boldsymbol{\Omega}) \tag{5}$$

For a *scalar* function, spatial isotropy implies that $\mathcal{G}$ depends only on the magnitude (Euclidean length) of the vector $\boldsymbol{\Omega}$:

$$\mathcal{G}(\Omega)=\mathcal{G}(\boldsymbol{\Omega}) \tag{6}$$

with $\Omega\stackrel{\text{def}}{=}\sqrt{\sum_{i=1}^{D}\Omega_i^2}$.

Let us choose σ to be such that for fixed ω the hypothetical zero-scale limit $\sigma\downarrow 0$ will leave the initial image unscaled, so:

$$\mathcal{G}(\Omega)\to 1\quad\text{as}\ \Omega\downarrow 0 \tag{7}$$

This means that we include the identity as a limiting, zero-scale kernel.

Also, we require the infinite-scale limit $\sigma\to\infty$ to give us a complete spatial averaging of the initial image:

$$\mathcal{G}(\Omega)\downarrow 0\quad\text{as}\ \Omega\to\infty \tag{8}$$

Performing several rescalings in succession should be consistent with performing a single, effective rescaling. More specifically, if σ_1, σ_2 are the scale parameters associated with two rescalings $\mathcal{G}(\Omega_1)$, $\mathcal{G}(\Omega_2)$ respectively, then the concatenation of these should be a rescaling $\mathcal{G}(\Omega_3)$ corresponding to an effective scale parameter $\sigma_3=\sigma_2\oplus\sigma_1$, in which the additive operator '$\oplus$' relates the effective scale parameter σ_3 to the parameters σ_1, σ_2. It is important to note that '$\oplus$' need not coincide with the familiar additive operator '+'. All that is required by consistency is that the set $\{\mathbf{R}_0^+;\oplus\}$ constitutes a *commutative semigroup* isomorphic to the commutative semigroup of image rescalings:

- *Semigroup operation*:
  $\forall\sigma_1,\sigma_2\,\exists\sigma_3\qquad\sigma_3\stackrel{\text{def}}{=}\sigma_2\oplus\sigma_1$
- *Associativity*:
  $\forall\sigma_1,\sigma_2,\sigma_3\qquad\sigma_3\oplus(\sigma_2\oplus\sigma_1)=(\sigma_3\oplus\sigma_2)\oplus\sigma_1$
- *Null element*:
  $\exists\sigma_0\,\forall\sigma\qquad\sigma\oplus\sigma_0=\sigma_0\oplus\sigma=\sigma$
- *Commutativity*:
  $\forall\sigma_1,\sigma_2\qquad\sigma_1\oplus\sigma_2=\sigma_2\oplus\sigma_1$

The consistency requirement that there is a one-to-one correspondence $\gamma(\sigma)\leftrightarrow\sigma$ manifests itself mathematically by the existence of an isomorphism $\{\gamma(\sigma),\circ\}\simeq\{\mathbf{R}_0^+;\oplus\}$, i.e. a one-to-one map between these two semigroups preserving the semigroup structure:

$$\gamma(\sigma)\circ\gamma(\tilde{\sigma}=\gamma(\sigma\oplus\tilde{\sigma}) \tag{9}$$

This isomorphism poses a very strong constraint on the form of the scale-space kernels.

We will now derive an explicit formula for the semigroup operation $\oplus$. On dimensional grounds (manifest scale invariance), any allowable reparametrisation of σ must be *homogeneous*, i.e. it must have the form $\sigma\to\mapsto\lambda\sigma^p$ for some dimensionless parameters $\lambda>0$ and $p\neq 0$. Without loss of generality we may put $\lambda=1$, since it is merely a scaling factor and (9) must be insensitive to the choice of units. So any allowable reparametrisation of the scale-parameter σ can be realized by an automorphism $\mathcal{P}$:

$$\mathcal{P}:\{\mathbf{R}_0^+;\oplus\}\to\{\mathbf{R}_0^+;+\}:\sigma\mapsto\sigma^p \tag{10}$$

Its inverse is given by:

$$\mathcal{P}^{-1}:\{\mathbf{R}_0^+;+\}\to\{\mathbf{R}_0^+;\oplus\}:\sigma\mapsto\sigma^{1/p} \tag{11}$$

If we assume that ordinary addition applies to $\{\mathbf{R}_0^+;+\}$, then the following identity holds (see (9)):

$$\gamma(\sigma)\circ\gamma(\tilde{\sigma})=\gamma(\mathcal{P}^{-1}(\mathcal{P}\sigma+\mathcal{P}\tilde{\sigma})) \tag{12}$$

Note, however, that (12) still makes sense in the limiting case $p\to\pm\infty$, for which (10) and (11) by themselves have no meaning. It is easy to see that this singular case corresponds to the singular *idempotent semigroups* $\{\mathbf{R}_0^+;\max\}$ and $\{\mathbf{R}_0^+;\min\}$ defined by:

$$\sigma_1\oplus\sigma_2\stackrel{\text{def}}{=}\max(\sigma_1,\sigma_2)\quad\forall\sigma_1,\sigma_2\in\mathbf{R}_0^+ \tag{13}$$

and:

$$\sigma_1\oplus\sigma_2\stackrel{\text{def}}{=}\min(\sigma_1,\sigma_2)\quad\forall\sigma_1,\sigma_2\in\mathbf{R}_0^+ \tag{14}$$

respectively, which emerge as limiting cases from the

*The vector notation for $\boldsymbol{x}$ and $\boldsymbol{\omega}$ is legitimate, once having chosen an arbitrary, fiducial origin for reference.

regular *monotonic semigroup* defined by:

$$\sigma_1 \oplus \sigma_2 \stackrel{\text{def}}{=} \sqrt[p]{\sigma_1^p + \sigma_2^p} \tag{15}$$

We stress the fact that the null elements of these semigroups arise from *non-physical* limiting procedures: for $p<0$ we have $\sigma_0=\infty$, whereas for $p>0$ we have $\sigma_0=0$. Note that we have already decided on the null element $\sigma_0=0$ by our limiting requirement (7), hence only positive p-values will be of interest ot us*.

We now turn to the derivation of scale-space kernels compatible with (12). It is convenient to consider the frequency representation; if we define:

$$\mathcal{G}(\Omega) \stackrel{\text{def}}{=} \tilde{\mathcal{G}}(\Omega^p) \tag{16}$$

we get from (12):

$$\tilde{\mathcal{G}}(\Omega)\tilde{\mathcal{G}}(\tilde{\Omega}) = \tilde{\mathcal{G}}(\Omega+\tilde{\Omega}) \tag{17}$$

The general solution to this constraint is a normalized exponential function:

$$\tilde{\mathcal{G}}(\Omega) = \exp(\alpha\Omega) \tag{18}$$

or:

$$\mathcal{G}(\Omega) = \exp(\alpha\Omega^p) \tag{19}$$

in which α is an arbitrary, negative constant (see (8)), whose absolute value can be absorbed into the definition of the scale parameter.

For the limiting case we have $\mathcal{G}(\Omega) = \lim_{p\to\infty} \tilde{\mathcal{G}}(\Omega^p)$, i.e. $\mathcal{G}(0)$ if $0\leqslant\Omega<1$ and $\tilde{\mathcal{G}}(\infty)$ if $\Omega>1$. Together with the limiting conditions (7) and (8) (and taking $\mathcal{G}(1) \stackrel{\text{def}}{=} \lim_{\Omega\uparrow 1} \mathcal{G}(\Omega)$ for definiteness), we thus find the following idempotent kernel:

$$\mathcal{G}(\Omega) = \chi_{[0,1]}(\Omega) \tag{20}$$

in which χ_I is the indicator function defined by:

$$\chi_I(x) \stackrel{\text{def}}{=} \begin{cases} 1 & \text{if } x \in I \\ 0 & \text{if } x \notin i \end{cases} \tag{21}$$

In dimensionful coordinates this becomes:

$$\mathcal{G}(\omega;\sigma) = \chi_{[0,1/\sigma]}(\omega) \tag{22}$$

i.e. an ideal low-pass filter with cut-off frequency $\omega = 1/\sigma$. In his article, Mallat proposes such an idempotent semigroup requirement as a starting point for a so-called 'multiresolution approximation'[6]; the operator which approximates a given signal at a resolution σ is a linear projection, satisfying (9) and (13).

The general, regular case comprises a so-called (strongly) *continuous semigroup* of operators for each value of p, as opposed to the idempotent semigroup (22):

$$\mathcal{G}(\omega;\sigma) = \exp(\alpha\sigma^p\omega^p) \tag{23}$$

*This is not a restriction: a reflection $p \leftrightarrow -p$ merely amounts to an interchange of the complementary concepts of scale and resolution, i.e. inverse scale. The reader may verify that by expressing all physical requirements in terms of resolution will yield the same result.

To single out a unique scale-space kernel, we need a final constraint on the parameter p. For a consistent interpretation of $\mathcal{G}(\Omega)$ as a spatial rescaling it is natural to impose the condition of *separability* in $D>1$ dimensions:

$$\mathcal{G}(\Omega) = \prod_{i=1}^{D} \mathcal{G}(\Omega_i) \tag{24}$$

in which Ω_i is given by the magnitude of the projection vector $(\Omega\cdot\hat{e}_i)\hat{e}_i$. This condition states that an isotropic rescaling can be obtained either directly through $\mathcal{G}(\Omega)$ or through a concatenation of rescalings $\mathcal{G}(\Omega_i)$ by the same amount in each of the *independent* spatial directions e_i, $i=1\ldots D$ separately. Indeed, only in this way we can think of σ as a natural length unit in an isotropic space. The separability requirement fixes $p=2$, so $s \stackrel{\text{def}}{=} \sigma^2$, not σ itself, is the 'additive' parameter:

$$\sigma \oplus \tilde{\sigma} = \sqrt{\sigma^2+\tilde{\sigma}^2} \tag{25}$$

Note that the idempotent kernel (20) is not separable. A convenient choice for α is obtained by letting scale coincide with Gaussian width in the spatial domain, so that $\alpha = -1/2$.

So we have finally established the unique scale-space kernel. In the Fourier domain it is given by:

$$\mathcal{G}(\Omega) = \exp\left(-\frac{1}{2}\Omega^2\right) \tag{26}$$

or, in dimensionful coordinates:

$$\mathcal{G}(\omega;\sigma) = \exp\left(-\frac{1}{2}\sigma^2\omega^2\right) \tag{27}$$

In the spatial domain it is given by the normalized convolution kernel:

$$G(x;\sigma) = \frac{1}{\sqrt{2\pi\sigma^{2D}}} \exp\left(-\frac{x^2}{2\sigma^2}\right) \tag{28}$$

Note that in the spatial domain the Gaussian kernel $G(\boldsymbol{x};\sigma)$ has a scale dependent amplitude. It has the dimension of an inverse D-dimensional volume: we may write it as a product of an explicit volume factor and a dimensionless, scaled Gaussian:

$$\mathrm{G}(x;\sigma) = \frac{1}{V(\sigma)} G(X) \tag{29}$$

with:

$$G(X) = \frac{1}{\sqrt{2\pi^D}} \exp\left(-\frac{1}{2}X^2\right) \tag{30}$$

Therefore:

$$d^D\boldsymbol{X}\,G(X) = d^D\boldsymbol{x}\,G(x;\sigma) \tag{31}$$

is a scale invariant measure.

The prefix *semi* in 'semigroup' expresses the intrinsically irreversible nature of rescaling. Put differently, rescaling gives rise to *irreversible catastrophes* in the

topological structure of the original image. In 'forward direction', i.e. when increasing scale, there are no 'acausal' bifurcations (no creation of spurious detail).

It is important to stress that filtering a given image with $G(\boldsymbol{x}; \sigma)$ does *not* yield an image with an inner scale σ, but with an inner scale $\sigma \oplus \sigma_0$, if σ_0 is the inner scale of the original image. Each layer $L(\boldsymbol{x}; \sigma)$ in scale-space can in turn be regarded as an initial image with an inner scale equal to $\sigma \oplus \sigma_0$, and it is only when $\sigma \gg \sigma_0$ that the inner scale of $L(\boldsymbol{x}; \sigma)$ approximately equals σ. This observation is especially important if one is interested in scales only slightly larger than one pixel, for we can at most associate some 'effective' inner scale σ_0 with the originally sampled image, but we cannot expect the scale-space requirements to be nicely fulfilled near pixel scale[4].

Having defined natural length units it seems rather trivial to remark that we now have a *natural distance measure* for the separation of two points $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ on a given level σ:

$$\Delta X \stackrel{\text{def}}{=} \| \boldsymbol{X}_2 - \boldsymbol{X}_1 \| \tag{32}$$

Note its singularity at the highest (fictitious) resolution $\sigma = 0$. When viewed with an infinite resolution, two distinct points are always 'infinitely' far apart, since there can be an arbitrarily large amount of structure inbetween.

Significant changes due to rescaling will occur only when we increase scale by an *order of magnitude* rather than by some absolute amount. Hence it is more natural to reparametrize our scale parameter, thus removing the artificial singularity at $\sigma = 0$:

Natural scale parameter:

A natural, dimensionless scale parameter τ is obtained by the following reparametrization of σ:

$$\sigma = \epsilon \exp\{\tau\} \quad or \quad \tau = \ln\{\sigma/\epsilon\} \quad \tau \in (-\infty, +\infty) \tag{33}$$

Note that we are forced to introduce, on dimensional grounds, a 'hidden scale' ϵ, which carries the dimension of a length. It is a property of the image, not of the universal scale-space kernel. An intrinsic scale inherent to any imaging device that presents itself is the sampling width or pixel width. Now we have a dimensionless scale parameter τ that indicates in a continuous manner the order of magnitude of scale relative to ϵ and that can take on, at least in theory, any real value. If we take ϵ to be the sampling width, then $\tau = 0$ corresponds to a resolution, where the width of the blurring kernel is of the same order of magnitude as the pixel width ϵ, i.e. the inner scale of the original image. This sets a practical lower limit to the kernel widths, at which discretization effects will start to contribute to a significant degree. The range $\tau \in (-\infty, 0)$ corresponds to subpixel scales that are not represented in the image and in which all structure has been averaged out. When building up a scale-space it is most natural to use an equidistant sampling of τ, because it is this parameter that precisely formalizes the physical notion of scale. An equidistant sampling of absolute scale σ would violate scale-invariance.

In this section we have derived the unique *scalar* scale-space kernel that satisfies all our front-end vision symmetries, as well as some additional constraints, noticeably the concatenation or semigroup requirement (9) and the separability condition (24). But it is important to stress that the very assumption that it has to be a scalar subject to these scaling properties has been explicitly added by our desire to find a filter that merely scales its input, but is *not* part of our fundamental front-end vision requirements. Indeed, it is only by virtue of these extra constraints that we were able to single out the Gaussian as the unique solution.

In the next section we show that the Gaussian scale-space kernel is merely the lowest order member of a complete, hierarchically ordered family of scale-space filters, all of which are compatible with our front-end vision requirements.

Complete hierarchical family of higher order operators

Although in principle the one-parameter Gaussian kernel is all one needs to generate a scale-space, it is highly insufficient for a *complete, local* description of image structure. In fact, this filter is the physical counterpart of the trivial mathematical identity operator in the sense that it extracts a *scaled copy* of a given input, representing the same scene merely on a different inner scale.

In this section we show that the front-end vision requirements sec admit many more scale-space operations beyond mere scaling. We derive a *complete, hierarchically ordered family* of *n-th order tensorial scale-space filters* $\{\gamma_{i_1 \ldots i_n}(\sigma)\}_{n=0}^{\infty}$ (in both spatial and Fourier representation), and discuss their role in front-end image analysis. The previously established Gaussian scale-space kernel naturally fits into this family as just the zeroth order, *scalar* member.

Since the kernels are diagonal in the Fourier domain, it is easiest to consider their Fourier representations. It is a common misconception to think that rotational invariance of the kernels implies that they only depend on the length $\|\boldsymbol{\omega}\|$ of the vector $\boldsymbol{\omega}$. This only holds for *scalar* kernels. It is easy to construct other, *tensorial* kernels within the isotropy constraint. In fact, any tensorial kernel must be proportional to a tensor product containing n factors $\boldsymbol{\omega}$, with $n = 0, 1, 2 \ldots$, since $\boldsymbol{\omega}$ is the only independent vector available. The proportionality constant must be a scalar. Putting in a scalar multiplier $\mathcal{G}(\boldsymbol{\omega}; \sigma)$ to account for proper scale fixing we can formulate the following claim:

Claim 1 *A complete, hierarchically ordered family of multiplicative scale-space kernels is given in the Fourier representation by the set*:

$$\{\mathcal{G}_{i_1 \ldots i_n}(\boldsymbol{\omega}; \sigma) = i\omega_{i_1} \ldots i\omega_{i_n} \mathcal{G}(\boldsymbol{\omega}; \sigma)\}_{n=0}^{\infty} \tag{34}$$

Alternatively, in the spatial representation, by the set of convolution filters:

$$\{G_{i_1 \ldots i_n}(\boldsymbol{x}; \sigma) = \partial_{i_1 \ldots i_n} G(\boldsymbol{x}; \sigma)\}_{n=0}^{\infty} \tag{35}$$

Note that the zeroth order kernel G is the only scalar kernel. All higher order kernels are tensorial quantities. For example, the first order kernel, i.e. the *gradient*, is a vector.

The proof of this claim is given below, where we

show that this *cartesian family* is sufficient for a complete determination of local image structure. Consider a given image $L(\boldsymbol{x}; \sigma)$ at a fixed scale. If we are interested in the geometrical structure of this image in the neighbourhood of some fixed point $\boldsymbol{x} \in \mathbf{R}^D$ we may consider its Taylor approximation up to a sufficient order N:

$$L(\boldsymbol{x}+\delta\boldsymbol{x}; \sigma) = \sum_{n=0}^{N} \frac{1}{n!} L_{i_1 \ldots i_n}(\boldsymbol{x}; \sigma)\, \delta x_{i_1} \ldots \delta x_{i_n} + \mathcal{O}\left(\left(\frac{\delta \boldsymbol{x}}{\sigma}\right)^{N+1}\right) \quad (36)$$

Note that a scaled image $L(.; \sigma) \stackrel{\text{def}}{=} L_0 * G(.; \sigma)$ is a *smooth* function for all $\sigma > 0$, no matter how 'dirty' the initial condition L_0 may be (within certain very weak restrictions). Digitized images are, by the very fact of being digitized, always 'dirty' in the differential geometric sense (even in the absence of noise). In theory, L_0 may even be everywhere discontinuous. Scaled differentiation, as opposed to ordinary ('unscaled') differentiation, is well-posed by nature. Note the following identity:

$$L_{i_1 \ldots i_n}(\boldsymbol{x}; \sigma) = L_0 * G_{i_1 \ldots i_n}(\boldsymbol{x}; \sigma) \quad (37)$$

In other words, we have obtained the following important result:

Result 1 *To obtain the Cartesian partial derivatives of order n of a* rescaled *image* $L(\boldsymbol{x}; \sigma)$ *one only needs to convolve the* original *image* $L_0(\boldsymbol{x})$ *with the corresponding partial derivatives of the zeroth-order Gaussian* $G(\boldsymbol{x}; \sigma)$.

The calculation of derivatives is most easily done in the Fourier domain, in which the filters are diagonal:

$$L_{i_1 \ldots i_n} = \mathcal{F}^{-1}[\mathcal{L}_0 \mathcal{G}_{i_1 \ldots i_n}] \quad (38)$$

Since (36) represents the image's local geometry at $\boldsymbol{x}$ and at scale σ up to any desired order of precision N, we have proven the completeness of the constructed Cartesian kernel family. Each essential kernel component in the family corresponds to an independent degree of freedom. Thus is is also a minimal set. The zeroth order member of the Cartesian family represents the scaled identity operator, the higher order members constitute the physical, scaled counterpart of a *complete family of mathematical linear differential operators*. In Figure 1 the spatial profiles of a number of operators are shown.

An alternative, but equivalent way of looking at the completeness of this set of filters has been given by J. Koenderink and A. Van Doorn, who took the isotropic diffusion equation as a fundamental starting point for the derivation of the complete family of scale-space filters or *local neighbourhood operators*[13,20], since this equation uniquely prohibits the generation of 'spurious detail' in scale-space[7].

We end this section by noting that, although *in theory* nth order derivatives of a *scaled* image are all well-defined, there is only one operational way of calculating them, *viz*, by a convolution of a lower scale image with their corresponding tensor components $G_{i_1 \ldots i_n}(\boldsymbol{x}; \sigma)$ (cf. (37)). This brings us to another important result:

Result 2 *The operations of scaling and differentiation are intrinsically related.*

Because of the one-to-one correspondence between the Gaussian kernels $\gamma_{i_1 \ldots i_n}(\sigma)$ (on a fixed scale: 'horizontal image structure') and the Cartesian partial differential operators, it is straightforward to invoke the powerful machinery of well-established mathematical disciplines, like *differential geometry*, *tensor calculus* and *invariants theory* in a robust way. This enables us to study the visual system as a 'geometry engine'[22].

Differential invariants in scale-space

Once we have calculated the N-jet, we are provided with all partial derivatives of the image up to and including Nth order. However, one such derivative, L_x say, does not represent any geometrically meaningful property, since the choice of the coordinate axes is completely arbitrary. If we restrict ourselves to an orthonormal basis, we still have the possibility of

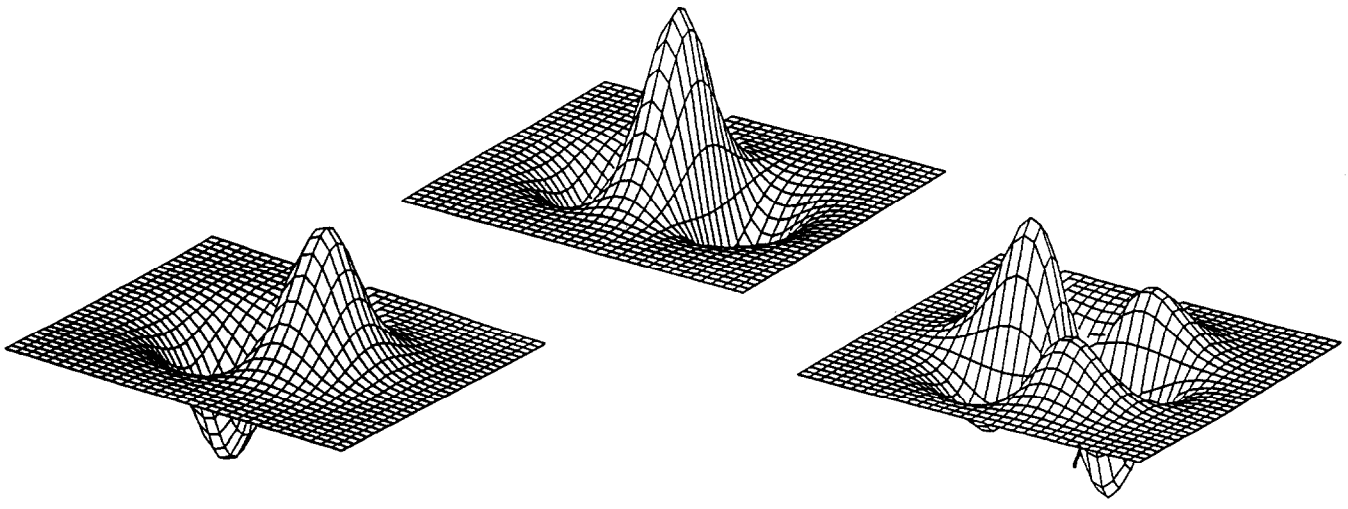

Figure 1. Some Gaussian derivative profiles: G_x, G_{xx} *and* G_{xyy}

rotating a given coordinate frame over any angle. Clearly, such a choice does not have anything to do with the image. On the other hand, it is also clear that any property that is invariant under such a coordinate transformation, must be connected to the image 'itself' and therefore can be given a geometric interpretation. The reverse is also true: every image property can be expressed through an invariant function (or 'invariant', in short). This shows that there is an intimate relation between *invariants theory* and *differential geometry*.

Although it is possible to give a coordinate-free description of image geometry in principle, we do need coordinates in actual calculations. To this end we simply choose any allowable coordinate system, but at the same time assure that the functions of interest are independent of that particular choice. Note that the term 'invariant' always implies the existence of a *group* of allowable transformations. In our case this is a rather 'minimal' group, *viz.* the product group of $SO(D)$, the *special orthogonal group* of all rotations in D dimensions (sometimes extended to $O(D)$, the *full orthogonal group*, by admitting reflections) and $T(D)$, the *translation group*. This is a very basic group, which we believe is especially important in medical imaging.

Those special combinations of image derivatives that exhibit such an invariance under cartesian coordinate transformations are called *cartesian differential invariants*. To get a basic understanding of the theory of cartesian invariants, it is necessary to understand the basics of *cartesian tensor calculus*[23]. In this section we briefly outline how to construct functions describing true image properties.

It turns out that we can in fact construct an infinite number of invariants in each point of the image, but we argue that there is only a small number of independent ones among these. This is to say that, to a given order N, we can build any geometrically meaningful quantity as a function of those (typically very few) independent or *irreducible invariants*.

Manifest invariant index notation

It is clear that we cannot form an invariant out of a single derivative like L_x, whose value always crucially relies on the choice of the x-axis and thus varies among coordinate systems. It is the subject of tensor calculus to describe the transformation behaviour of quantities like L_x, called *tensor components*. A 'closed set' of tensor components, although given with respect to some arbitrarily chosen coordinate system, does however constitute a coordinate independent object, called a *tensor*. The meaning of the word 'closed' in this context is that, after a change of coordinates, each tensor component acquires a new value that can be expressed as some function of the old tensor components. This function only depends on the transformation parameters involved, i.e. a set of rotation angles and translation components (such a smoothly parametrized group of transformations is called a *Lie group*[16]). For example, in 2D the partial derivative L_x changes, after a rotation over an angle α, according to the following rule:

$$L_{x'} = \cos\alpha\, L_x + \sin\alpha\, L_y \tag{39}$$

This shows that L_x cannot be the single component of a tensor. We should at least add the component L_y to it. Indeed, this suffices to obtain a 2-component tensor $\{L_x, L_y\}$, since these two components do transform in a closed way. Their transformations can be written in matrix form:

$$\begin{pmatrix} L_{x'} \\ L_{y'} \end{pmatrix} = \begin{pmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{pmatrix} \begin{pmatrix} L_x \\ L_y \end{pmatrix} \tag{40}$$

It is clear that also the coordinates $\{x, y\}$, taken as a pair relative to some fixed origin, constitute a tensor. Let R_{ij} stand for the (i, j)th element of the transformation matrix, then the abovementioned transformation can be conveniently written in condensed form as:

$$L'_i = R_{ij} L_j \tag{41}$$

Because L_i has only one free index it is called a 1-tensor or *vector*. But we can also consider tensors with more free indices (*n-tensors*). An example of a 2-tensor is the *Hessian*, i.e. the set of all second order partial derivatives: its transformation is given by:

$$L'_{ij} = R_{ik} R_{jl} L_{kl} \tag{42}$$

It has exactly three *essential components*, *viz.* L_{xx}, $L_{xy} = L_{yx}$, L_{yy}. This means that because of the symmetry of the tensor, we cannot choose all its components independently.

By now it may be obvious that all partial derivatives of a given order n form the components of an n-tensor. For each of its free indices its transformation law contains a transformation matrix with one free and one contracted index:

$$L'_{i_1 \ldots i_n} = R_{i_1 j_1} \ldots R_{i_n j_n} L_{j_1 \ldots j_n} \tag{43}$$

These derivative tensors share the additional property of being *symmetric*, i.e. we can freely inter-change indices without any effect, e.g. $L_{ij} = L_{ji}$. Thus there is a significant reduction of essential components.

Of great importance are the following two *constant tensors*: the symmetric *Kronecker tensor* δ_{ij}, which is always a 2-tensor, and the antisymmetric *Lévi-Civita* tensor $\varepsilon_{i_1 \ldots i_D}$, which is a D-tensor in D dimensions. These tensors have invariant components, independent of the choice of the coordinate axes (this property makes them well-defined). They are defined as follows:

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases} \quad \text{and}$$

$$\varepsilon_{i_1 \ldots i_D} = \begin{cases} 1 & \text{if } (i_1 \ldots i_D) \text{ is even} \\ -1 & \text{if } (i_1 \ldots i_D) \text{ is odd} \\ 0 & \text{otherwise} \end{cases} \tag{44}$$

When including reflections into the transformation group, the ε-'tensor' is not a true tensor in the above sense anymore. Its significance still remains as a so called *relative* or *pseudo-tensor*. Its transformation law is then slightly modified so as to render its components invariant again:

$$\varepsilon'_{i_1 \ldots i_D} \stackrel{\text{def}}{=} (\det R)^{-1} R_{i_1 j_1} \ldots R_{i_D j_D} \varepsilon_{j_1 \ldots j_D} = \varepsilon_{i_1 \ldots i_D} \quad \text{with} \quad \det R = \pm 1 \tag{45}$$

Once we understand the transformation behaviour of the derivatives, we can try and combine them into (absolute or pseudo-) invariant combinations. This is, in fact, very easy. Given a set of tensors, the way to form an invariant is by means of *full contractions and alternations of indices* in a tensor product. A contraction is a procedure that pairwise reduces the number of free indices in a tensor by performing a restricted summation over them. More precisely, a contraction of i, j in L_{ij} by definition yields $L_{ii} = L_{ij}\delta_{ji}$ (this contraction is also referred to as the *trace* of L_{ij}).

An alternation of D tensor indices is defined as a full contraction of these indices onto the D indices of the ε-tensor as, for example, (in 2D) in:

$$\varepsilon_{ij}L_iL_kL_{jk} = (L_x^2 - L_y^2)\,L_{xy} + L_xL_y(L_{yy} - L_{xx}) \quad (46)$$

Of course, functions of several invariants are themselves invariants.

If we consider reflections as well as rotations (after all, these also respect orthonormality of the coordinate basis), then the ε-tensor becomes a relative tensor and invariants containing an odd number of these become *relative invariants*, i.e. quantities that are invariant up to a possible minus sign (which shows up only when the orientation of the coordinate basis is reversed). In fact, we can always write a relative invariant using exactly one ε-factor. This follows from the fact that any tensor product of an even number of ε-tensors can be written in terms of δ-tensors. In 2D:

$$\varepsilon_{ij}\varepsilon_{kl} = \delta_{ik}\delta_{jl} - \delta_{il}\delta_{jk} \quad (47)$$

Similar relations hold in arbitrary dimensions.

Relative invariants are related to *oriented* geometrical objects. We will often speak of invariants, but silently admit relative invariants, too. The term *absolute invariant* is then used to explicitly exclude relative ones.

Gauge coordinates

In the index notation one refrains from choosing any particular coordinate frame. The invariance of a function then manifests itself through a full contraction of indices in the tensor products that make up the function. For this reason we speak of *manifest invariance* when using this notation.

Another way of forming manifest invariants is by singling out one particular, geometrically meaningful coordinate frame and using directional derivatives along its axes. There are several ways to set up such a coordinate frame. One useful way is to require, in each point of the image separately, one axis to coincide with the image gradient direction (called the w-axis henceforth). The other axis (v-axis) is then automatically directed tangentially along the isophote. By its very definition, L_v vanishes identically. This is precisely the motivation for this particular gauge. Because of the rotation freedom, this kind of requirement is always allowed, provided the image gradient does not vanish*.

We call such a requirement a *gauge condition*, and the resulting coordinates *gauge coordinates*. It should always be checked whether a gauge condition is *admissible*, i.e. realizable through a suitable transformation provided by the transformation group at hand.

The directional derivative operators applied to the image will yield invariants. Invariance becomes manifest by writing differential invariants using these invariant differential operators.

To illustrate the use of differential geometry and at the same time show the power of gauge coordinates that are tuned to a particular problem, let us derive an expression for the *isophote*† *curvature* κ (cf. Clark[24]). The meaning of curvature of a planar curve may be intuitively clear; it is a measure for the local deviation from its tangent line[25]. A useful definition is the following one: put a coordinate frame with its origin in the point P of interest on the curve. The x-axis should be tangent to the curve. The curve can then locally be described by a function $y(x)$ on an open interval around $x = 0$. In this system the curvature in the origin is defined as the second derivative $y''(0)$. So in the (v, w)-system centred at P we have, by definition, $\kappa = w''(0)$, in which $w(v)$ denotes the function describing the isophote locally near $P(v = 0,\ w = 0)$. Now the isophote passing through the point P is implicitly given by the equation $L = L_P$. Taking first and second implicit derivatives of this equation with respect to v yields:

$$L_v + L_w w' = 0 \quad \text{and}$$

$$L_{vv} + 2L_{vw}w' + L_{ww}w'^2 + L_w w'' = 0 \quad (48)$$

In P we have, by our suitable choice of gauge, $L_v(0) = 0$, hence also $w'(0) = 0$. So in P we have $\kappa \overset{\text{def}}{=} w''(0) = -L_{vv}/L_w$. In a similar way, one may proof that the *flow line curvature* μ, i.e. the curvature of the integral curves of the gradient vector field (the orthogonal trajectories of the isophotes), is given by the formula $\mu = -L_{vw}/L_w$. So we have the following result:

$$\kappa = -\frac{L_{vv}}{L_w} \quad \text{and} \quad \mu = -\frac{L_{vw}}{L_w} \quad (49)$$

Because these invariants are closely related to simple isophote properties, they look simplest when written in this particular gauge.

It may come as a surprise to learn that, although we have calculated the isophote curvature in a *simplifying* coordinate system, it takes hardly any effort to arrive at the general expression in *arbitrary* coordinate systems. The method goes as follows: write down an invariant in manifest index notation that reduces to the simplified expression evaluated in gauge coordinates (simply guessing in combination with a modest amount of foresight usually does the trick in one go). By invariance, the two expressions are guaranteed to represent the same geometrical property. The following index notation for κ and μ (49) can be easily justified this way:

$$\kappa = \frac{L_i\varepsilon_{ij}L_{jk}\varepsilon_{kl}L_l}{(L_mL_m)^{3/2}} \quad \text{and} \quad \mu = \frac{L_i\varepsilon_{ij}L_{jk}\delta_{kl}L_l}{(L_mL_m)^{3/2}} \quad (50)$$

*The (v, w)-gauge is ill-defined in points with a vanishing gradient, but these points form a countable set, at least in *generic images*, images that are topologically stable against local distortions. Blurring a nontrivial image to a certain level of resolution always yields a generic image (which may, however, become obscured in a computer implementation by truncations due to the finite precision of L-values).

†An isophote is a contour of constant image values.

To see how this works, consider the isophote curvature expression in (50). Its denominator is the third power of the gradient magnitude, i.e. $L_w^3 = (L_m L_m)^{3/2}$, while its numerator can be translated from index into gauge notation as follows:

$$\kappa L_w^3 = L_i \varepsilon_{ij} L_{jk} \varepsilon_{kl} \quad \text{(arbitrary system)} = -L_{vv} L_w^2 \quad \text{(gauge system)} \quad (51)$$

(the last equality follows from $L_v = 0$, $\varepsilon_{vw} = -\varepsilon_{wv} = 1$ and $\varepsilon_{vv} = \varepsilon_{ww} = 0$, so that the only non-trivial term is the one with indices $i = l = w$ and $j = k = v$). Similar arguments hold for μ in (50). Evaluating the contractions in some Cartesian coordinate system will give us the explicit formulas for κ and μ in (50):

$$\kappa = \frac{2L_x L_y L_{xy} - L_x^2 L_{yy} - L_y^2 L_{xx}}{(L_x^2 + L_y^2)^{3/2}} \quad (52)$$

$$\mu = \frac{(L_x^2 - L_y^2)\, L_{xy} + L_x L_y (L_{yy} - L_{xx})}{(L_x^2 + L_y^2)^{3/2}} \quad (53)$$

It is clear that this explicit notation obscures the Cartesian invariance property and can become very cumbersome when there are many contractions to be performed.

Another example of a manifest invariant is given by the well-known Laplacean:

$$\Delta L = L_{ii} = L_{vv} + L_{ww} = L_{ww} - \kappa L_w \quad (54)$$

in which we have used the expression for isophote curvature (49). This example shows that, in general, invariants can be interrelated. More specifically, the (v, w)-gauge nicely reveals the shortcomings of edge detection methods based on Laplacean zero crossings often encountered in the literature[26,27]. The term L_{ww} is the second order image derivative along the gradient direction, i.e. normal to the isophote. If we define an edge as the locus of points of maximum gradient magnitude, which seems a quite natural choice, then the zero crossings of ΔL can only accurately describe edges if the isophotes are sufficiently straight, so that the curvature term can be ignored. It is well-known that this condition ceases to hold near corners and this is one deficiency of this zero crossings method. Another deficiency is the detection of *phantom edges*[24], i.e. non-edge points detected by this zero crossings method. Even if the isophotes are straight, the L_{ww} zero crossings detect not only local maxima of L_w (*true edges*), but also local minima, which are the least likely candidate edge-points of all.

Complete sets of differential invariants

It may be evident that we can construct an infinite number of invariants from any finite set of tensors by means of tensor multiplications and contractions. But it is also clear that the N-jet in a given point only has a finite number of independent degrees of freedom. For example, in 2 dimensions local image structure up to second order is completely determined by five independent 2-jet components, which is 1 less than the number of essential components because of the gauge degree of freedom. In the (v, w)-gauge in which $L_v = 0$, these correspond to the set $\{L, L_w, L_{vv}, L_{vw}, L_{ww}\}$. Therefore, we might foresee the existence of a finite number of so called *irreducible polynomial invariants*, i.e. a set of basic polynomial invariants in terms of which all other invariants can be expressed. However plausible this argument may seem the proof of it in the general case is far from trivial. A proof of existence was established by Hilbert, although the mathematical literature does not seem to provide an algorithm for the actual construction of such an irreducible set. In the simple case of the 2-jet, however, such an irreducible set can readily be given[28]. In 2D:

$$\mathcal{I} = \{L, L_i L_i, L_i L_{ij} L_j, L_{ii}, L_{ij} L_{ji}\} \quad (55)$$

An example of reducibility is given by the following identity:

$$L_{ij} L_{jk} L_{ki} = \left(\frac{3}{2} L_{ij} L_{ji} - \frac{1}{2} L_{ii} L_{jj}\right) L_{kk} \quad (56)$$

dwhich can be verified most easily in the (p, q)-gauge, defined by the gauge condition $L_{pq} = 0$, i.e. the coordinate system in which the Hessian matrix of all second order derivatives is diagonal. This gauge is admissible, since it can always be realized by a suitable rotation. Because of invariance, this reducibility property holds in arbitrary coordinate systems.

In the next section we will take one or two examples for each of the lowest order jets ($N = 0 \ldots 3$). For the sake of presentation we only consider the 2D case, but it must be stressed that the dimensionality does not pose a fundamental restriction to the concepts introduced. In fact, in much of the previous theory we refrained from specifying the dimension of space explicitly wherever this was irrelevant.

APPLICATIONS

A trivial 0-jet example of a differential invariant is L, the local image intensity (in an implicitly given point P and on an implicitly given scale σ). A simple 1-jet example is $\sqrt{L_i L_i} = \|\nabla L\|$, the image gradient magnitude. It is most pronounced on edges, where there is a relatively strong change of intensity values over a relatively short distance. Note that this is just the Canny edge detector[29] (see also De Micheli *et al.*[30]). A simple 2-jet example is the familiar Laplacean $L_{ii} = \Delta L$. Figure 2 shows some differential invariants as they were calculated for noisy test images on several scales.

We already pointed out that there are basically only two independent, pure second order 'irreducibles', which can be taken as L_{ii} and $L_{ij} L_{ji}$ (55). Any other pure second order property can be expressed as some combination of these two. From a geometrical point of view this is clear, since a second order image property is always related to its 'deviation from flatness'. The local image intensity profile can deviate from its first order behaviour in two directions independently. There are two *principal directions*, corresponding to the coordinate axes of a system in which the mixed derivatives vanish (the (p, q)-gauge). The invariants L_{pp} and L_{qq} can then be regarded as measures for the deviation of flatness in these principal directions. This way the

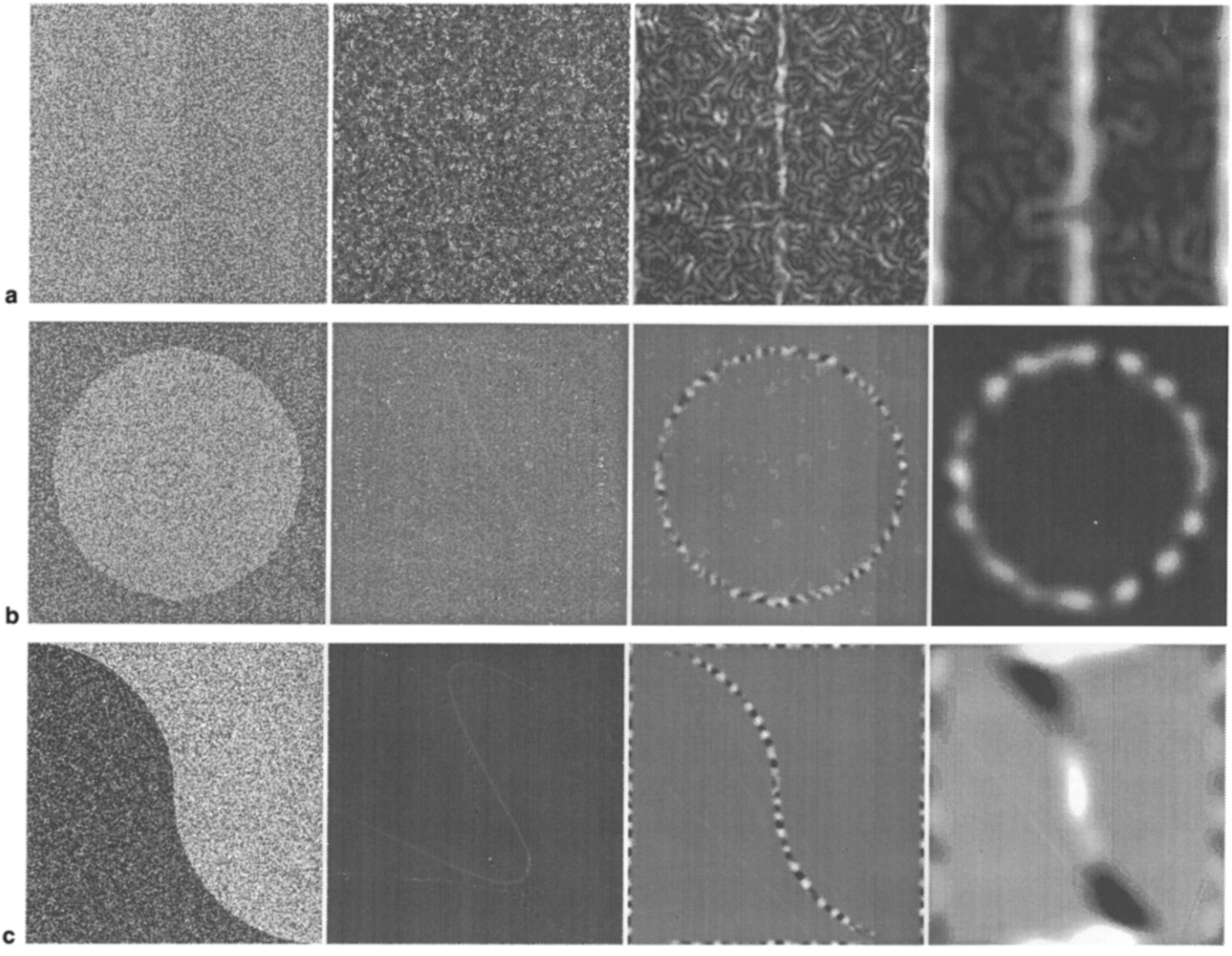

Figure 2. Some simple examples of invariants calculated for noisy test images on various scales. The additive gaussian noise imposed serves to illustrate the robustness of scale-space differentiation. (a) 1st order invariant L_w *for noisy straight edge: 'edgeness'; (b) 2nd order invariant* $L_{vv}L_w^2$ *for noisy polygon (16 corners): 'cornerness'; (c) 3rd order invariant* $L_{vvv}L_w^5 - 3L_{vv}L_{vw}L_w^4$ *for noisy inflexion: 'bendedness'*

Laplacean $L_{ii} = L_{pp} + L_{qq}$ turns out to be twice the mean deviation from flatness, whereas the square root of $L_{ij}L_{ji} = L_{pp}^2 + L_{qq}^2$ is an absolute measure for the total deviation from flatness[13].

Another geometrical property closely related to this 'deviation from flatness' property is the notion of light and dark 'blobs' in the image. These 'blobs' can be given an exact meaning by looking at the sign of the following invariant, called *umbilicity* (U):

$$U = \frac{\varepsilon_{ij}\varepsilon_{kl}L_{ik}L_{jl}}{L_{mn}L_{nm}} = \frac{2L_{pp}L_{qq}}{L_{pp}^2 + L_{qq}^2} \tag{57}$$

Note that we have normalized U such that $-1 \leqslant U \leqslant +1$. Dark and light blobs (or 'hills' and 'dales') now correspond to patches with equally signed principal deviations, i.e. with positive U, whereas the complementary, indifferent ('saddle-like') patches have negative U. The blobs are separated from the indifferent regions by the zero-crossings of U. We can single out the light blobs by looking at the sign of the Laplacean in addition: blobs with negative (positive) Laplacean are light (dark) blobs.

Figure 3 shows the light blobs of a NMR image on various scales. The complementary dark blobs and indifferent patches have been suppressed.

The ordered light blob patches are reasonable primitives for a *fixed-scale segmentation*, and an across-scale linkage algorithm might be set up using these ordered fixed-scale segments to define linkage criteria for a more realistic, multiscale image segmentation.

DISCUSSION AND CONCLUSIONS

In this paper we have shown that the fundamental motivation for the construction of a scale-space is given by the physical nature of images and the universal law of scale invariance. Constraints arising from the lack of *a priori* geometrical knowledge naturally lead to the Gaussian kernel and its derivatives in $D > 1$ dimensions. The operations of scaling and differentiation are essentially intimately related. The study of invariants under a certain group of image transformations gives a robust mathematical basis for the study of image structure.

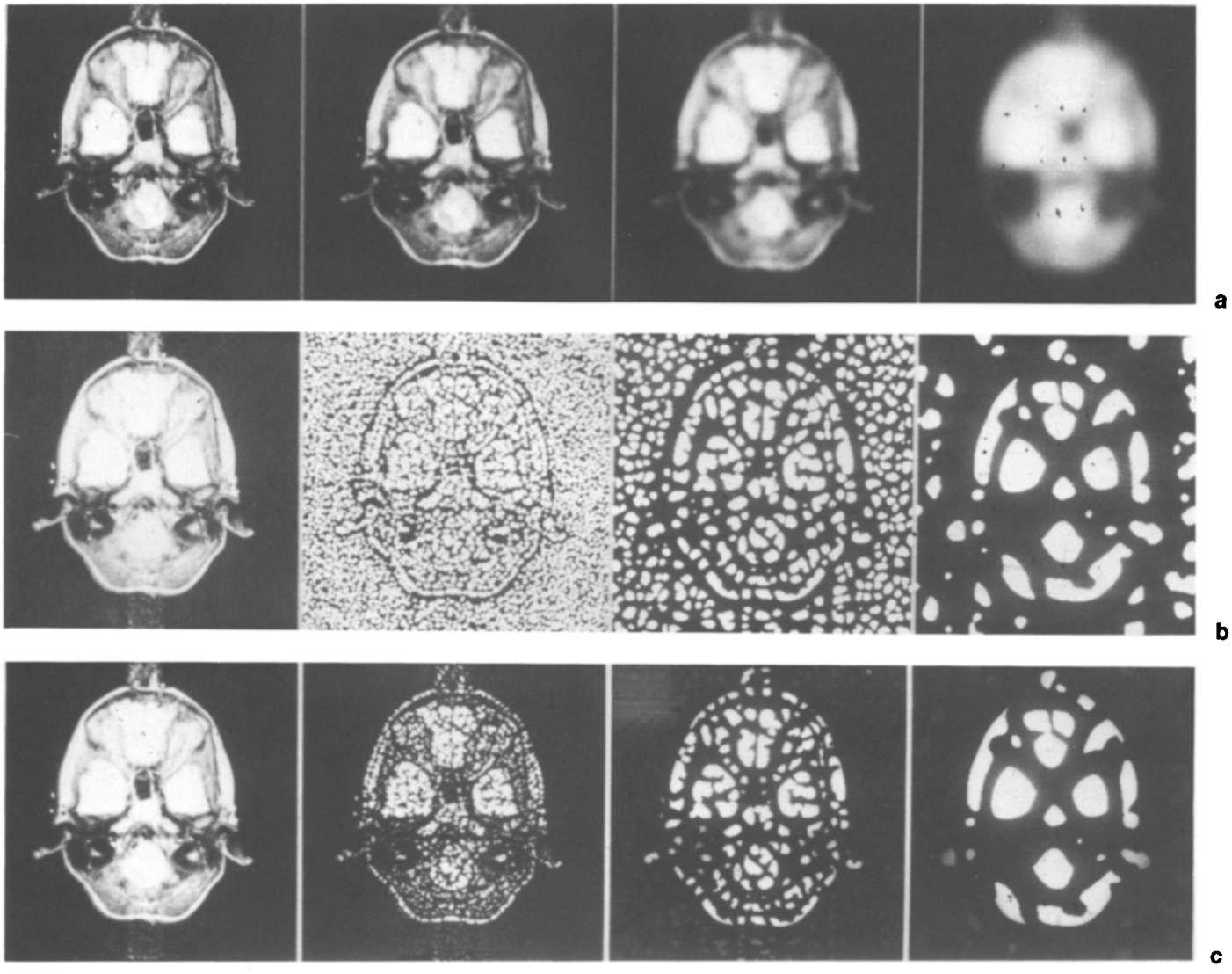

Figure 3. Patch classification based on 2-jet differential invariants calculated for an NMR image on various scales. (a) Scaled original image; (b) light blobs, i.e. patches with positive umbilicity and negative Laplacean; note that these binary invariants are by their very nature unstable in (nearly) flat regions, since a small perturbation δL *of* L *may easily cause them to flip. This is seen to occur in particular in the flat background; (c) as (b) but now the patches have been weighted by the zeroth order images from (a) (on corresponding scales) so as to obtain continuous invariants again. This trick preserves the patches and may be used to obtain a hierarchically labelled patch classification by assigning a priority number to each patch corresponding to the relative ranking of its average intensity value (this ranking is not shown here). The patch with the highest average intensity value will then be the 'most pronounced blob' at that scale, etc. It is clear that the unstable patches surrounding the skull will acquire a low priority label in this way*

The theory allows for the good understanding of many current available feature detection mechanisms, e.g. the Canny edge detector, Laplacean zero crossings, isophote curvature, etc., and puts these in the perspective of a broad class of differential invariants up to some order. The theory is applicable in many areas of computer vision.

This theory may be further developed by a more systematic study of the irreducible invariant 'building blocks' up to any order, by inclusion of the temporal domain[31], stereo, optic flow, and also by studying the 'deep structure' in scale-space (i.e. the structure across scales), incorporating our local theory into a global model. A particularly important, but still unanswered question is also how to operationally gauge the local measurements and how to establish a (multilocal) connection.

The resemblance between the complete family of scale-space kernels and mammalian receptive field profiles known from numerous neurophysiological data is encouraging: it suggests that our theory of differential invariants in scale-space is a promising attempt towards a robust simulation of some of the successful geometric routines actually working in the human visual system.

This theory may have an important impact on various topics in medical imaging, notably image segmentation, classification and pattern recognition.

ACKNOWLEDGEMENTS

This work was supported by the Dutch Ministry of Economic Affairs Grant [VS-3DM]-50249/89-01.

REFERENCES

1 **Burt, P J, Hong, T H and Rosenfeld, A** 'Segmentation and estimation of image region properties through cooperative hierarchical computation', *IEEE Trans. Syst. Man. & Cybern.*, Vol 11 No 12 (1981) pp 802–825

2 **Brunnstrom, K, Eklundh, O and Lindeberg, T** 'On scale and resolution in the analysis of local image structure', *1st Euro. Conf. Comput. Vision*, Antibes, France (1990) pp 3–13

3 **Korn, A** 'Toward a symbolic representation of intensity changes in images', *IEEE Trans. PAMI*, Vol 10 No 5 (1988) pp 610–625

4 **Lindeberg, T** 'Scale-space for discrete signals', *IEEE Trans. PAMI*, Vol 12 No 3 (1990) pp 234–245

5 **Babaud, J, Witkin, A P, Baudin, M and Duda, R O** 'Uniqueness of the gaussian kernel for scale-space filtering', *IEEE Trans. PAMI*, Vol 8 No 1 (1986) pp 26–33

6 **Mallat, S G** 'A theory for multiresolution signal decomposition: The wavelet representation', *IEEE Trans. PAMI*, Vol 11 No 7 (1989) pp 674–694

7 **Koenderink, J J** 'The structure of images', *Biol. Cybern.*, Vol 50 (1984) pp 363–370

8 **Witkin, A** 'Scale space filtering', *Proc. Int. Joint Conf. on Artif. Intell.*, Karlsruhe, Germany (1983) pp 1019–1023

9 **Yuille, A L and Poggio, T A** 'Scaling theorems for zero-crossings', *IEEE Trans. PAMI*, Vol 8 (1986) pp 15–25

10 **Poston, T and Steward, I** *Catastrophe Theory and its Applications*, Pitman, London (1978)

11 **Blom, J, ter Haar Romeny, B M and Bel, A** Spatial derivatives and the propagation of noise in Gaussian scale-space. Submitted, May 1991

12 **Jain, A K** *Fundamentals of Digital Image Processing*, Prentice Hall, NJ (1989)

13 **Koenderink, J J and van Doorn, A J** 'Representation of local geometry in the visual system', *Biol. Cybern.*, Vol 55 (1987) pp 367–375

14 **Fourier, J** *The Analytical Theory of Heat*, Dover Publications, New York (1955)

15 **Strutt, L R J W** 'The principal of similitude', *Nature*, Vol XCV (March 1915) pp 66–68, 644

16 **Olver, P J** *Applications of Lie Groups to Differential Equations*, Vol 107 of *Graduate Texts in Mathematics*, Springer-Verlag, Berlin (1986) pp 218–221

17 **Hubel, D H and Wiesel, T N** 'Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex', *J. Physiology*, Vol 160 (1962) pp 106–154

18 **Young, R A** 'The gaussian derivative model for machine vision: visual cortex simulation', Karlsruhe, Germany (1983) pp 1019–1023

19 **Young, R A** 'Simulation of human retinal function with the gaussian derivative model', *Proc. IEEE* CVPR Ch2290–5, Miami, FL (1986) pp 564–569

20 **Koenderink, J J and van Doorn, A J** 'Receptive field families', *Biol. Cybern.*, Vol 63 (1990) pp 291–298

21 **Danielsson, P-E and Seger, O** 'Rotation invariance in gradient and higher order derivative detectors', *Comput. Vision, Graph. & Image Process.*, Vol 49 (1990) pp 198–221

22 **Koenderink, J J** 'The brain a geometry engine', *Psychol. Res.*, Vol 52 (1990) pp 122–127

23 **Kay, D C** *Tensor Calculus*, Schaum's Outline Series, McGraw-Hill, New York (1988)

24 **Clark, J J** 'Authenticating edges produced by zero-crossing algorithms', *IEEE Trans. PAMI*, Vol 11 (1989) pp 43–57

25 **Spivak, M** *A Comprehensive Introduction to Differential Geometry (vol. I-V)*, Publish or Perish Inc., Berkeley, CA (1970)

26 **Marr, D C and Hildreth, E C** 'Theory of edge detection', *Proc. Roy. Soc. London B*, Vol 207 (1980) pp 187–217

27 **Torre, V and Poggio, T A** 'On edge detection', *IEEE Trans. PAMI*, Vol 8 No 2 (1986) pp 147–163

28 **Kanatani, K** *Group-Theoretical Methods in Image Understanding*, Vol 20 of *Series in Information Sciences*, Springer-Verlag, Berlin (1990)

29 **Canny, J** 'A computational approach to edge detection', *IEEE Trans. PAMI*, Vol 8 No 6 (1987) pp 679–698

30 **De Micheli, E, Caprile, B, Ottonello, P and Torre, V** 'Localization and noise in edge detection', *IEEE Trans. PAMI*, Vol 10 No 11 (1989) pp 1106–1117

31 **Koenderink, J J** 'Scale-time', *Biol. Cybern.*, Vol 58 (1988) pp 159–162