Tutorial

# Scale-Space Theory for Multiscale Geometric Image Analysis

***Bart M. ter Haar Romeny, PhD***
*Utrecht University, the Netherlands*
B.terHaarRomeny@isi.uu.nl

## Introduction

Multiscale image analysis has gained firm ground in computer vision, image processing and models of biological vision. The approaches however have been characterised by a wide variety of techniques, many of them chosen ad hoc. Scale-space theory, as a relatively new field, has been established as a well founded, general and promising multiresolution technique for image structure analysis, both for 2D, 3D and time series.

The rather mathematical nature of many of the classical papers in this field has prevented wide acceptance so far. This tutorial will try to bridge that gap by giving a comprehensible and intuitive introduction to this field. We also try, as a mutual inspiration, to relate the computer vision modeling to biological vision modeling. The mathematical rigor is much relaxed for the purpose of giving the broad picture. In appendix A a number of references are given as a good starting point for further reading.

## The multiscale nature of things

In **mathematics** objects have no scale. We are familiar with the notion of points, that really shrink to zero, lines with zero width. In mathematics are no metrical *units* involved, as in physics. Neighborhoods, like necessary in the definition of differential operators, are defined as taken into the limit to zero, so we can really speak of *local operators*.

In **physics** objects live on a *range* of scales. We need an instrument to do an observation (our eye, a camera) and it is the range that this instrument can see that we call the scale range. To expand the range of our eye we have a wide armamentarium of instruments available, like microscopes and telescopes. The scale range known to humankind spans about 50 decades, as is beautifully illustrated in the book (and movie) "Powers of Ten" [Morrison 1985]. The range one instrument can see is always necessarily bounded on two sides: the *inner scale* is the smallest detail seen by the smallest aperture (e.g. one CCD element of our digital camera, a cone or rod on our retina); the *outer scale* is the coarsest detail that can be discriminated, i.e. it is the whole image (field of view).
In physics dimensional units are essential: we express any measurement in these units, like meters, seconds, candelas, ampères etc. There is no such thing as a physical 'point'.

In mathematics the smallest distance between two points can be considered in the limit to zero, but in physics this reduces to the finite aperture separation distance (sampling distance). Therefore we may foresee serious problems with notions as differentiation, especially for high order (these problems are known as regularization problems), sub-pixel accuracy etc. As we will see, these problems are just elegantly solved by scale-space theory.

In **front-end vision** the apparatus (starting at the retina) is equipped just to extract multi-scale information. Psychophysically is has been shown that the threshold modulation depth for seeing blobs of different size is constant (within 5%) over more than two decades, so the visual system must be equipped with a large range of sampling apertures. There is abundant electrophysiological evidence that the receptive fields (RF's) in the retina come in a wide range of sizes[1] [Hubel '62, '79a, '88a].

In any image analysis there is a *task*: the notion of scale is often an essential part of the description of the task: "Do you want to see the leaves or the tree"?

**Linear Scale-Space Theory - Physics of Observation**

To compute any type of representation from the image data, information must be extracted using certain *operators* interacting with the data. Basic questions then are: What operators to use? Where to apply them? How should they be adapted to the task? How large should they be? We will derive the kernel from first principles (axioms) below.

These operators (or filters, kernels, apertures: different words for the same thing) come up in many tasks in signal analysis. We show that they are a necessary consequence of the physical process of *measuring* data. In this section we derive from some elementary axioms a complete family of such filters. As we will see, these filters come at a continuous range of sizes. This is the basis of scale-space theory.

If we start with taking a close look at the observation process, we run into some elementary questions:

- What do we mean with the 'structure of images' [Koenderink 1984]?
- What is an image anyway?
- How good should a measurement (observation) be?
- How accurately can we measure?
- How do we incorporate the notion of scale in the mathematics of observation?
- What are the best apertures to measure with?
- Does the visual system make *optimal* measurements?

*Any* physical observation is done through an aperture. By necessity this aperture has to be finite (would it be zero no photon would come through). We can modify the aperture

[1] It is not so that every receptor in the retina (rod or cone) has its own fiber in the optic nerve to further stages. In a human eye there are about $150.10^6$ receptors and $10^6$ optic nerve fibres. Receptive fields form the elementary 'apertures' on the retina: they consist of many cones (or rods) in a roughly circular area projecting to a single (ganglion) output cell, thus effective integrating the luminance over a finite area.

considerably by using instruments, but never make it zero width. This implies that we never can observe the physical reality in the outside world, but we can come close. We can speak of the (for us unobservable) infinite resolution of the outside world.

We consider here physical observations by an initial stage measuring device (also called front-end) like our retina or a camera, where no knowledge is involved yet, no preference for anything, and no nonlinearities of any kind. We call this type of observation *uncommitted*. Later we will relax this notion, among others by incorporating the notion of a model or make the process locally adaptive to the image content, but here, in the first stages of observation, *we know nothing*.
This notion will lead to the establishment of *linear* scale-space theory. It is a natural requirement for the first stage, but not for further stages, where extracted information, knowledge of model and/or task comes in etc. We then come into the important realm of *nonlinear* scale-space theory, which will be discussed in section 4.

Scale-space theory is the theory of kernels where the size (‘scale’) of the kernel is a free parameter. This multi-scale approach is actually the natural physical situation. A single constant-size aperture function may be sufficient in a controlled physical application (e.g. fixed measurement devices), but in the most general case *no a priori size* is determined. Control is needed over the scale. Scale-space theory comes around whenever we observe our system of study, and thus it is applied at feature detection, texture, optic flow, disparity, shape, etc.

Noise is always part of the observation. We cannot separate it from the data. It can only be extracted from the observation if we have a model of the observed entity or the noise, or both. Very often this is not considered explicitly. One e.g. often assumes the object comes from the 'blocks world' so it has straight contours, but often this is not known.

In the next paragraph we consider the aperture function as an operator: we will search for *constraints* to pin down the exact specification of this operator. For an unconstrained front-end there is a unique solution for the operator: the Gaussian kernel. If there is no preferred size of the operator, it is natural to consider them at all sizes, i.e. as a family, parameterized by the parameter for scale (one per dimension): the standard deviation (the 'width' or ‘scale’) of the Gaussian kernel.

**The aperture function of an uncommitted front-end**
To derive the aperture function we first establish the requirements and constraints appropriate for the physical situation at hand.

Let us study an example where things go wrong: if we look at an image through a square aperture, and make larger copies of the aperture to do the sampling at larger scales, we get the blocky effect seen in figure 1. [Koenderink 1984a] coined this *spurious resolution*, i.e. the emergence of details that were there not before. We seem to have created new detail such as lines and corners in the image. By looking at the result through your eyelashes, you blur the image and remove the extra detail again. As a general rule we want the information content only to decrease with larger aperture.

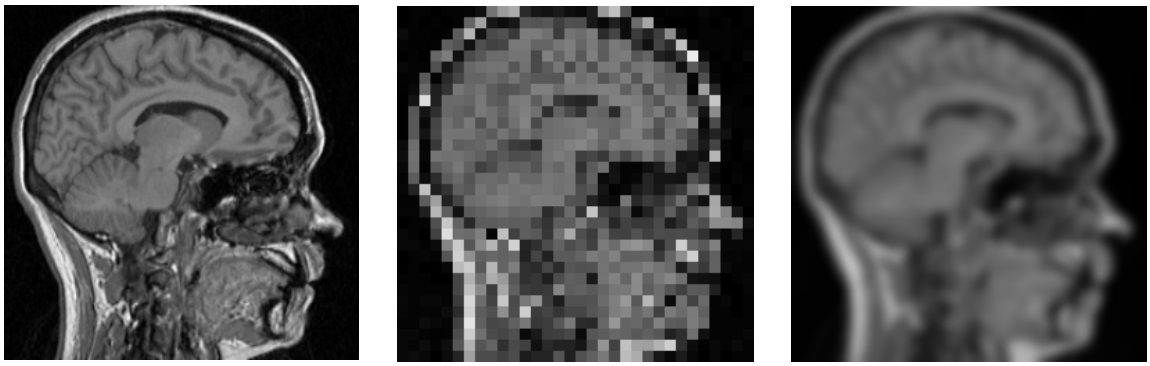

Figure 1: Spurious resolution due to square apertures. Left: Original sagittal MR. Middle: original image resampled at every $8^{th}$ pixel in x- and y direction. Representation with pixel replication, giving spurious resolution. Right: same resampling, representation with Gaussian kernel with $\sigma$ = 4 pixels. All images are represented as 256 x 256 pixels.

The following line of reasoning is due to [Florack et al. 1992k]. We should realize that the description of *any* physical system must be described in a way independent of the particular choice of coordinate system (Einstein got famous for it). If we change the coordinate system, then the description must still describe the same system. We want to be *invariant* under the actions of a particular coordinate transformation. There are many *groups* of coordinate transformations (we discuss an overview of the most relevant groups later on). We deal with medical images mostly, so we will consider the group of orthogonal coordinate transformations: translations, rotations and mirrorings, where the coordinate axes remain orthogonal. Further, we want no particular knowledge or model involved, so the system must be linear. Any knowledge or model would just *be* the nonlinearity, as we will see later in the nonlinear scale-space theories.

The requirements can be stated as axioms, or postulates for an uncommitted visual front-end. In essence it is the mathematical formulation for "we know nothing, we have no preference whatsoever":

- linearity (no knowledge, no model, no memory)
- spatial shift invariance (no preferred location)
- isotropy (no preferred orientation)
- scale invariance (no preferred size, or scale)

## Scale and Dimension

Every physical unit has a physical dimension, and it is this that mostly discriminates physics from mathematics. It was Baron Jean-Baptiste Fourier (yes, the same as of the Fourier Transform) who already in 1822 established the concept of dimensional analysis. This is what every physicist does first, when he inspects a new formula: are the dimensions correct? Are they the same in all expressions of the formula? It is impossible to add meters to meters/second.

***Law of Scale Invariance:*** *Physical laws must be independent of the choice of fundamental parameters.*

Dimensional analysis is an elegant tool to find out basic relations between physical entities, or even to solve a problem. It is often a method of first choice, when no other information is available. It is often quite remarkable how much one can deduct by just using this technique. We will use dimensional analysis to establish the expression defining the basic linear isotropic scale-space kernel.

First we set up a matrix of all dimensional units and physical variables involved in our physical system. We consider the N-dimensional case, in the Fourier domain (this will turn out to lead to easier expressions). The variables are: $\sigma_x$ the 'width' of the aperture function in the spatial domain, $\omega$ the spatial frequency, $L_0$ the luminance distribution of the outside world (the driving conditions), and $L$ the luminance after observation through the aperture. Here is the matrix:

| | $\sigma_x$ | $\omega$ | L | $L_0$ |
|---|---|---|---|---|
| **length** | +1 | -1 | 0 | 0 |
| **intensity** | 0 | 0 | +1 | +1 |

Note that we have 4 physical variables expressed in 2 fundamental physical units. We can make a number of dimensionless combinations. Dimensional analysis states that the dimensionless quantities should be functions of each other, which may be raised to arbitrary power. There is a theorem in physics, the Pi Theorem, that states how many dimensionless quantities can maximally be formed, given the matrix of unit dimensions and physical variables: this maximum number is the number of physical variables minus the rank of the matrix. The rank of the matrix above is 2, so in our case: 4-2=2.

For the two independent dimensionless quantities[2] we choose $L/L_0$ and $\sigma\vec{\omega}$, so

$$\frac{L}{L_0} = G(\sigma\vec{\omega})^p$$

where $G$ is the function (operator, kernel, filter) to be determined. The power $p$ comes from the fact that $\sigma\vec{\omega}$ is dimensionless. So here we have the first important relation: $L/L_0$ is some function of $\sigma\vec{\omega}$, raised to some power of $p$. Let us now determine the unknown function $G$, and $p$.

Spatial shift invariance implies convolution, i.e. we scan the aperture over any possible location of the image: $L(\vec{x},\sigma) = L_0 \otimes G(\vec{x},\sigma)$, where $\otimes$ stands for convolution:

$$f \otimes g = \int_{-\infty}^{\infty} f(u)g(x-u)du\,.$$

In the Fourier domain convolution becomes multiplication: $\hat{L}(\vec{\omega},\sigma) = \hat{L}_0 \cdot \hat{G}(\vec{\omega},\sigma)$.

[2] For the mathematically inclined: this can be calculated from the nullspace of the matrix, e.g. in Mathematica: NullSpace[m].

The isotropy axiom states that there is no preference for direction, so only the length of the vector $\vec{\omega}$ matters so we get a scalar: $\sigma\vec{\omega} \Rightarrow \sigma\omega$. When we study the asymptotic behavior of the system at the border scales, we find at the inner scale that the image is not scaled at all, i.e. our operator becomes the identity operator: $\lim_{\sigma\omega\downarrow 0} G(\sigma\omega) \rightarrow 1$, and at the outer scale complete spatial averaging occurs: $\lim_{\sigma\omega\rightarrow\infty} G(\sigma\omega) \rightarrow 0$.

The most important constraint is the requirement that the operation must be *self-similar*: the observation through an aperture blurs the image, i.e. increases the inner scale, and when we observe this observation again, the inner scale is again increased with the scale of the aperture. The total scaling of such a cascade of scalings must be consistent with performing just a single rescaling with a larger scale and, this is the important fact, which has the same mathematical shape.

So, a concatenation of two rescalings $G(\sigma_1\omega)$ and $G(\sigma_2\omega)$ should be a rescaling $G(\sigma_3\omega)$ with the effective scale parameter $\sigma_3 = \sigma_1 \oplus \sigma_2$. Note that $\oplus$ need not to be ordinary addition. In mathematical terminology, the operator $\oplus$ and the group of positive numbers constitute a *commutative semigroup*. A result from the theory of semigroups is that $G(\omega\sigma_1)\, G(\omega\sigma_2) = G((\omega\sigma_1 + \omega\sigma_2)^p)$. A general solution to this constraint where a product of functions givs a function with the sum off the variables is the exponential function: $G(\omega\sigma) = \exp(\alpha\omega\sigma)^p)$. We still have our power of *p* as $\alpha\omega\sigma$ is dimensionless.

The *D* spatial dimensions are independent, thus *separable*: $G(\omega\sigma) = \prod_{i=1}^{D} G(\omega_i\sigma_i)$, so scaling in each direction separately is the same as an isotropic scaling with $G(\sigma\omega)$. This fixes *p*=2, because the length of the total projection vector in our Euclidean space is calculated by the Pythagoras formula from the magnitudes of the constituent projections $\vec{\omega}\sigma \cdot \hat{e}_i\hat{e}_i$.

We demand a real solution, so we want $\alpha^2$ to be real. Because of $\lim_{\sigma\omega\rightarrow\infty} G(\sigma\omega) \rightarrow 0$ $\alpha^2$ is negative. We choose $\alpha^2 = -\frac{1}{2}$, because we then get a kernel with area of unity, so it does not multiply our observation by some factor. The kernel is then called a normalized kernel.

So finally we get our *normalized Gaussian kernel* in the Fourier domain:

$$\hat{G}(\vec{\omega}, \sigma) = \exp(-\tfrac{1}{2}\sigma^2\omega^2)$$

which is in the spatial domain

$$\hat{G}(\vec{x}, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}^D} \exp(-\frac{\vec{x}\cdot\vec{x}}{2\sigma^2})$$

So the surprising final result is that we can derive from a set of very basic axioms (no preference for anything) the Gaussian kernel as the *unique* kernel for an uncommitted front-end. Note that this is only true for the choice of uncommittment: when we relax that requirement, other kernels appear [Pauwels 93a, Florack 92a].

Looking at, or sampling, an image blurs the image. In mathematical terminology: convolution with a Gaussian necessarily increases the inner scale. The Gaussian is the *operator* that transforms the inner scale of the image. The *cascade property* states that it

is the same if one reaches a final inner scale in a single step from the input image by a given Gaussian aperture, or apply a sequence of (smaller) Gaussian kernels, to reach the same scale.

The stack of images as a function of increasing inner scale is coined a *linear 'scale-space'* [Witkin 1983, Koenderink 1984], see figure 2.

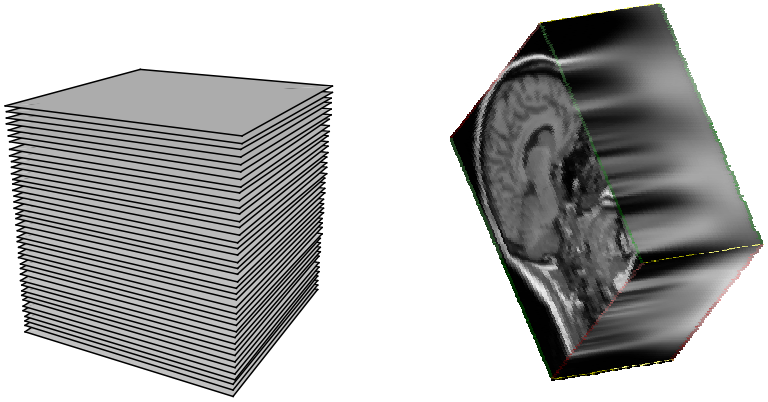

Figure 2. Left: A scale-space of an image is the stack of images at all scales between the inner and outer scale of the image. Right: zero-order or intensity scale-space (just blurring) of a sagittal MR image, increasing scale to the right.

It was first realized by Koenderink [Koenderink 1984a] that the generating equation of a linear scale-space is the linear *diffusion equation*:

$$\frac{\partial L}{\partial s} = \vec{\nabla} \cdot \vec{\nabla} L = \Delta L = L_{xx} + L_{yy}$$

(for 2D), stating that the derivative to scale equals the divergence of the gradient of the luminance function, which is the Laplacean, the sum of the second partial derivatives. The blurring can be considered as the diffusion of the intensity over time, but time is now taken as our scale. The Gaussian is the Green's function of the diffusion equation. When the diffusion is equal for all directions, i.e. the sigma's of the Gaussian are equal, we call the process *homogeneous*. When the diffusion is equal for every point of the image, we call the process *isotropic*. Because of the diffusion equation, the process of generating a multiscale representation is also known as *image evolution*. Later we will encounter many nonlinear diffusion processes with their nonlinear diffusion equations.

The derivation of the diffusion equation has been accomplished in many ways [Babaud 1986], among which we mention:

- Causality [Koenderink 1984a]: coarser scales can only be the causal result of what happened at finer scales;
- Maximum principle [Hummel1984]: at any increase of the inner scale (blurring) the maximal luminance at the coarser scale is always lower then the maximum intensity at the finer scale, the minimum is always larger;
- No new extrema are formed at larger scales [Lindeberg 1990a]: this holds *only* for one-dimensional signals;
- The linear (isotropic) diffusion process always reduces the information content (expressed as the log of the entropy of the signal) with increasing scale [Jagersund 1994a];

- Physics of the diffusion process (see textbooks in physics, e.g. [Feinmann]): the decrease of the luminance with time (or scale, which is here equivalent) is equal to the divergence of a *flow*, i.e. equal to the divergence of the luminance gradient.

Note that the $s$ in the diffusion equation has the dimension of the *squared* spatial dimension, so it takes the role of the Gaussian variance, not the standard deviation. The relation to the standard deviation is: $\sigma^2 = 2s$. One can see this also by considering the dimensions of each term in the diffusion equation.

### Gaussian derivatives and regularization

The Gaussian kernel is now established as the unique scale-space operator to change scale. There is an important additional result: One of the most useful results in linear scale-space theory is that also *all partial derivatives* of the Gaussian kernel are solutions of the diffusion equation, and together with the zero-th order Gaussian they form a *complete family of scaled differential operators* (fig.3).

Figure 3: Gaussian derivative profiles as multiscale differential operators. From left to right: the Gaussian kernel $G(x,y;\sigma)$, $\frac{\partial G(x,y;\sigma)}{\partial x}$, $\frac{\partial^2 G(x,y;\sigma)}{\partial x^2}$, $\frac{\partial^2 G(x,y;\sigma)}{\partial x dy}$, and the Laplacean $\frac{\partial^2 G(x,y;\sigma)}{\partial x^2} + \frac{\partial^2 G(x,y;\sigma)}{\partial y^2}$.

If we want to take the derivative of a discrete image, this is the derivative of an *observed* image $L \otimes G$, where the observation process is expressed as the convolution of the outside world image with the aperture function. We may interchange (commute) the differential and the convolution operator because both are linear (this is easily prooved in the Fourier domain):

$$\frac{\partial}{\partial x}(L \otimes G) = L \otimes \frac{\partial G}{\partial x}$$

So the derivative is found by convolving the image with the derivative of a Gaussian. This means that the derivative outcome is given at a given *scale*: We call the derivative operator the *scaled derivative operator* or *Gaussian derivative operator*. Differentiation and observation is done in a single process. The visual system seems to exploit this strategy by looking at the outside world through receptive fields that have the sensitivity profile of Gaussian derivative functions.

Note particularly that differentiation and scaling are intrinsically connected: it is impossible to differentiate discrete/sampled data without increasing the inner scale, i.e. without blurring the image a little. This is a natural consequence of a very important property of this family of differential operators: the *regularization* of the differentiation

process. Differentiation is now done by integration. This regularization is one of the most important results of scale-space theory.

Differentiation is known to be *ill-posed*. What does this mean?
For any physical process one wants the output to vary only slightly when some operator applied to the data is varied slightly. Take for example a small high frequency disturbance on our luminance distribution, on our image:

$$\hat{L}(x) = L(x) + \varepsilon \cos \omega x$$

The Gaussian derivative is given by

$$\hat{L}_x(x,\sigma) = L_x(x,\sigma) + \varepsilon\omega\, e^{-\omega^2\sigma^2/2} \cos \omega x\,.$$

The exponential term is the regularizing factor. The mathematical derivative would have been obtained for the case where the scale goes to zero, i.e. the exponential term would not be there. Note that we have to include the extra parameter $\sigma$ in our writing of the derivative function $\hat{L}_x(x,\sigma)$. We see that disturbances can be made arbitrarily small provided that the derivative of the signal is computed at sufficiently coarse scale $\sigma$ in scale-space.

The notion of increasing the inner scale with any differentiation is counterintuitive. We are so used to the mathematical notion of a zero-scale derivative that it takes some reflexion to see what happens here. *Looking at* the data *is* the regularization process. Important is to realize that the operator is regularized, and not the data. The data should never be modified. This rule, however, is often seen to be violated: many approaches exist to regularize the data: cubic splines, thin plate approximations, deformable templates, graduated nonconvexity, blurring etc. In essence the outcome may of course be the same or similar, the philosophy differs. We now understand why so many approaches first blur the data and then they differentiate. A good review of scale-space operators in the context of regularization is given by [Nielsen 1996b].
For the mathematicians: [Florack 1992d] showed that the problem of ill-posedness was originally solved by the famous mathematician Schwartz [Schwartz 1951a, Schwartz 1966], who showed that the derivative of a mathematical *distribution* is obtained by convolving the distribution with the derivative of a (any) smooth *test function*. In our case the discrete image is the distribution, the test function is the smooth aperture function, i.e. the (infinitely differentiable) Gaussian kernel. So images can be considered 'regular tempered distributions' in the mathematical language of distribution theory.

The notion of self-similarity of the kernels becomes especially clear when we express them in so-called *natural coordinates*, i.e. dimensionless spatial units. We then make all measurements relative to the scale at hand: we can consider the scale of the operator as the natural yardstick to measure spatial properties at that scale. Expressed in the natural coordinate $\tilde{x}$, where $\tilde{x} \mapsto \frac{x}{\sigma}$. The Gaussian derivatives get a factor in front:

$$\frac{\partial^n G}{\partial \tilde{x}^n} \mapsto \sigma^n \frac{\partial^n G}{\partial x^n}$$

The limiting case, when the scale of the operator shrinks to zero, leads to the familiar mathematical differential operator. The zero-th order operator, the Gaussian kernel itself, shrinks to the well known delta Dirac function (also called the Kronecker delta function, or identity operator). So the Gaussian kernel itself can be seen as the *scaled identity operator*, it does nothing but changing the inner scale, i.e. it only blurs.

This intrinsically regularizing framework now enables the extraction of in principle any order of derivative for discrete data. As it turns out, for higher order derivatives we need larger scales. There we encounter a limit: there is a fundamental relation between the order of differentiation, the scale of the operator and the accuracy of the result (for details see [ter Haar Romeny et al. 1991]).

Another important concern is: what scale to take? We will discuss this later in the context of (automatic) *scale selection*, which of course may even involve a different scale in any point of the image.

But first we consider the application of differential geometry and tensor analysis on discrete image data with our regularized derivatives.

## Geometric Structure and Invariance

The set of partial derivatives to order $N$ in a particular point in the image is called the local *N-jet*, and the set of scaled (i.e. Gaussian) derivatives the multiscale local N-jet. The only relevant geometric entities to consider are properties that are not changed when we change our coordinate system, i.e. that are *invariant* under a particular coordinate transformation. Gaussian partial derivatives alone are not invariant, in general they change when we e.g. rotate the coordinate system. So we need to construct particular combinations of derivatives as invariant features, i.e. local image properties. These local image properties, expressed as some (mostly polynomial) combination of the N-jet elements, are called the local invariants, or, because we incorporate derivatives, local differential invariants. They are always scalars.
A first example is the inner product of the gradient vector with itself, so we get the *length* of the gradient. Length, in Euclidean geometry, may be defined without reference to coordinate axes, so the dot product of two vectors is a geometric invariant. As an intuitive notion it is useful to realize the fact that the invariant property is 'attached' to the object, not to the coordinate frame. Hermann Weyl (1885-1955 [Weyl 1946]), famous for his contributions to invariant theory, stated it clearly: "Every invariant has a geometric meaning".

Coordinate transformations are expressed as elements of *groups*. Some important groups of coordinate transformations in computer vision are:

- the group of orthogonal coordinate transformations: rotations and translations of an orthogonal coordinate frame;
- the affine coordinate transformations: linear transformations of the coordinate axes: rotations, translations, shear and expansion;
- the similarity group: the orthogonal group expanded with scaling;

- the perspective transformations: projective coordinate transformations, most conveniently expressed in homogeneous coordinates.

Coordinate transformations are characterized by the matrix of the transformations. When the determinant of this matrix is unity, we 'create no extra volume', we call the group *special*. In the following we will focus primarily on medical imaging, where the special group of orthogonal transformations is of most interest.

## Tensors

Tensor analysis forms a useful and handy mechanism to describe, construct and manipulate properties in coordinate independent physics. As Einstein's philosophers' stone (the absolute differential calculus) tensor analysis gives us an easy way to construct the invariants to our need.

Tensors are actually lists of lists of lists of ... etc. A vector is a one-tensor, a single list of numbers. A two-tensor is represented as a matrix, and its physical meaning can be seen as a linear operator or linear transformation, transforming a vector in a vector. Tensors are equipped with indices, one for each list-index. An excellent introduction to the tensor analysis needed as concise introduction to scale-space theory is [Simmonds 1995a].

In scale-space theory, we consider spatial partial derivatives as tensor elements, so the index of the tensor denotes the derivative to each dimension. E.g. the one-tensor $L_i$ denotes the gradient vector of the (e.g. 3-dimensional) luminance function $L(x,y,z)$, consisting of the first order derivatives to $x$, $y$ and $z$:

$$L_i \equiv \tilde{\nabla} L \equiv \left\{ \frac{\partial L}{\partial x}, \frac{\partial L}{\partial y}, \frac{\partial L}{\partial z} \right\}$$

In the same fashion the matrix of second order partial derivatives (the *Hessian matrix*) is defined as the two-tensor $L_{ij}$. Both the index $i$ and $j$ run independently over the dimensions. The Hessian matrix has 9 elements in 3D. The trace (sum of diagonal elements) of the Hessian is called the Laplacean of $L$.

There are two constant tensors, not involving derivatives of $L$:

- the Kronecker delta tensor $\delta_{ij}$, defined as the unity or symmetry operator. It is also called the identity matrix. This tensor has the value 1 when the indices have the same value, and zero otherwise. So only the diagonal elements are 1, all others are zero.
- the Levy-Civita epsilon tensor $\epsilon_{ij\ldots}$, defined as the anti-symmetry operator. The elements of this tensor take the value of the sign of the permutations of the indices $ij\ldots$ There are as many indices as there are dimensions, so in 3D: $\epsilon_{ijk}$.

Examples in 2D: $\delta_{ij} \equiv \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, $\varepsilon_{ij} \equiv \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$

There is much to say about tensors, and, as it turns out, the theory simplifies substantially when we consider Cartesian frames only: Consider for example what happens to the image gradient $L_i$ when we apply a linear transformation $a_{ij}$ on the displacement vector $\tilde{x}_i = a_{ij} x_j$ (and thus $x_i = a_{ij}^{-1} \tilde{x}_j$). The chain rule says that $\tilde{\nabla}_i L = a_{ij}^{-1} L_j$. This shows that in general, $x_i$ and $L_i$ transform differently: $x_i$ is said to be a contravariant vector, and $L_i$ a

covariant vector. Contravariant vectors are given upper indices, so $x_i$ becomes $x^i$. The good news is that for Cartesian orthogonal coordinate frames the notion of covariant (lower indices) and contravariant (upper indices) tensors disappears (see for a clear explanation [Simmonds 1995a] or [Misner 1973]), because of the fact that the frame is orthogonal: $a_{ji}^{-1} = a_{ij}$. From now on we will only use lower indices.

Invariant properties are obtained when we *contract* tensors, i.e. we form inner products in such a way that scalar values crank out. This is done by *summing over paired indices*. The summation symbol is often omitted which is known as the *Einstein convention*. E.g. in 3D indices run over *x*, *y* and *z* and we get the following contractions:

$$L_i L_i \equiv \sum_{i=x,y,z}^{D} L_i L_i = L_x^2 + L_y^2 + L_z^2 \qquad \text{(squared gradient of } L\text{)}$$

$$L_{ii} \equiv \sum_{i=x,y,z}^{D} L_{ii} = L_{xx} + L_{yy} + L_{zz} \qquad \text{(Laplacean of } L\text{)}$$

where $L_x$ denotes $\frac{\partial L}{\partial x}$. Of course, contraction can also incur the constant tensors. Loosely spoken, contraction with the Kronecker delta tensor gives an inner product, with the Levy-Civita tensor an outer product.

We can of course make many other examples of contracted tensors: $L_i L_{ij} L_j$, $\varepsilon_{ij} L_i L_j$, $L_{ij} L_{ji}$.

In this way an endless number of invariant geometric properties can be constructed for all points of the image at a particular scale or range of scales. In practice, the partial derivative images are calculated first, and then assembled to an invariant image as the Cartesian formula indicates.
For clear reasons, such contracted invariants from indices are called *manifest invariants*. Many papers have appeared applying the successful application of (often complex) differential geometry to images [Florack 1992b, ter Haar Romeny 1994c, Thirion 1995a, Weber 1993a, Monga 1992].

Invariant theory had a major development about a century ago, and is now again fully actual in many important computer vision areas (see the excellent review book [Mundy 1992]. Differential geometry is a very suitable language to describe geometrical properties, such as shape, texture, motion etc. The brain can also be seen as an ‘geometry engine’ [Koenderink 1989].

It was shown by Hilbert [Hilbert 1893a] that *any* invariant of finite order can be expressed as a polynomial expression of a set of *irreducible invariants*. This set is typically very small, e.g. for 2D to second order the list of irreducible invariants counts only five elements, i.e. $L$, $L_i L_i$, $L_{ii}$, $L_i L_{ij} L_j$ and $L_{ij} L_{ji}$. The irreducible invariants form in each pixel the most concise *basis* of structural descriptors, and as such are important in texture description and pattern classification [Schmidt 1996b].

## Gauge coordinates

So far we considered so-called *extrinsic geometry*, i.e. we used an external coordinate frame. Formulas markedly clean up when we apply *intrinsic geometry*, i.e. when we are able to define *locally* such a coordinate frame, that one or more of the partial derivatives is zero (in physics this is known as the gauge condition). So we adapt in each point the local frame to the object. We call this new frame $\{v,w\}$. The easiest example in 2D is to line up one axis of the coordinate frame with the direction tangential to the isophote (line of constant intensity) in each point. This gives the *v* axis, and the *w* axis then aligns with the gradient, perpendicular to the isophote. See figure 4.

An infinitesimal deviation along the *v* axis gives no variation in the luminance as we move along the isophote, so by definition $L_v \equiv 0$. This is called the "*fixing of the gauge*". Also by this definition, $L_w$ gives the direction along which the luminance changes most: the gradient.

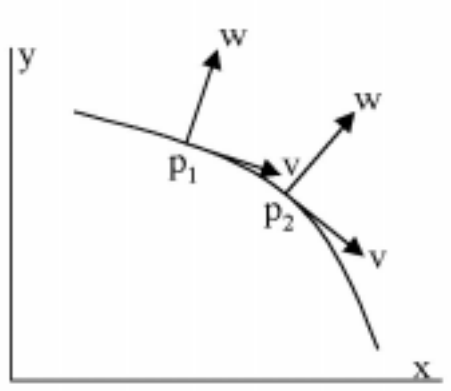


Figure 4: The gauge frame coordinates $\{v,w\}$ in two points $p_1$ and $p_2$ on an isophote are defined as having the *v* unit-vector as tangential to the isophote in each point, and the *w* unit-vector in the direction of the gradient of the luminance, i.e. perpendicular to the isophote in each point.

We have frozen the extra degree of freedom that we had for our local coordinate system: a rotation. Therefore we get the handy result: *any* differential expression expressed in partial derivatives with respect to the gauge coordinates *v* and/or *w* is an orthogonal invariant. So e.g. $L_{vv}$ is an invariant property invariant for translation or rotation (it is the local 'ridgeness' in a point), and $L_{vv}L_w^2$ is an invariant property for the same reason (it is the local 'cornerness').

Of course we can go from one representation to the other. This gives the transformation from gauge derivatives to manifest invariant notation. We recognize the direction of the *w* axis in the gradient direction $L_j$, and the *v* axis perpendicular to that, through the use of the contraction with the Kronecker tensor $\delta_{ij}$ and the Levi-Civita tensor $\varepsilon_{ij}$:

$$\frac{\partial}{\partial v} = \frac{\partial_i \varepsilon_{ij} L_j}{\sqrt{L_k L_k}}, \quad \frac{\partial}{\partial w} = \frac{\partial_i \delta_{ij} L_j}{\sqrt{L_k L_k}}.$$

In gauge coordinates the diffusion equation reads $L_s = L_{vv} + L_{ww}$, in manifest invariant notation $L_s = L_{ii}$.

As an example of the convenience of gauge coordinates, let us derive the expression for isophote curvature by studying the local isophote given by its defining formula *L(v,w(v))*

= $c$ (where $c$ is a constant) in the point $P$ by implicit differentiation to $v$ and application of the chain rule, using the gauge condition $L_v \equiv 0$:

$$L_v + L_w w' = 0 \quad \rightarrow \quad w'(P) = -\frac{L_v}{L_w}, \quad w'(P) = 0$$

Differentiating again:

$$L_{vv} + 2L_{vw}w' + L_{ww}w'^2 + L_w w'' = 0$$

The isophote curvature $\kappa$ by definition is $w''$: $\kappa(P) = -\frac{L_{vv}}{L_w}$, a convenient and compact equation. The curvature, a second order scalar property, is the reciprocal of the radius of the touching ('osculating') circle to the isophote. For a straight like the curvature is zero. It can take positive (convex) and negative (concave) values. In extrema, where the gradient $L_w$ vanishes, the curvature goes to infinity. To get rid of the division by the gradient, we can construct a new invariant by multiplication by powers of the gradient. In particular, the expression $L_{vv}\ L_w^2 = L_{xx}\ L_y^2 - 2\ L_x L_y L_{xy} + L_{yy} L_x^2$ turns out to be a good *corner detector*, see figure 5, and is invariant under affine transformations. The right hand term of the equation immediately shows how in practice we calculate this property from the Cartesian partial derivative images. Note that we don't need to construct the isophote itself. Due to the regularized nature of the scaled derivatives, the image is now fully differentiable, it is continuous. The quantity $L_{vv}$ is a good ridge detector, see fig. 6.

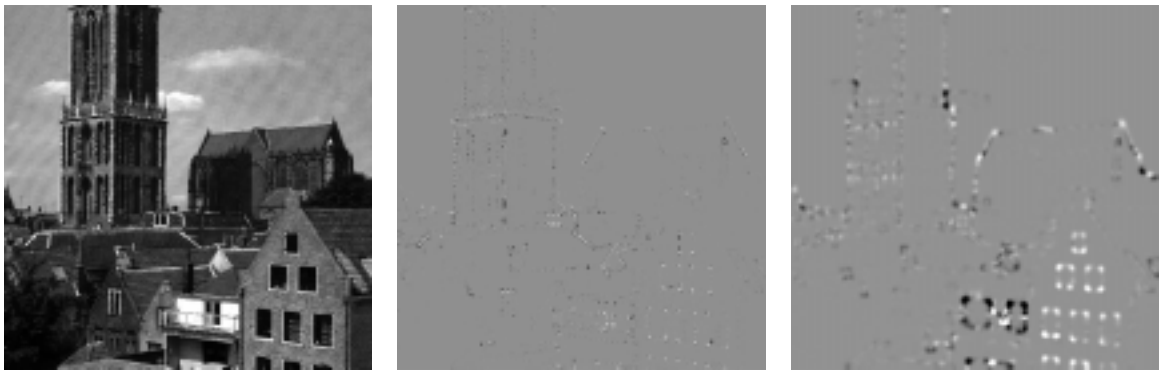

Figure 5: Corner detection by calculating $L_{vv}L_w^2$ in each pixel. Left: Image of the cathedral of Utrecht, resolution $512^2$. Middle: invariant output at a scale of 1 pixel. Right: idem at a scale of 4 pixels. Note the signs of concave and convex corners, the zero output in all non-corner pixels, and the differences for the different scales.

With differential geometry one has a natural language to do *geometric reasoning* e.g. about local shape properties. In 3D one can study the curvature of the isophote *surface* in all compass directions over the surface starting from the point in study. It turns out that the largest and the smallest curvatures are always found in perpendicular directions: these are the *principal curvatures,* along the *principal directions*. The product of the two curvatures is coined the Gaussian curvature, and is an important invariant feature. The sign of the Gaussian curvature e.g. indicates whether the surface is convex or concave. On local flat surface points the Gaussian curvature is zero, the connections of these points forming the *parabolic lines*. Surfaces that can be composed of straight lines, such as

cones and cylinders, have zero Gaussian curvature. The arithmetic mean of the principal curvatures is called the *mean curvature*. Surfaces that have zero mean curvature at any point are called *minimal surfaces*.

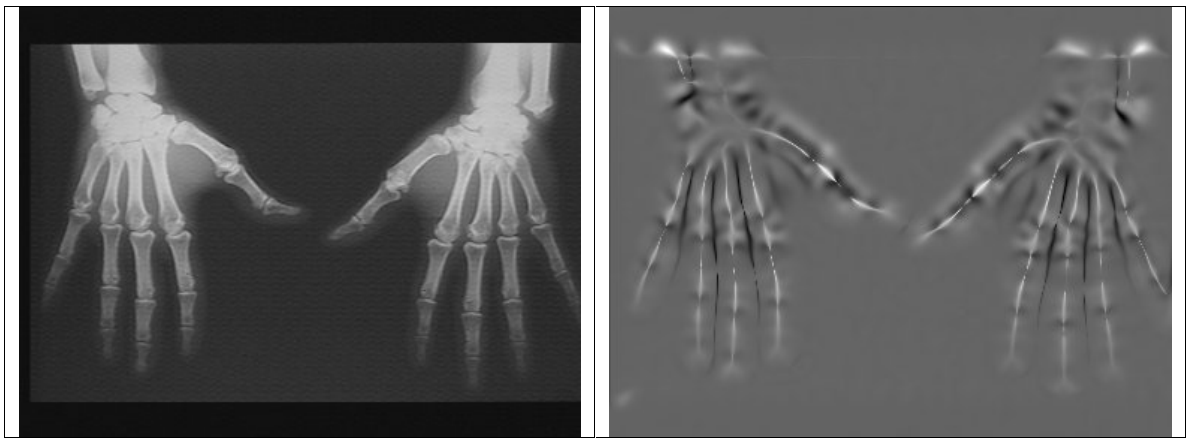

Figure 6: Ridge detection with the *Lvv* operator at a scale of 4 pixels. Resolution of hand radiograph 376x288 pixels.

In figure 7 the 3D Gaussian curvature $\frac{\varepsilon_{ijk}\varepsilon_{pqr}L_i L_{jp} L_r L_{kq}}{2(L_s L_s)^2}$ of the inner (isophote) surface of the left ventricle of a dog is shown for 16 frames of a time sequence, showing clearly the convex (dark) and concave (light) parts of the surface as well as the parabolic lines as their boundaries.

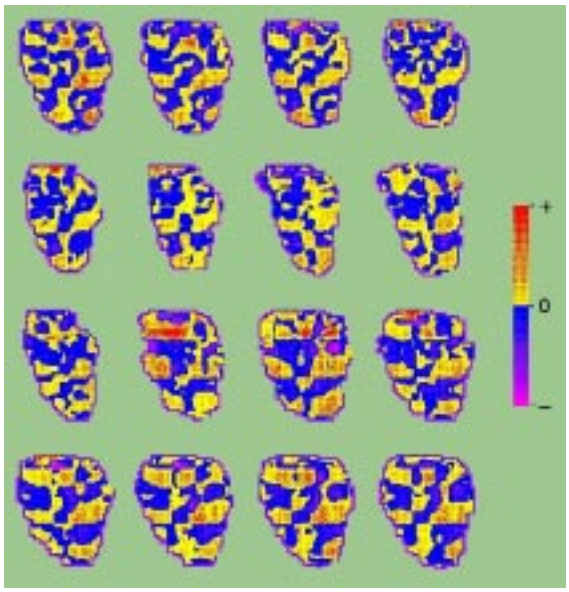


Figure 7: Gaussian curvature on the surface of a dogheart's left ventricle. 16 rowwise frames of a time series, 1/16 of a second apart of one heart cycle of one second. Yellow-red indicates concave regions, blue-purple indicates convex regions. The parabolic lines of zero Gaussian curvature are the separating lines. The change of 3D curvature over time is a sensitive measure of heartwall contraction, which may be severely diminished at infarct regions. From [Niessen 1997].
Images: Yale University 1996. Gated MR sequence.

The mesh formed by the lines (integral paths) of just the maximal or minimal curvature directions on a surface (the *extremal mesh*) has successfully been used in 3D image matching (registration) applications [Thirion 1995a].

To make structure analysis independent of e.g. brightness and contrast variations (e.g. the analysis should be the same even when you put up your sunglasses), invariance under general intensity transformations forms an important class. Evidently, the structural properties of isophotes (their curvature(s) and higher order derivatives of the curvature(s)) form the natural invariants under this requirement.

## Computer implementation

Gaussian derivatives are calculated by convolution of the image function with the derivative of the Gaussian. This can conveniently be handled in the Fourier domain, where a convolution reduces to a regular product, and the *n*-th order derivative operator reduces to multiplication with a factor $(i\omega)^n$. The Fourier transform of a derivative is $i\omega$ times the Fourier transform of the function:

$\hat{F}\left\{\frac{\partial G}{\partial x}\right\} = i\omega_x \hat{F}\{G\}$, and $\hat{F}\left\{\frac{\partial^n G}{\partial x^n}\right\} = (i\omega_x)^n \hat{F}\{G\}$ where $\hat{F}\{\ \}$ denotes the Fourier transformation. So the scaled (Gaussian) derivative of a function $L$ in the Fourier domain becomes:

$\hat{F}\{L \otimes G_{x_{1\ldots n}}\} = \hat{F}\{L\} \cdot (i\omega)^n \hat{F}\{G\}$, and finally $L \otimes G_{x_{1\ldots n}} = \hat{F}^{-1}\left\{\hat{F}\{L\} \cdot (i\omega)^n \hat{F}\{G\}\right\}$.

## Relation between order of differentiation, scale and accuracy

We can expect problems when we take high order derivatives at a small scale. High order derivatives have as many zero-crossings as the order dictates, and we are dealing with a discrete representation on a regular grid. There may just not be enough room under the Gaussian window to represent this function correctly. There exist a fundamental relation between the order of differentiation of an operator, its scale and the required accuracy [ter Haar Romeny 1994e]. For smaller scale $\sigma$ the Fourier transform of the kernel increases in width, and at a certain scale this gives rise to *aliasing*. In theory this occurs at all scales due to the infinite extent of the exponential function, but it becomes apparent at smaller scales. The 'leaked' information is folded back, in theory even from all further periods as well, so the amplitude no longer represents the accurate value of 'pure' differentiation.

To quantify the effect, we consider the powerspectrum, i.e. the square of the signal. The aliasing error can be defined as the relative integrated energy of the aliased frequencies over the total energy (note the integration boundaries):

$$error(n,\sigma) = \frac{\Delta E_n(\sigma,\omega)}{E_n(\sigma,\omega)} = \frac{\int_{\pi}^{\infty} \omega^{2n} e^{-\sigma^2\omega^2} d\omega}{\int_{0}^{\infty} \omega^{2n} e^{-\sigma^2\omega^2} d\omega} = \frac{\Gamma(\frac{1}{2}+n) - \Gamma(\frac{1}{2}+n), 0, 4\pi)}{\Gamma(\frac{1}{2}+n)}$$

where $\Gamma(n)$ is the Euler gamma function, and $\Gamma(n,z_0,z_1)$ is the generalized incomplete gamma function $\Gamma(n,z_0) - \Gamma(n,z_1)$. We used Mathematica to calculate this. Figure 6 shows the aliasing error as a function of scale $\sigma$ and order $n$ for orders up to 10. Note that indeed for higher order a larger scale must be selected.

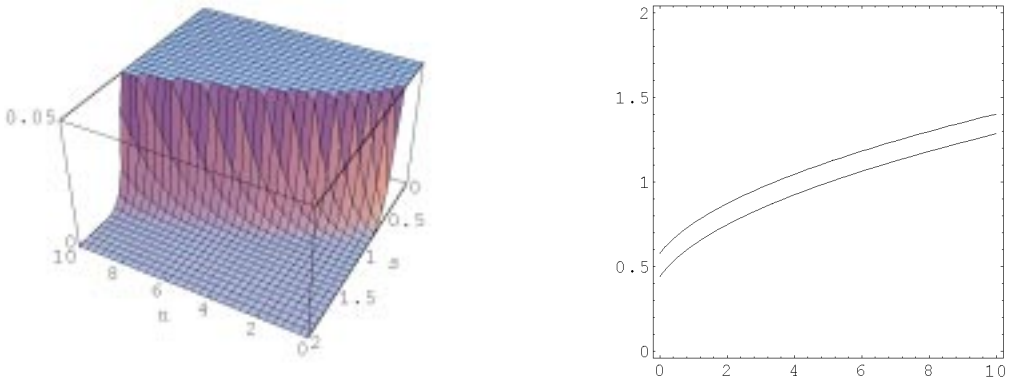


Figure 8. Left: Error of aliasing as a function of differential order (*n*) and scale (*s*) of the Gaussian derivative operator, expressed in pixels. Orders (*n*) up to 10, scales (*s*) up to 2 pixels. Top level is equal to 5% error. Right: Maximum differential order that can be extracted with Gaussian derivative operators without aliasing as a function of scale of the Gaussian derivative kernels expressed in pixels. Horizontal: orders up to 10, vertical: scales up to 2 pixels. Allowed error: 5% (upper line), 1% (lower line).

The error allowed is a matter of choice, dependent on the application. The error as a function of $\sigma$ and *n* is rather steep, indicating that the maximum allowed order for a given $\sigma$ (or the reverse) is rather independent of the particular choice of the error-level.

This error occurs due to limits in the *inner scale*. I.e. the kernel size approaches the resolution of the measurement. A similar line of reasoning can be put up for the error due to limits in the *outer scale*, when the kernel gets too wide in the *spatial domain*.

## Multiscale shape description and segmentation

The differential properties described so far are stricty *local*. We can describe it as keyhole image descriptions. Nothing is said about the relations between two neighboring points. This is the well-known problem of *perceptual grouping*. A number of approaches are proposed to attack this fundamental and important problem: [Pizer et al. 1994a] suggest the use of *symmetry* in shapes. The study of the medial axis (which is a result of processing multi-local information: the boundaries on each side) over scale gives a useful shape descriptor, enabling the quantification, matching and deformation studies of shape. The multi-scale 'medialness' or ridge function is coined the *core* of an object. It intrinsically carries the information of the 'width' of the object.

Intuitively it is easy to imagine that image structure has a hierarchical structure over scale. The decreasing information in the scale-space 'stack' gives rise to a graph-like structure, e.g. of the *singularities* (points where the gradient vanishes, e.g. intensity extrema, saddle points) and catastrophes (points where higher order derivatives vanish). The structure of this tree, also referred to as the *deep structure*, may be exploited to *link* pixels according to certain criteria to create *segments*. This is exploited in the study of the 4-dimensional (*x, y, z, s*) *hyperstack* [Vincken 1996, Niessen 1996b]. It enables the *top-down* hierarchical analysis of image structure: the number of required segments determines at what level in scale-space the down-projection to the ground level starts. Figure 7 shows a segmentation of the human cortex applying this idea.

Another important 'grouping' paradigm is the use of multiscale image features in *geometrical reasoning* or statistical grouping paradigms, like the Hough transform [Karssemeier 1995a].

## Scale Selection

The scale so far is a *free* parameter: if we don't know what scale to choose, just calculate them all. For an initial observation stage (like the retina) this may seem a logical strategy, for later stages there must be some *scale selection* for a number of reasons: efficiency of representation, and matching the captured data to the *task* (do we want the leaves or the tree).

There are many ways to select a most appropriate scale. Lindeberg pioneered this field by considering the ranking of the volumes objects span in scale-space, the *scale-space primal sketch*. An operator gives maximal output if its size is tuned best to the object. Recently other approaches have been proposed: the study of the variation of the information content (as the logarithm of the entropy of the voxel intensities) over scale [Jägersund 1994a], or the condition number of the coefficients when sets of simultaneous equations occur, e.g. in multiscale optic flow [Niessen 1995c].

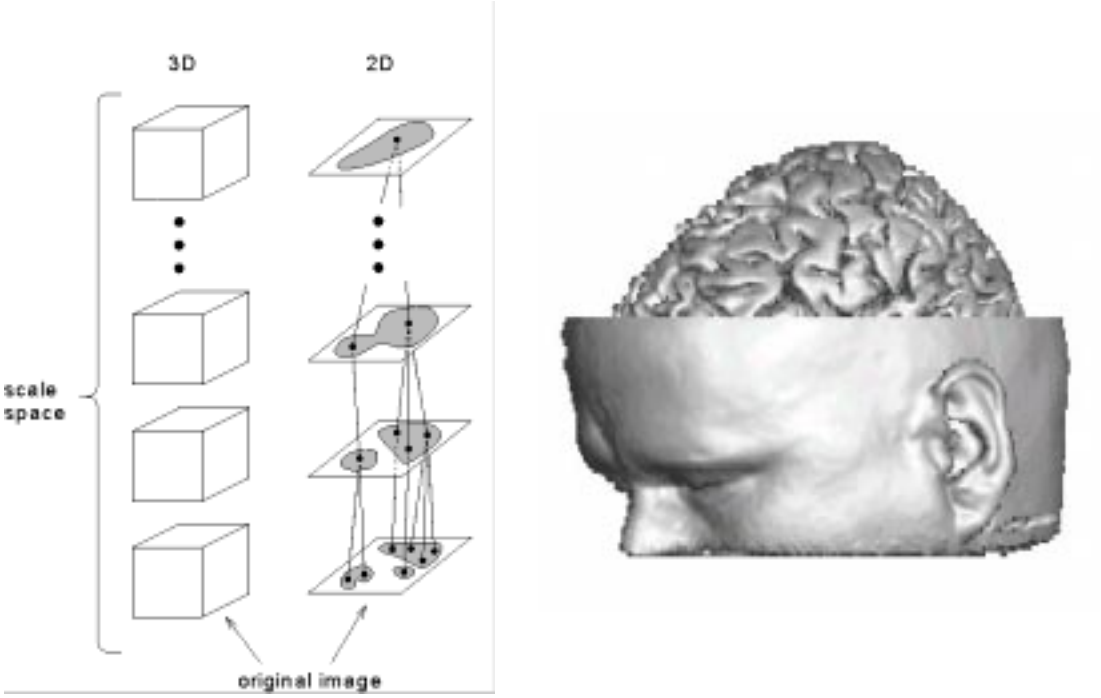


Figure 9. Exploiting the *deep structure* in scale-space to link 'similar' voxels in segments: segmentation of cortical areas. Left: the structure of links over scale. The 3D image datasets are blurred to a range of scales, forming the hyperstack. First the links are created in a bottom-up fashion as in a pyramid from the intensity values of e.g. neighboring 4 pixels. From a toppoint of choice a down-projection along the links can be made in order to group the pixels in the highest resolution image. Right: example of a segmented cortical surface from 3D MR data. From [Vincken 1996].

## The human visual front-end

There are many neurophysiological and psychophysical data indicating the multiscale analysis by the human visual front-end system. Front-end is here defined as the pre-attentive stage. The cones and rods in the retina form more or less circular *receptive fields* (RF) of a rather large range of sizes. In the fifties their sensitivity profiles have been discovered and measured by the pioneers Kuffler, Hubel and Wiesel. The measured 'center-surround' sensitivity profile closely resembles a Laplacean of a Gaussian. An excellent introduction to the early visual system is the Nobel price-winning work of Hubel and Wiesel [Hubel 1962] and the highly recommended books by [Hubel 1988a], [Zeki 1993] and [Rodieck 1998]

The receptive fields in the retina project to the *lateral geniculate nucleus* (LGN), a peanut shaped kernel in the thalamus in our mid-brain. From there an extensive bundle of projections run to the *visual cortex* in the back of our head. The receptive field profiles of the so-called simple cells in the cortex have a very close resemblance to the Gaussian derivatives [Young 1986]. These have even be proposed [Koenderink 1987a, 1987b, 1988b] to give a good *taxonomy* for the many kind of simple cells found. The cortex has many more complicated cells, such as complex cells, hyper complex cells, end-stopped cells etc.

The multi-scale 'stack' model enables an elegant model for the distribution of RF sizes over the retina [Lindeberg 1994i]. It is well known that our visual acuity decreases rather strongly with eccentricity from the fovea. When the paradigm of self-similarity is taken into account for the *processing capacity* per scale, all scales should be treated in a similar manner, i.e. with the same amount of 'wetware'. This leads to the *same number* of retinal receptive fields applied per scale, tesselated in a nice hexagonal array over a roughly circular area. The smallest RF's together lay down the foveal region, while the same number of larger receptive fields necessarily must occupy a larger area, but again centered in a circular area around the fovea. This superposition of receptive field 'carpets' elegantly explains the observed linear decrease with excentricity of many RF-size related phenomena, as acuity, velocity detection thresholds etc.

The Gaussian model only holds for the linear, uncommitted front-end. As soon as the possibility exist to use extracted geometrical information, kernels can be *tuned* to a particular task, leading to *nonlinear* kernels. The general framework for nonlinear diffusion equations is that the conductance term becomes a function of differential properties of the image, i.e. of the $N$-jet $J^N$:

$$\frac{\partial L}{\partial s} = \vec{\nabla} \cdot c(J^N) \vec{\nabla} L$$

It is interesting to consider the front-end connections in this nonlinear scale-space framework. In figure 10 a schematic is given of the initial *optical tract*. The first stage, retina to LGN, is completely uncommitted. Here we cannot have something else as a linear stage. The RF's in the LGN show a center-surround structure. Often not or barely mentioned, there is a massive amount of fibers projecting *backwards* (retrograde) from primary cortex to LGN [Sherman 1990a, 1993a]: about 50% of all LGN synapses derive from corticogeniculate neurons, while only 10-20% are from the retina, a striking

minority! Local inhibitory cells take care of about 25% of the synaptic input. These strong feedback channels, together with the multiscale oriented differential *N*-jet representation the cortical columns are capable of, provide input to the idea that the front-end visual system may modify the RF's as a function of the image itself. Recent accurate measurements of RF's show that indeed many RF's show strongly varying sensitivity profiles over time [DeAngelis 1995a].

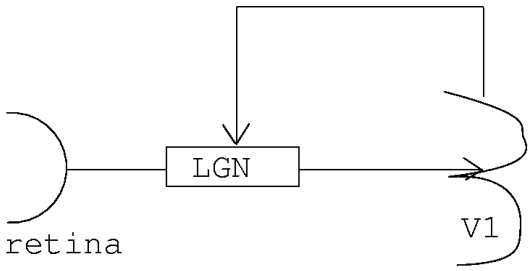


Figure 10. Schematic drawing of the initial optic tract with the forward and feedback connections between the Lateral Geniculate Nucleus (LGN) and the primary visual cortex (V1).

## Nonlinear Scale-Space Theory

Nonlinear scale-space theory is gaining much current interest. The classical paper initializing the field in computer vision was by Perona and Malik [Perona 1987], who proposed to make the conductance function a decreasing function of the weighted square of the gradient. The result is then that at point with high gradient output, there will be little diffusion (a small scale kernel is applied there), while at homogeneous areas the large kernel take care of good noise averaging. It was shown by Catté that the original partial differential equation (PDE) was ill-posed, and that the regularized extraction of the gradient with Gaussian derivatives makes the equation well behaved.

Many nonlinear adaptive image evolution schemes have followed. An overview of current theories is presented in [ter Haar romeny 1994f]. Broadly, they can be classified in a number of distinct catagories:

- scalar valued PDE formulations, like the original Perona and Malik proposal;
- vector valued diffusion [Whittaker 1994a], where the simultaneous influence of differential properties is studied in a set of coupled PDE's;
- tensor valued diffusion [Weickert 1994b], where the conductance is made up of the second moment matrix (or 'structure matrix') $D = \begin{pmatrix} L_x^2 & L_x L_y \\ L_y L_x & L_y^2 \end{pmatrix}$. The eigenvectors of this matrix indicate the 'main' directions of the local structure (it is a well known instrument in oriented texture analysis). The consequence is that diffusion is enhanced in e.g. the principal ridge detection, usefull in e.g. enhancing noisy fingerprint images. The Gaussian kernels become ellipsoids;
- affine and projective scale-spaces [Olver 1994d]: here the differential properties are expressed in the *invariant arclength* under the particular group. E.g. the affine arclength is used to calculate affine curvature in the affine scale-space paradigm.
- axiomatic morphological scale-spaces: Alvarez showed in a set of classical papers [Alvarez 1992, 1992a] an axiomatic approach to come to the equation $L_s = L_{vv}$, which

he coined the 'fundamental equation of image processing'. This equation was simultaneously found by Sapiro et al. [Sapiro 1992].

Nonlinear scale-space formulations by nonlinear PDE's can easily be implemented by explicit numerical approximations, e.g. forward Euler schemes [Niessen 1994a]. Recently, implicit schemes are developed, giving a must faster evolution 'timestepping' [Weickert 1996c].

[Mumford and Shah 1985a] initialized a new field by introducing *energy minimizing functionals*: the image evolution is considered as driven by an energy functional specified as $E(f,K) = \beta\int_{\Omega}(f-g)^2 dx + \int_{\Omega-K}|\nabla f|^2\, dx + \alpha|K|$ where $g$ is the observed grey value image, $f$ the approximation of $g$, $\Omega$ the image domain, $K$ the (closed) set of edges, and $|K|$ is the total length of the edgeset $K$. The penalty term ($f$-$g$) takes care that the image does not deviate too much from the original. Many modifications have been proposed on this general scheme.

Another important group of nonlinear evolution schemes is *curve evolution*. The curves here are the isophotes, as an image can be fully described as a nested set of isphotes. Originally derived for the study of propagating fronts, like flames, a robust mathematical framework has been set up [Sethian 1989a, Sapiro 1992a]. Osher and Sethian [Osher 1988] published a widely used numerical method to efficiently calculate curve evolutions of greyscale isophotes.

## Conclusions

Scale-Space theory is a solid mathematical/physical framework for multiscale image analysis. It has firm roots in the physics of observation. The scaled operators guarantee a regularized differentiation process, so we can apply the full breadth of differential geometry on discrete images of any dimension.


## Acknowledgements

The author gratefully acknowledges all members of the Utrecht University Image Sciences Institute ("Tools of Geometry in Vision" team: http://www.isi.uu.nl) on which work most of this tutorial material is based, and the members of the EC-US "Diffusion" collaboration.


September 26-27, 1999 the Second International Conference on Scale-Space Theory in Computer Vision will be held in Corfu, Greece, as a follow-up to the successful First Conference in Utrecht, July 1997.

Recommended books for reading on scale-space theory: [5], [12], [11], [23], [28], [56].

**Utrecht, the Netherlands. Summer 1999**

## References


1. L. Alvarez, F. Guichard, P. L. Lions, and J. M. Morel, “Axiomes et équations fondamentales du traitement d'images,” C. R. Acad. Sci. Paris, vol. 315, pp. 135--138, 1992. Also: report 9216, CEREMADE, Université Paris Dauphine, 1992, presented at Arch. for Rat. Mech., July 1992.
2. L. Alvarez, P.-L. Lions, and J.-M. Morel, “Image selective smoothing and edge detection by nonlinear diffusion. II,” SIAM J. Num. Anal., vol. 29, pp. 845--866, June 1992.
3. J. Babaud, A. P. Witkin, M. Baudin, and R. O. Duda, “Uniqueness of the Gaussian kernel for scale-space filtering,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 8, no. 1, pp. 26--33, 1986.
4. G. C. DeAngelis, I. Ohzawa, and R. D. Freeman, “Receptive field dynamics in the central visual pathways,” Trends Neurosci., vol. 18, pp. 451--458, 1995.
5. L. M. J. Florack, Image Structure. Computational Imaging and Vision Series, Dordrecht: Kluwer Academic Publishers, 1996.
6. L. M. J. Florack, B. M. ter Haar Romeny, J. J. Koenderink, and M. A. Viergever, “Linear scale-space,” Journal of Mathematical Imaging and Vision, vol. 4, no. 4, pp. 325--351, 1994.
7. L. M. J. Florack, B. M. ter Haar Romeny, J. J. Koenderink, and M. A. Viergever, “Families of tuned scale-space kernels,” in Proceedings of the European Conference on Computer Vision (G. Sandini, ed.), (Santa Margherita Ligure, Italy), pp. 19--23, May 19--22 1992.
8. L. M. J. Florack, B. M. ter Haar Romeny, J. J. Koenderink, and M. A. Viergever, “Images: Regular tempered distributions,” in Proc. of the NATO Advanced Research Workshop Shape in Picture - Mathematical description of shape in greylevel images (Y.-L. O, A. Toet, H. J. A. M. Heijmans, D. H. Foster, and P. Meer, eds.), vol. 126 of NATO ASI Series F, pp. 651--660, Springer Verlag, Berlin, 1994.
9. L. M. J. Florack, B. M. ter Haar Romeny, J. J. Koenderink, and M. A. Viergever, “Cartesian differential invariants in scale-space,” Journal of Mathematical Imaging and Vision, vol. 3, pp. 327--348, November 1993.
10. J. Fourier, The Analytical Theory of Heat. New York: Dover Publications, Inc., 1955. Replication of the English translation that first appeared in 1878 with previous corrigenda incorporated into the text, by Alexander Freeman, M.A. Original work: “Théorie Analytique de la Chaleur”, Paris, 1822.
11. B. M. ter Haar Romeny, Front-End Vision and Multiscale Image Analysis: Introduction to Scale-Space Theory. Dordrecht, the Netherlands: Kluwer Academic Publishers, in preparation.
12. B. M. ter Haar Romeny, ed., Geometry-Driven Diffusion in Computer Vision. Dordrecht: Kluwer Academic Publishers, 1994.
13. B.M. ter Haar Romeny, L.M.J. Florack, J.J. Koenderink, M.A. Viergever: Scale-Space Theory in Computer Vision. Proc. First Intern. Conf., Lecture Notes in Computer Science 1252, Springer Verlag, Berlin, 1997.
14. B. M. ter Haar Romeny, L. M. J. Florack, A. H. Salden, and M. A. Viergever, “Higher order differential structure of images,” Image and Vision Computing, vol. 12, pp. 317--325, July/August 1994.
15. B. M. ter Haar Romeny, W. J. Niessen, J. Wilting, and L. M. J. Florack, “Differential structure of images: Accuracy of representation,” in Proc. First IEEE Internat. Conf. on Image Processing, (Austin, TX), pp. 21--25, IEEE, November, 13-16 1994.
16. D. Hilbert, Theory of Algebraic Invariants. Cambridge Mathematica Library, Cambridge University Press, 1890. Reprinted lecture notes 1897.
17. D. H. Hubel, Eye, Brain and Vision, vol. 22 of Scientific American Library. New York: Scientific American Press, 1988.
18. D. H. Hubel and T. N. Wiesel, “Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex,” Journal of Physiology, vol. 160, pp. 106--154, 1962.
19. D. H. Hubel and T. N. Wiesel, “Brain mechanisms of vision,” Scientific American, vol. 241, pp. 45--53, 1979.
20. R. A. Hummel and B. C. Gidas, “Zero crossings and the heat equation,” Tech. Rep. 111, New York Univ. Courant Institute of Math. Sciences, Computer Science Division, 1984.
21. M. Jägersand, “Saliency maps and attention selection in scale and spatial coordinates: An information theoretic approach,” in Proc. Fifth Intern. Conf. on Computer Vision, (MIT Cambridge, MA), pp. 195--202, IEEE, June 20-23 1995. Catalog number 95CB35744.

22. N. Karssemeijer, “Detection of stellate distortions in mammograms using scale-space operators,” in Proc. Information Processing in Medical Imaging 1995, pp. 335--346, 1995.
23. J. J. Koenderink, “The structure of images,” Biol. Cybern., vol. 50, pp. 363--370, 1984.
24. J. J. Koenderink and A. J. van Doorn, “Two-plus-one dimensional differential geometry,” Pattern Recognition Letters, vol. 21, pp. 439--443, May 1994.
25. J. J. Koenderink and A. J. van Doorn, “Representation of local geometry in the visual system,” Biol. Cybern., vol. 55, pp. 367--375, 1987.
26. J. J. Koenderink, “Design principles for a front-end visual system,” in Neural Computers (R. Eckmiller and C. v. d. Malsburg, eds.), Springer-Verlag, 1987. Proceedings of the NATO Advanced Research Workshop on Neural Computers, held in Neuss, Fed. Rep. of Germany, September 28--October 2.
27. J. J. Koenderink and A. J. Van Doorn, “Operational significance of receptive field assemblies,” Biol. Cybern., vol. 58, pp. 163--171, 1988.
28. T. Lindeberg: Scale-Space Theory in Computer Vision. Kluwer Academic Publishers, Boston, 1994.
29. T. Lindeberg, “Scale-space for discrete signals,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 12, no. 3, pp. 234--245, 1990.
30. T. Lindeberg and L. Florack, “Foveal scale-space and linear increase of receptive field size as a function of eccentricity,” Tech. Rep. ISRN KTH/NA/P--94/27--SE, Dept. of Numerical Analysis and Computing Science, Royal Institute of Technology, August 1994.
31. C.W. Misner, K.S. Thorne, and J. A. Wheeler, Gravitation. San Francisco: Freeman, 1973.
32. O. Monga, N. Ayache, and P. T. Sander, “From voxel to intrinsic surface features,” Image and Vision Computing, vol. 10, pp. 403--417, July/August 1992.
33. P. Morrison, Powers of Ten: About the Relative Size of Things in the Universe. W. H. Freeman and Company, 1985.
34. J.L. Mundy and A. Zisserman, eds., Geometric Invariance in Computer Vision. Cambridge, Massachusetts: MIT Press, 1992.
35. M. Nielsen, L. M. J. Florack, and R. Deriche, “Regularization, scale-space, and edge detection filters,” in Proc. Fourth European Conference on Computer Vision, (Cambridge, UK), April 14-18 1996.
36. W. J. Niessen, K. L. Vincken, A. S. E. Koster, and M. A. Viergever, “A comparison of multiscale image representations for image segmentation,” in Proceedings IEEE Workshop on Mathematical Methods in Biomedical Image Analysis, (San Francisco), pp. 263--272, 1996.
37. W. J. Niessen, B. M. ter Haar Romeny, L. M. J. Florack, and M. A. Viergever, “A general framework for geometry-driven evolution equations,'' *International Journal of Computer Vision*, vol. 21, no. 3, pp. 187-205, 1997.
38. W. J. Niessen, B. M. ter Haar Romeny, and M. A. Viergever, “Numerical analysis of geometry-driven diffusion equations,” in Geometry-Driven Diffusion in Computer Vision (B. M. ter Haar Romeny, ed.), vol. 1 of Computational Imaging and Vision, pp. 393--410, Dordrecht: Kluwer Academic Publishers, 1994.
39. W. J. Niessen, J. S. Duncan, and M. A. Viergever, “A scale-space approach to motion analysis,” in Computing Science in the Netherlands 95 (J. C. Van Vliet, ed.), pp. 170--181, Stichting Mathematisch Centrum, Amsterdam, 1995.
40. P. Olver, G. Sapiro, and A. Tannenbaum, “Differential invariant signatures and flows in computer vision: A symmetry group approach,” in Geometry-Driven Diffusion in Computer Vision (B. M. ter Haar Romeny, ed.), Computational Imaging and Vision, pp. 255--306, Dordrecht: Kluwer Academic Publishers, 1994.
41. S. Osher and S. Sethian, “Fronts propagating with curvature dependent speed: algorithms based on the Hamilton-Jacobi formalism,” J. Computational Physics, vol. 79, pp. 12--49, 1988.
42. E. J. Pauwels, P. Fiddelaers, T. Moons, and L. J. van Gool, “An extended class of scale-invariant and recursive scale-space filters,” Tech. Rep. KUL/ESAT/MI2/9316, Catholic University Leuven, 1993.
43. P. Perona and J. Malik, “Scale-space and edge detection using anisotropic diffusion,” in IEEE Computer Society Workshop on Computer Vision, (Miami, FL), pp. 16--22, 1987.
44. S. M. Pizer, C. A. Burbeck, J. M. Coggins, D. S. Fritsch, and B. S. Morse, “Object shape before boundary shape: Scale-space medial axes,” Journal of Mathematical Imaging and Vision, vol. 4, no. 3, pp. 303--313, 1994.
45. G. Sapiro and A. Tannenbaum, “Affine shortening of non-convex plane curves,” Tech. Rep. 845, Technion Israel Institute of Technology, Department of Electrical Engineering, Technion, IIT, Haifa 32000, Israel, July 1992. EE Publication.

46. C. Schmidt and R. Mohr, “Object recognition using local characterization and semi-local constraints,” tech. rep., INRIA, 1996.
47. L. Schwartz, Théorie des Distributions, vol. I, II of Actualités scientifiques et industrielles; 1091-1122. Paris: Publications de l'Institut de Mathématique de l'Université de Strasbourg, 1950--1951.
48. L. Schwartz, Théorie des Distributions. Hermann & Cie, 1966.
49. J. A. Sethian, “A review of recent numerical algorithms for hypersurfaces moving with curvature dependent speed,” J. Differential Geometry, vol. 31, pp. 131--161, 1989.
    G. Sapiro and A. Tannenbaum, “On affine plane curve evolution,” Journal of Functional Analysis, vol. 119, pp. 79--120, January 1994.
50. S. M. Sherman and C. Kock, “Thalamus,” in The Synaptic Organization of the Brain (G. M. Shepherd, ed.), pp. 246--278, New York: Oxford University Press, 1990. Third Edition.
51. S. M. Sherman, “Dynamic gating of retinal transmission to the visual cortex by the lateral geniculate nucleus,” in Thalamic Networks for Relay and Modulation (D. Minciacchi, M. Molinari, G. Macchi, and E. G. Jones, eds.), pp. 61--79, Oxford: Pergamon Press, 1993.
52. J. G. Simmonds, A Brief on Tensor Analysis. Undergraduate Texts in Mathematics, Springer-Verlag, 1995. Second Edition.
53. J.-P. Thirion and A. Gourdon, “Computing the differential characteristics of isodensity surfaces,” Computer Vision, Graphics, and Image Processing: Image Understanding, vol. 61, pp. 109--202, March 1995.
54. K. L. Vincken, W. J. Niessen, A. S. E. Koster, and M. A. Viergever, “Blurring strategies for image segmentation using a multiscale linking model,” in IEEE Conference on Computer Vision and Pattern Recognition, CVPR'96, (San Francisco, CA), IEEE Computer Society Press, 1996.
55. J. Weber and J. Malik, “Robust computation of optical flow in a multi-scale differential framework,” in Proc. Fourth International Conference on Computer Vision, pp. 12--20, 1993.
56. J. Weickert, ‘‘Anisotropic diffusion in image processing,”. ECMI Series, Stuttgart: Teubner-Verlag, 1998.
57. J. Weickert, ``A review of nonlinear diffusion filtering,'' in *Scale-Space Theory in Computer Vision* (B. M. ter Haar Romeny, L. Florack, J. Koenderink, and M. Viergever, eds.), vol. 1252 of *Lecture Notes in Computer Science*, pp. 3-28, Springer, Berlin, 1997.
58. H. Weyl, The Classical Groups, their Invariants and Representations. Princeton, NJ: Princeton University Press, 1946.
59. R. Whitaker and G. Gerig, “Vector-valued diffusion,” in Geometry-Driven Diffusion in Computer Vision (B. M. ter Haar Romeny, ed.), Computational Imaging and Vision, pp. 93--134, Kluwer Academic Publishers B.V., 1994.
60. A.P. Witkin, “Scale-space filtering,” in Proc. International Joint Conference on Artificial Intelligence, (Karlsruhe, Germany), pp. 1019--1023, 1983.
61. R.A. Young, “The Gaussian derivative model for machine vision: Visual cortex simulation,” publication GMR-5323, General Motors Research Labs, Computer Science Dept., 30500 Mound Road, Box 9055, Warren, Michigan 48090-9055, July 7 1986.