CWI Quarterly Volume 11 (4) 1998, pp. 385 – 412

Linear and Non-Linear Scale-Spaces in Computer Vision [1]

Eric J. Pauwels
ESAT-PSI, Department of Electrical Engineering, K.U.Leuven
K. Mercierlaan 94, B-3001 Leuven, Belgium
e-mail: eric.pauwels@esat.kuleuven.ac.be

1. Motivation and background

1.1. Introduction

Computer vision is a multi-disciplinary research domain that aims to design systems capable of realising automatic, robust and intelligent image enhancement and interpretation. The intrinsic and unpredictable complexity of observed natural images means that — lots of effort and progress notwithstanding — we have barely begun to explore the tremendous potentials of this emerging technology. In this paper we will review some longstanding challenges in computer vision that have given rise to a particular set of mathematical problems in the theory of filter design and (non-linear) partial differential equations (PDE's).

Our aim is to give a flavour of the current research direction in this area of computer vision. Given the vast amount of available material and the limited size of this paper, emphasis will be on intuition rather than formal or rigorous derivations. As it is impossible to give a complete overview of this field, this presentation will necessarily be somewhat eclectic, the choice of topics often reflecting our own preferences. However, we have tried to provide sufficient pointers to the literature so that the interested reader can follow up strands that capture his attention.

1.2. The role of scale-spaces in computer vision

For the ease of discourse, computer vision is traditionally (and slightly artificially) divided into three main research areas:

1. **Low-level vision** focuses on operations at the pixel level. Examples that spring to mind are edge detection or morphology-based image restoration. What characterises these operations as low-level is their very limited

[1] The author gratefully acknowledges partial support by the Belgian Fund for Scientific Research (F.W.O. Vlaanderen), under grant G.0366.98.

spatial range: Typically, processing only depends on a small connected neighbourhood of each pixel at the time. Although the limitations of this myopic approach are obvious, it is nevertheless an exceedingly important first step, the main advantages being the high *processing speed* due to the *restricted connectivity.* Indeed, the insistence of low-level vision on small neighbourhoods means that local information exchange suffices to compute the output. In terms of implementation (e.g. on VLSI-chips) this is of paramount importance as it entails that each processor only needs to be connected to its immediate neighbours, thus ensuring easy scalability (compare this with the global connectivity in neural networks which is a major limiting factor for up-sizing them). But also from a mathematical point of view this has important consequences. For one thing, it motivates the requirement for a recursivity condition on the operators (see Subsection 1.4), allowing large-scale operators to be generated by successive applications of operators that have a much smaller support. Yet another albeit related issue linked to this restricted connectivity, is the central role played by differential equations as they too can be evaluated by comparing neighbouring values only.

2. **Intermediate-level vision** concerns itself with segmenting larger image regions and grouping them together based on features such as *colour, texture, shape, etc* .... Typical problems that are addressed comprise figure-ground separation, detection of regions-of-interest, object detection and shape recognition. Given the current state of technology, it looks as if the main role for this sort of processing is to be found at an intermediate triage level. At this stage one tries to figure out where in the image interesting or salient regions can be found. Once these regions have been delineated they can be passed on to more dedicated modules that are specialised in particular forms of processing or reasoning. Without this intermediate triage stage these higher levels algorithms would get bogged down by the huge amount of raw and diverse data.

3. **High-level vision** finally tries to combine all the strands of information to come up with a semantically correct interpretation of the image. Obviously, at this stage a lot of additional expertise often with more of an AI-flavour, needs to be brought to bear on the problem. Needless to say that algorithms addressing this level of processing are still very rudimentary.

Scale-space theories fit mainly into the category of low-level vision and can be situated in the broader context of *multi-scale analysis*, other examples of which are *pyramids, wavelets,* .... In the case of scale-space, the reasoning is as follows. Instead of linking up individual pixels at large distances (thus sacrificing the cherished local connectivity), the idea is to look at "king-sized" pixels by lumping together the information stored at individual pixels. Successive iterations of this process result in an image that is summarised at different *scales*, where all of the image-details are still present at the finest scale, while

they start disappearing at increasingly coarser scales. The rationale behind this is twofold:

- *Creations of a natural visual hierarchy:* the importance of structures is reflected in their *survival time* under these "summarisation"-operations. The motivation at the heart of this approach is that progressively increased processing will cause unimportant details to blend in with the background, thus focusing attention on the structure and their interrelations of the more prominent features.
- *Combining descriptions across scales yields additional information:* the information at high levels can be used to re-interpret information that might be ambivalent at the myopic pixel-level. E.g. local edge enhancement might be steered by the detection of an edge at a higher level. In the words of WITKIN [31]: "Coarse scale may be used to *identify* structures [extrema], and fine scale to *localise* them."

The importance of these points had long been recognised in computer vision and early attempts in this direction had resulted in constructions such as pyramids and quad-trees. In fact, most of the early work in scale-space theory was driven by the desire to come up with a principled (read "axiomatic") way to underpin some of the work going on in Gaussian pyramids. Although interesting in its own right, it has turned out that the main merit of this development was the introduction of the diffusion equation which provided an alternative way of looking at convolutions. Indeed, later on the emphasis shifted away from the filter-paradigm to take full advantage of the extra flexibility provided by the partial differential equations (PDE) framework. Researchers then took a more pragmatic stance and started looking at extensions of the diffusion equation that were designed with particular applications in mind.

In this paper we will retrace and highlight some of these developments. First, we will concentrate on an axiomatic approach to scale-space, where we will explore two sets of slightly different assumptions on filter families. The concept of the infinitesimal generator provides a natural stepping-stone to the corresponding evolution equation, of which the heat equation is the prototype associated with Gaussian filters. Further refinement of the PDE paradigm naturally leads to the seminal work of Perona and Malik that was the starting point for later contributions on geometry-driven diffusion.

1.3. Some definitions and notation

Let us begin by fixing some notation that will be used throughout the rest of this paper. An *image* u is defined to be a bounded function $u : D \subset \mathbb{R}^n \longrightarrow \mathbb{R}$ on some open domain D (for reasons of convenience we will usually take $D = \mathbb{R}^n$). The function u is called the *grey-value function* and in most cases $n = 2$, but especially in medical imaging, 3-dimensional images are becoming increasingly common-place. A point in the image domain D is often referred to as a *pixel*.

Starting from the original image u on $\mathbb{R}^n$ we want to construct a one-parameter family of derived functions $\overline{u}(x,t)$ (where $t \geq 0$) which are meant to represent the image at increasingly larger scales. Mathematically speaking this amounts to the determination of a family of *scale-space operators* $\{K_t;\ t \geq 0\}$ which will map the image u to a unique function of two arguments $\overline{u} : \mathbb{R}^n \times \mathbb{R}^+ \longrightarrow \mathbb{R}$, such that for each $t \geq 0$ fixed,

$$\overline{u}(x,t) = K_t(u(x)) \qquad \text{and} \qquad \overline{u}(x,0) = u(x).$$

Each of the derived images $\overline{u}(x,t)$ can be thought of as a "blurred" version of the original, with t measuring the amount of blurring. The function $\overline{u}$ on the $(n+1)$-dimensional half-space $\mathbb{R}^n \times \mathbb{R}^+$ will be called *scale-space* associated with $u(x)$ and generated by the family of scale-space operators $\{K_t;\ t \geq 0\}$. Moreover, we will hereafter no longer distinguish between u and $\overline{u}$; this should pose no problems. Also, adopting common engineering terminology, the operators K_t are often referred to as a filter family.

1.4. Connection with semi-group theory

In the remainder of this paper a number of conditions will be spelled out that to a large extent will determine the structure of the scale-space operators. These conditions will vary depending on the particular point of view of what the first stages in visual processing (the so-called "front-end") should accomplish. However, one condition that crops up in all scale-spaces is the *recursivity principle* (see e.g. [1]). This finds its origin in the fact that it seems reasonable to assume that the blurring-process should have an additive property: using an already slightly blurred picture and adding more blurring should yield the same result as subjecting the original image to the cumulated amount of blurring. Recall that this ties in with the underlying philosophy of low-level vision that operations at larger scales can be obtained by concatenating filters at smaller scales.

In terms of the operators K_t this amounts to the *semi-group property*:

$$K_t K_s = K_{t+s} \qquad \text{for all } t, s \geq 0, \tag{1}$$

which (in conjunction with $K_0 = I$) turns the operator family $\{K_t;\ t \geq 0\}$ into an *(additive) semi-group*. This semi-group property is extremely useful as it allows the introduction of the important concept of an *infinitesimal generator* which links the operators to an evolution equation, a link that has been the inspiration for much of the more recent work on image enhancement (see Sections 4 and 5).

We remind the reader that the infinitesimal generator is obtained by "differentiating" the semi-group with respect to its parameter t. More precisely, we notice that since the semi-group condition implies that $u(t) = K_t(u_0) = K_{t-s}K_s(u_0) = K_{t-s}u(s)$, it follows that (for $h > 0$)

$$u(t+h) - u(t) = (K_{t+h} - K_t)u_0 = (K_h - I)K_t u_0 = (K_h - I)u(t).$$

If we now divide by h and let $h \to 0^+$ we get

$$\frac{\partial u}{\partial t} = Lu \qquad \text{with initial condition} \qquad u(0) = u_0 \tag{2}$$

where (at least formally)

$$L := \lim_{h \to 0^+} \frac{K_h - I}{h} \tag{3}$$

is defined to be the *infinitesimal generator* of the semi-group K_t ($t \geq 0$). If this limit exists, it is possible to give a precise meaning to the formal exponential relation $K_t = \exp(tL)$ between the two operators. In its turn this can be used to confirm that

$$\frac{\partial u}{\partial t} = \frac{\partial}{\partial t}(K_t u_0) = \frac{\partial}{\partial t}(e^{tL} u_0) = L e^{tL} u_0 = Lu,$$

whence we can conclude that (2) is indeed the appropriate evolution equation for the transformation $u(t) = K_t u_0$.

The difficulty of course, is to find an appropriate topology and corresponding conditions under which the above operator limit (3) is well-defined. This is by no means a trivial problem and has been extensively studied in the literature on functional analysis (see e.g. YOSIDA [33] or HILLE & PHILLIPS [9]). Suffice it to say that as far as linear scale-spaces are concerned, the operators are sufficiently well-behaved so that the relevant evolution equation can be worked out (also see [1]).

2. LINEAR GAUSSIAN SCALE-SPACES

2.1. Zero-crossings as edge-detectors

Witkin was the first to forge a link between the well-behaved evolution of inflection points and convolution with a Gaussian kernel. His original contribution [31], which was further elaborated in BABAUD *et.al* [3], focused on 1-dimensional signals $u_0(x)$ for which he created a multi-scale representation by convolution with a Gaussian kernel:

$$u(x,t) = G_t * u_0(x) \qquad \text{where} \qquad G_t(x) = \frac{1}{\sqrt{2\pi t}}\, e^{-x^2/2t}.$$

He was especially interested in zero-crossings of the second derivative $\partial^2 u / \partial x^2$ and observed that in this scale-space representation the number of zero-crossings never increased.

To explain this longstanding pre-occupation of the computer vision community with zero-crossings and scale-spaces, consider the simplest version of the problem of edge-detection, arguably the most elementary problem in image-processing. Let the function $u : \mathbb{R} \longrightarrow \mathbb{R}$ represent a 1-dimensional image (see Fig. 1). If u is a simple step-function (dashed curve), localising the edge is trivial. If the edge is blurred (left panel), localisation can still be achieved by

determining the inflection point (i.e. solving $u''(x) = 0$). However, in realistic cases, edges are both blurred and corrupted by noise (right panel), which means that the number of inflection points increases dramatically, complicating the edge-search. One way to distinguish inflection points due to genuine edges from those induced by noise, is to look at their stability under smoothing. Clearly, if the inflection point is persistent under increasing amounts of smoothing, the chances are that it represents a real image structure.

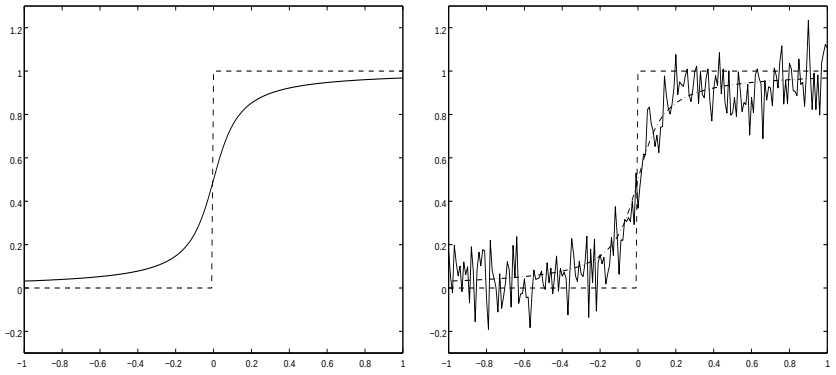

FIGURE 1. *Left:* Simple 1-dimensional step edge $u(x)$ (dashed) and blurred version (solid). *Right:* Typical edge profile (solid) after signal corruption due to blurring and noise.

Extension to two dimensions is straightforward if we realise that (if necessary, after an appropriate coordinate transformation) an edge can locally be modelled by a function $u(x, y)$ that has a 1-D edge in the x-direction while it is constant in the y-direction. Clearly, the edge locus of such a function is still determined by the vanishing of the Laplacian (a condition invariant under Euclidean transformations):

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0$$

or equivalently, for Gaussian smoothed edge:

$$\Delta(G_\sigma * u) = 0.$$

It is for this reason that much of the early research in scale-space theory focused on the structure of the zero-crossings for the Laplacian under various smoothing operators.

2.2. Selecting Gaussian kernels

As mentioned above, the link between the Gaussian convolution-kernels and the well-behaved evolution of inflection points was originally spotted by Witkin.

He proved that for 1-dimensional signals, the Gaussian convolution kernel is the only kernel (under mild additional restrictions) for which zero-crossings (of any order of derivative) are never created as scale increases. In terms of the underlying signal, this means that Gaussian smoothing never creates extrema ($u'(x) = 0$) or edges ($u''(x) = 0$) that are "spurious" in the sense that they are artifacts of the convolution-process rather than features of the underlying signal. Furthermore, as the scale-parameter increases these extrema tend to collapse and annihilate each other, creating a tree in scale-space that captures the structure of the image in terms of edges and bright or dark blobs. As the connection between the input image and the scale-space tree is a canonical one, his work suggested an efficient and natural way to summarise the primitive semantics underlying an image. However, when this line of thought was extended to arbitrary dimensions by YUILLE and POGGIO [32], it turned out that for these more realistic cases the results were far less satisfactory. We will discuss these issues in more detail below.

KOENDERINK [14] too argued that the Gaussian convolution kernel and its associated diffusion equation took up a central role in the creation of scale-spaces for images. He introduced the concept of *causality* which expresses that any grey-level at a coarser level can be traced back to a (not necessarily unique) "cause" at a finer resolution (although the reverse need not be true). In Koenderink's words: "no spurious detail should be generated when the resolution is diminished". HUMMEL [11] observed that this is equivalent to a version of the *maximum principle* for parabolic partial differential equations (PDEs). Therefore, apart from the heat- or diffusion-equation which represents the simplest of all possibilities, a number of (possibly non-linear) parabolic PDEs will satisfy Koenderink's criterion and might give rise to interesting computer vision algorithms. In sections 4 and 5 some likely candidates will be explored in more detail.

Let us now return to YUILLE and POGGIO'S paper [32] and sketch a brief outline of their exposition to give the reader a flavour of the sort of arguments used to single out Gaussian kernels. For reasons of simplicity we restrict ourselves to the 1-dimensional case, but the extensions to higher dimensions run along similar lines.

Consider the initial signal $u_0(x)$ that through convolution with a kernel $k_t(x) \equiv k(t,x)$ creates the scale-space signal $u(\cdot,t) = K_t(u_0) \equiv k(t,\cdot) * u_0$. Since we are interested in the evolution of the inflection-points, we introduce

$$E = \frac{\partial^2 u}{\partial x^2} = \frac{\partial^2 k_t}{\partial x^2} * u_0.$$

Yuille and Poggio now proceed by imposing a number of straightforward conditions on the kernel k_t:

1. **YP-1:** The filter at different scales is a simple rescaling of a fixed profile:

$$k_t(x) = \frac{1}{t}\,\kappa(x/t) \tag{4}$$

(for an extension of this, see the principle of scale-invariance in section 3).

2. **YP-2:** The kernel is symmetric about its centre which is independent of t. Otherwise, zero-crossings of a step edge would change their position with changes in scale.

3. **YP-3:** The filter vanishes at infinity ($|x| \longrightarrow \infty$) and recovers the whole image at sufficiently small scales:

$$\lim_{t\to 0+} K_t = I \text{ or equivalently, } \lim_{t\to 0+} k_t(x) = \delta(x).$$

These are rather technical assumptions and not very stringent. However, as Yuille and Poggio pointed out, what really clinches the argument is the condition that zero-contours of E should never be "created" as t increases (see also Fig.2). More precisely, at points (x_0, t_0) on such contours for which

$$E(x_0, t_0) = 0 \quad \text{and} \quad \frac{\partial E}{\partial x}(x_0, t_0) = 0$$

holds, the following *monotonicity condition* should be satisfied:

$$\frac{\partial E}{\partial t} \cdot \frac{\partial^2 E}{\partial x^2} > 0. \tag{5}$$

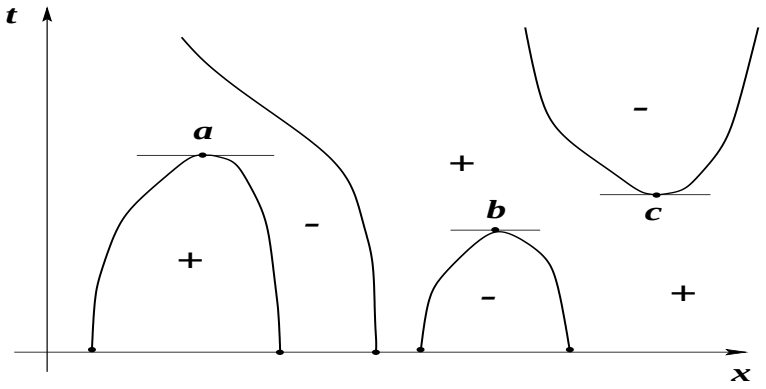

FIGURE 2. Behaviour of zero-crossings of E in scale-space. Observe that due to the signs of the regions bounded by the zero-crossings we can conclude that in a (b) both E_t and E_{xx} are negative (positive), whereas the configuration in c is excluded by the monotonicity condition.

As this condition must hold for any integrable input function u_0, the conditions on E can be pushed through to yield conditions on k_t. More precisely, the authors show that a judicious choice of the "test-function" u_0 reveals that there is a functional dependence between the derivatives of k:

$$\lambda_1 k + \lambda_2 \frac{\partial k}{\partial x} + \lambda_3 \frac{\partial^2 k}{\partial x^2} + \lambda_4 \frac{\partial k}{\partial t} = 0,$$

where the coefficients λ_i possibly depend on both x and t, and furthermore, λ_3 and λ_4 are of opposite signs. However, as the operator is shift-invariant, so is the PDE and therefore one can conclude that the coefficients cannot depend on the space variable x. Moreover, since t and x have the same dimensions (see condition YP-1), a straightforward dimensional analysis specifies the functional dependence of the coefficients in t:

$$\theta t \frac{\partial k}{\partial t} = \gamma t^2 \frac{\partial^2 k}{\partial x^2} + \beta t \frac{\partial k}{\partial x} + \alpha k, \tag{6}$$

where $\alpha, \beta, \gamma, \theta$ are numerical constants such that $\gamma/\theta > 0$. So it transpires that one can construct a counter-example unless the filter $k_t(x)$ satisfies the PDE (6). The rest of the proof hinges on the fact that straightforward changes of variables compatible with the above-mentioned conditions, can reduce the PDE (6) to the standard diffusion equation. Indeed, rescaling $k_t \longrightarrow e^{-\alpha/\theta} k_t$ will remove the term in k, while we can conclude that $\beta = 0$ since it only adds a shift-term (which is excluded by condition YP-2). We therefore end up with

$$\frac{1}{t}\frac{\partial k}{\partial t} = \frac{\gamma}{\theta}\frac{\partial^2 k}{\partial x^2}$$

which upon the introduction of $\tau = t^2/2$ and $\sigma^2 = 2\gamma/\theta > 0$ reduces to the standard diffusion equation:

$$\frac{\partial k}{\partial \tau} = \frac{\sigma^2}{2}\frac{\partial^2 k}{\partial x^2}. \tag{7}$$

The only solution to this equation that vanishes at infinity (cfr. YP-3) is indeed the Gaussian kernel:

$$k_t(x) = \frac{1}{\sqrt{2\pi\sigma^2\tau}} e^{-x^2/2\sigma^2\tau} = \frac{1}{\sqrt{\pi}\,\sigma t} e^{-(x/\sigma t)^2}.$$

Extending this reasoning to two or more dimensions the results by Yuille and Poggio can be summarised in the following theorem (for more details see [32]):

PROPOSITION 1. *For filters obeying conditions 1–3 and the* monotonicity condition *detailed above,*

1. *In one dimension, the Gaussian is the only filter which never creates zero-crossings of the second derivative as the scale increases;*
2. *In two dimensions, the Gaussian is the only filter which never creates zero-crossings of the Laplacian as the scale increases;*
3. *More generally, if L is a differential operator (in any dimension) that commutes with the diffusion equation, then solutions of*

$$L(k_t * u_0) = const$$

will not be created if and only if the filter is Gaussian. In particular, ridges or ravines (characterised by $a\partial_x u + b\partial_y u = 0$) which correspond to salient features in the image, are among those protected structures. However, in two or more dimensions local extrema such as peaks don't enjoy this property (see remark below).

Some remarks are in order here:

- Unfortunately, zero-crossings in 2 dimensions lack the simple tree-like behaviour that they exhibit in 1D. Although new zero-crossings cannot be created as t grows, they can both split and merge so that the number of zero-contours can both increase and decrease. For instance, consider the initial function
$$u_0(x, y) = (x^4 - x^2) + 2y^2 - \epsilon$$
(for $\epsilon > 0$ sufficiently small). This function has a saddle point at the origin. Therefore u_0 has only a single zero-contour. However, time evolution under the diffusion equation at the origin and at time $t = 0$ yields
$$\left.\frac{\partial u}{\partial t}\right|_{(0,0,0)} = 2$$
which will push the value at the origin up sufficiently so that zero-contour splits into two separated contours.

- It is also interesting to note how contrary to the situation in 1 dimension, diffusion in two dimensions can create new extrema. At first sight, this seems odd as diffusion is usually taken to be synonymous with smoothing, and creating a new extremum seems to run against the grain. The geometric intuition however, is quite straightforward. Consider two Gaussian-like peaks which are well-separated and of which the first (centred at A) is slightly higher than the second (centred at B). Now, further assume that the peaks of both Gaussian are connected by a narrow ridge. Clearly, this means that the peak at B is no longer an extrema, so that the only maximum of this structure is situated at A. Now we start diffusion on this initial configuration. Due to the slender structure of the ridge, erosion at C is far more drastic than at the other points of the structure, resulting in the creation of a saddle-point at C and an emerging maximum at B!

From an image processing point of view this last remark is particularly disappointing as it was originally thought that a stable image-structure could be built based on dark and bright "blobs", corresponding to regions around local extrema. As the above results show, the early expectations based on promising 1-dimensional results, failed to generalise to higher dimensions. For this reason, other avenues of research were explored some of which we will briefly discuss in the next sections. However, due to their attractive simplicity, Gaussian scale-spaces remain an area of active interest and the interested reader

is referred to e.g. FLORACK *et.al.* [7, 8], and especially LINDEBERG [18] and SPORRING *et.al.* [26] where he will find an excellent and comprehensive overview of both continuous and discrete Gaussian scale-spaces, along with numerous applications.

3. NON-GAUSSIAN RECEPTIVE FIELDS FOR LINEAR SCALE-SPACES

Since convolution with a Gaussian kernel in more than one dimension cannot guarantee that new extrema will not be created, we will in this section re-evaluate the conditions imposed on the front-end. In particular, if we assume (as we did in the previous section) scale-space filters to be *linear, isotropic convolution filters,* then two conditions (viz. recursivity and scale-invariance, see below) suffice to narrow down the collection of possible filters to a family that essentially depends on one parameter which determines the *qualitative* shape of the filter.

Gaussian filters correspond to one particular value of this shape-parameter. For other values the filters exhibit a more complicated pattern of positive ("excitatory") and negative ("inhibitory") regions. Intriguingly, this ties in with data from neurophysiology which show that receptive fields (read "filters") along the first stages of the visual pathway act as convolution operators, the kernels of which exhibit a centre-surround structure: an excitatory (inhibitory) centre is bordered by an inhibitory (excitatory) surround. Furthermore, in some cases additional concentric regions of alternating sensitivity have been observed [15, 16]. So it seems that a set of simple and straightforward axioms predicts the qualitative shape of filters that actually occur in the front-end of biological vision systems. (For more information on the neurophysiology of mammalian visual systems we refer to the excellent books by HUBEL [10] and ZEKI [34]).

Let us now retrace our steps and give a systematic overview of the conditions. We will concentrate on the one-dimensional case only, as the extension to more variables is straightforward. We will assume as before that at each scale t the corresponding operator K_t is a convolution-operator, the kernel of which will be denoted by $k_t(x)$. Following Yuille and Poggio's lead in the previous section, the following straightforward conditions are imposed on k_t:

A0: k_t is *mass-preserving*: $k_t * 1 = 1$ or equivalently $\int_{I\!R} k_t(x)dx = 1$, otherwise the filter would change a constant signal;

A1: k_t is even (the 1D-equivalent of rotation-invariance): $k_t(-x) = k_t(x)$;

A2: k_t is integrable ($k_t \in L^1$), otherwise the convolution would not be well-defined;

A3: k_t is a continuous function of both x and t (but not necessarily as a function of (x,t)).

An important consequence of the second condition is the existence of the

Fourier-transform of the kernel k_t:

$$\tilde{k}_t(\omega) \equiv \mathcal{F}(k_t(x)) := \int_{\mathbb{R}} e^{-ix\omega} k_t(x)\ dx.$$

Condition A1 implies that the Fourier transform is actually a cosine-transform of a real-valued function and therefore both even and real-valued:

$$\tilde{k}_t(\omega) = \int_{\mathbb{R}} \cos(\omega x) k_t(x)\ dx.$$

To pin down the form of the kernel we introduce two further conditions which capture the notion that blurring needs to be additive and scale-invariant.

1. **The Recursivity Principle**

   It seems reasonable to assume that the blurring-process should have an additive property: using an already slightly blurred picture and adding more blurring should yield the same result as subjecting the original image to the cumulated amount of blurring. As pointed out in the introduction, this is known in the literature as the *recursivity principle* (see e.g. [1]) and it turns the operator family into an additive semi-group:

   $$\forall\, t, s \geq 0 : K_t K_s = K_{t+s} \qquad \text{or equivalently,} \qquad k_t * k_s = k_{t+s}. \quad (8)$$

   Translated in terms of Fourier-transforms, the recursivity principle becomes:

   $$\tilde{k}_t \cdot \tilde{k}_s = \tilde{k}_{t+s}. \quad (9)$$

   Since we assumed continuity of k_t as a function of t, the same holds for $\tilde{k}_t$ and it therefore follows that the functional equation (9) has only one possible solution (see e.g. RUDIN [25] p.207-8):

   $$\tilde{k}_t(\omega) = e^{-g(\omega)t} \quad (10)$$

   where $g(\omega)$ is both real-valued and even (since $\tilde{k}_t(\omega)$ is). Moreover, since we postulated that $k_t(x) \in L^1$ it follows that $\tilde{k}_t(\omega)$ is continuous and vanishes at infinity (i.e. $\tilde{k}_t(\omega) \in C_0$) (cfr. RUDIN [25] p. 197). This implies that $g(\omega)$ is continuous and $g(\omega) \to \infty$ as $|\omega| \to \infty$.

   If we could prove that $g(\omega)$ increases sufficiently rapidly to allow us to conclude that $\tilde{k}_t$ is in fact *integrable* ($\tilde{k}_t \in L^1$), then we could invert the Fourier transform and the assumption that $k_t(x)$ is a continuous function of x entails that the inverse Fourier-transform of $\tilde{k}_t$ reproduces $k_t(x)$. We would then obtain the following (quite explicit) representation of the kernel:

   $$k_t(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{i\omega x} e^{-g(\omega)t}\, d\omega = \frac{1}{2\pi} \int_{\mathbb{R}} \cos(\omega x) e^{-g(\omega)t}\, d\omega. \quad (11)$$

   The second equality is due to the "even-ness" of $g(\omega)$. At this stage however such conclusions cannot yet be drawn as we still have an enormous amount

of leeway in the choice of the function g: basically, every even function of ω would do. The principle of scale invariance which is discussed in the next paragraph will further narrow down the possible choices for g.

2. **The Principle of Scale Invariance**

When there is no a priori information as to what structure in an image we are looking for, it follows that there should be no preferred scale. It therefore seems imperative that all kernel-functions should be *qualitatively identical*. If this wasn't the case, one could select a particular scale t by specifying that its corresponding kernel k_t made it stand out from the rest of the one-parameter family. This means that in order for the kernel k_t to be *scale invariant* there should exist a *fixed kernel function* (or *"parent-kernel"*) ϕ such that at different levels of the scale-parameter, k_t is a simple rescaling of this parent-kernel by means of a *rescaling-funtion* $\psi : \mathbb{R}^+ \longrightarrow \mathbb{R}^+$,

$$k_t(x) = k(x,t) = \frac{1}{\psi(t)}\,\phi\left(\frac{x}{\psi(t)}\right). \tag{12}$$

Two remarks are in order here.

Firstly: The rescaling introduces a new function ψ, extending the condition (YP-1) put forward by Yuille and Poggio (see above). This may seem strange at first but on second thoughts it becomes clear that there is no compelling reason why one should limit oneself to linear rescalings (i.e. $\psi(t) = t$) in which space- and scale-coordinates play an equivalent role. There are however some mild conditions which we will impose on ψ:

(a) Since we assumed continuity of k_t as a function of t we will also require ψ to be continuous;

(b) In view of its interpretation as an indicator of the scale we will assume that it is strictly increasing with $\psi(0) = 0$ and $\psi(t) \to \infty$ as $t \to \infty$. In particular, this implies that the inverse function ψ^{-1} exists and is a strictly increasing, continuous function.

Secondly: The global rescaling of the parent-kernel ϕ (i.e. the multiplication of ϕ by the coefficient $1/\psi(t)$) is necessary to make the integral

$$\int_{\mathbb{R}} k(x,t)\; dx = \int_{\mathbb{R}} \frac{1}{\psi(t)}\,\phi\left(\frac{x}{\psi(t)}\right)\; dx = \int_{\mathbb{R}} \phi(x)\; dx = 1$$

independent of the scale t. We point out that due to assumption A2 it follows that $\phi \in L^1$ and hence we can conclude that its Fourier-transform $\tilde{\phi}$ is well-defined. Notice moreover that the even-ness of $k_t(x)$ implies that of ϕ.

It is fairly straightforward to work out that scale invariance restricts the possible form of the function g in (10) to mononomials. More precisely, we have the following proposition:

PROPOSITION 2. *Let us assume that at each scale $t \geq 0$ the kernel-function $k_t(x) \in L^1(\mathbb{R}, dx)$ is mass-preserving, even (as a function of x) and continuous as a function of x and t (separately).*
If the family $\{k_t \mid t \geq 0\}$ satisfies both the recursivity principle (8) and is scale invariant (12), then its Fourier transform is equal to

$$\tilde{k}_t(\omega) = e^{-a\,|\omega|^\alpha\, t}, \tag{13}$$

whence,

$$k_t(x) = \frac{1}{2\pi} \int_{\mathbb{R}} cos(\omega x)\ e^{-a\,|\omega|^\alpha\, t}\ d\omega, \tag{14}$$

where $\alpha > 0$. This kernel family is in fact a rescaling of a fixed kernel ϕ:

$$k_t(x) = \frac{1}{\psi(t)}\ \phi\left(\frac{x}{\psi(t)}\right),$$

where the Fourier transform of ϕ has the form:

$$\tilde{\phi}(\omega) = e^{-a\,|\omega|^\alpha} \tag{15}$$

The corresponding rescaling-function is given by $\psi(t) = t^{1/\alpha}$.

For a proof we refer to PAUWELS *et.al* [20]. Notice that we can recover the "unscaled" kernel ϕ by putting $t = 1$;

$$\phi(x) = k_1(x) = \frac{1}{2\pi} \int_{\mathbb{R}} cos(\omega x) e^{-a\,|\omega|^\alpha}\ d\omega \quad \text{and} \quad \psi(t) = t^{1/\alpha}$$

The Gaussian kernel is obtained by taking $\alpha = 2$:

$$\phi(x) = \frac{1}{2\pi} \int_{\mathbb{R}} cos(\omega x) e^{-a\omega^2}\ d\omega = \frac{1}{\sqrt{2\pi\sigma^2}}\ e^{-x^2/2\sigma^2}, \qquad \text{where } \sigma^2 = 1/2a,$$

and $\psi(t) = \sqrt{t}$. This last result was also obtained by Yuille and Poggio in the preceding section and is well-documented in the literature, but from the above analysis it transpires that, contrary to what one might expect, recursivity and scale-invariance are *not* sufficient to single out the Gaussian kernel. In Fig. 3 we give the graphs (obtained by numerical Fourier inversion of (14) after setting $a = t = 1$) of the kernels for different values of the α-parameter. Finally, it should be mentioned that results based on a similar axiomatics were independently derived and published (unfortunately in Japanese) by T. IIJIMA [12]; a recent translation of this paper can be found in [30].

REMARKS

1. Apart from $\alpha = 2$, there are a few other values of the parameter for which Fourier inversion is possible in an analytically closed form. One such example is $\alpha = 1$, for $\tilde{\phi}(\omega) = e^{-|\omega|}$ is the Fourier transform of the Cauchy density. Hence ϕ and the kernel functions are given by

$$\phi(x) = \frac{1}{\pi}\,\frac{1}{1+x^2} \qquad \text{and} \qquad k_t(x) = \frac{1}{\pi}\,\frac{t}{t^2+x^2}.$$

2. Another example for which an analytic result is available, corresponds to the limiting value $\alpha = \infty$. This transpires from the observation that in (15)

$$\tilde{\phi}(\omega) \longrightarrow \begin{cases} 1 & \text{if } |\omega| < 1 \\ 0 & \text{if } |\omega| > 1 \end{cases} \qquad \text{as} \qquad \alpha \longrightarrow \infty.$$

   Up to a factor $1/\pi$, this is the Fourier transform of the sinc-function $sinc(x) \equiv (\sin x)/x$.

3. For most values of α however, we have to take recourse to numerical Fourier inversion to obtain a detailed description of the corresponding filter kernel. However, the functional form of the Fourier transform does provide us with some global information concerning the kernel, since there is a well-known relation between the (central) moments of a function and the derivatives of its Fourier-transform at $\omega = 0$:

$$\int_{I\!R} x^n \phi(x)\; dx = (-i)^n\; \tilde{\phi}^{(n)}(0). \tag{16}$$

   It is straightforward to check that if $\alpha > 2$, then ϕ is at least twice differentiable and $\phi''(0) = 0$. In view of (16) this implies that the second moment of ϕ vanishes and that ϕ therefore cannot be everywhere positive. Hence, when $\alpha > 2$ the kernel ϕ has zero-crossings. This is also borne out by the numerical inversions shown in Fig. 3.

4. Finally, it is worth mentioning that whenever the kernel has regions on which it becomes negative (as for $\alpha > 2$), the causality condition propounded by Koenderink (cfr. Subsection 2.2) is violated: i.e. grey-values exceeding the original range can be observed. This is most obvious near sharp discontinuities where a Gibbs-like phenomenon can be observed. However, a similar effect is also observed in experiments on visual perception: bright regions seem even brighter near boundaries with dark regions and vice versa. In psychophysics this observation is known as the Mach effect.

As a concluding aside we point out that although for each value of α the corresponding operator semi-group $K_t^{(\alpha)}$ has a an infinitesimal generator $L^{(\alpha)}$ such that the scale-space can formally be generated by the evolution equation

$$\frac{\partial u}{\partial t} = L^{(\alpha)} u \qquad \text{with} \qquad u(0) = u_0,$$

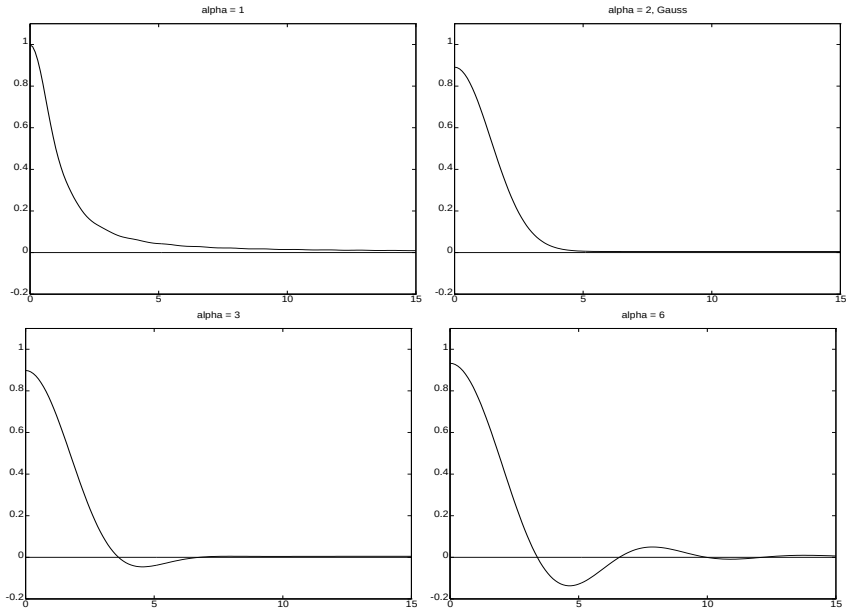

FIGURE 3. Graphs (for positive x-values) of the kernel function $\phi(x)$ for different values of parameter: $\alpha = 1, 2$ (top) and $\alpha = 3, 6$ (bottom). (A global rescaling of the y-axis by a factor $1/\pi$ has been discarded.) Notice the gradual emergence (for values $\alpha > 2$) of damped oscillations about the zero-level. In terms of *receptive fields* they correspond to "inhibitory" (i.e. *negative*) and "disinhibitory" (i.e. *positive*) regions flanking the "excitatory" centre.

only when $\alpha \in 2\mathbb{N}_0$ will this generator be a differential operator. In particular this means that insisting on an evolution equation that is (linear) PDE does not allow to recover the full gamut of filters k_t^α (for more details, see [20]). Also, it means that when α is not an even integer, the support of the filter cannot be restricted to a small neighbourhood which, in view of the remarks on restricted connectivity (see Section 1), is a serious drawback for a front-end filter.

4. PERONA-MALIK ANISOTROPIC DIFFUSION

One of the most important aspects of the the axiomatic approach to scale-spaces is the prominent rôle played by the diffusion equation, an equation that keeps cropping up in all sorts of guises. However, the limitations of this evolution equation are well-known. Surely, the most limiting in terms of practical usefulness is that although noise is gradually removed and the overall structure is simplified, edges are irrevocably blurred and displaced — from the point of view of image enhancement and interpretation a most unwelcome side-effect.

In an attempt to remedy some of the shortcomings of scale-spaces based on

the diffusion equation and other linear operators, Perona and Malik extended the scope of their investigations to include non-linear evolution equations. They enunciated three principles which in their view were crucial prerequisites for any front-end that intends to generate a "semantically meaningful" description of images:

1. **Causality:** Preventing the generation of "spurious detail", as described by KOENDERINK (see Subsection 2.2);
2. **Immediate localisation:** at each resolution region boundaries should coincide with semantically meaningful edges (i.e. edges should not shift away from the region boundaries to which they semantically belong);
3. **Localised smoothing:** smoothing should preferentially occur *inside* semantically meaningful regions, and *not across* their boundaries.

In their seminal paper [21] PERONA and MALIK claimed that a non-linear modification of the diffusion equation could achieve these three objectives. Their starting point was the observation that, in physics, the diffusion equation

$$\frac{\partial u}{\partial t} = c\,\Delta u \equiv c \sum_i \frac{\partial^2 u}{\partial {x_i}^2},$$

(with conductance c) is in fact obtained by combining a conservation law relating time-fluctuations in u to its (outward) flux $\mathbf{J}$,

$$\frac{\partial u}{\partial t} = -\mathrm{div}\ \mathbf{J},$$

with Newton's law, which links the flux to the gradient:

$$\mathbf{J} = -c\,\nabla u. \tag{17}$$

Large values for this conductance coefficient c indicate that the flux is highly sensitive to the gradient values, resulting in a fast diffusion. From this consideration it transpires that the divergence form

$$\frac{\partial u}{\partial t} = \mathrm{div}\,(c\,\nabla u). \tag{18}$$

is the most natural way to write the diffusion equation.

Perona and Malik's key insight was that this expression suggested a natural way to remedy some of the shortcomings that vexed ordinary diffusion. In particular, as the flux-equation (17) indicates how the speed of diffusion can be controlled by c they proposed to make c dependent on the *gradient* of the underlying image (grey-level function). In effect, large values for $\|\nabla u\|$ are indicative of edges, and one wants the diffusion at those locations to slow down

($c \to 0$), whereas in other regions a high c-value helps to iron out the noise. More concretely, they proposed to use a diffusion of the form

$$\frac{\partial u}{\partial t} = \operatorname{div}\left(c(\|\nabla u\|)\,\nabla u\right). \tag{19}$$

where $c(\cdot)$ is a symmetric, bell-shaped curve such as

$$c(v) = \frac{1}{1+(v/K)^2} \qquad \text{or} \qquad c(v) = e^{-(v/K)^2/2}. \tag{20}$$

Although this looks like an extremely attractive thing to do, the problem with this approach become obvious when we rewrite (19) as follows (for ease of notation we switch to the 1-dimensional case and denote $u_x = \partial u/\partial x$):

$$\begin{aligned}
\frac{\partial u}{\partial t} &= \frac{\partial}{\partial x}\left(c(u_x)\frac{\partial u}{\partial x}\right) = \frac{\partial}{\partial x}(c(u_x)\,u_x) \\
&= c(u_x)\,\frac{\partial^2 u}{\partial x^2} + c'(u_x)\frac{\partial^2 u}{\partial x^2}\,u_x \\
&= A(u_x)\,\frac{\partial^2 u}{\partial x^2},
\end{aligned}$$

where $A(p) = c(p) + pc'(p)$. The resemblance with the ordinary diffusion equation is only superficial as a straightforward computation immediately reveals that for the conductances in (20), the coefficient $A(u_x)$ becomes negative if the gradient exceeds the threshold $|u_x| > K$. The evolution equation then turns into the inverse heat equation, rightly notorious for its unstable behaviour. In fact, Catté *et.al.* [5] showed that positivity of the coefficient $A(p) = c(p) + pc'(p)$ is both necessary and sufficient to prove existence and uniqueness of a solution for the Perona-Malik equation (19) at all times $t > 0$. This, of course, is rather discouraging as it defies the original intent of the set-up.

Nevertheless, researchers in computer vision payed little heed to these less than promising mathematical forebodings and forged ahead anyway, often with surprisingly good and stable results. This was also explained in [5] where it was shown that the Perona-Malik equation could be regularised by introducing an arbitrarily small amount of smoothing with a Gaussian kernel G_σ of width σ before evaluating the conductance coefficient:

$$\frac{\partial u}{\partial t} = \operatorname{div}\left(c(\|\nabla G_\sigma * u\|)\,\nabla u\right), \qquad \text{with } \sigma > 0 \text{ arbitrarily small.} \tag{21}$$

Since discretizing in space implicitly involves some form of averaging, this suggests that the discrete implementations of (19) might intrinsically be more stable than is to be expected on the basis of theoretical results (a rigorous and in-depth discussion of the relevant mathematical framework can be found in Weickert's recent book [29]). However, as pointed out in [2], the hope that

a canonical weak solution can be obtained by letting $\sigma \to 0+$ proved to be erroneous as it turned out that there isn't a sufficient amount of stability to obtain a well-defined limit.

Considerable progress was made when work by a Paris group headed by MOREL and LIONS [2] showed how these equations could be modified so that they allowed so-called *viscosity solutions*. Unfortunately, a detailed discussion of this latter concept lies outside the scope of this review paper. Suffice it to say that viscosity solutions represent a generalised form of PDE solutions for which the differentiability conditions are relaxed so that merely continuous functions qualify as candidate solutions. However, in contradistinction to other forms of generalised solutions (such as distributions), viscosity solutions as a class are still sufficiently restricted so that in many cases *both existence and uniqueness* results can be proven.

In essence, the Paris group concentrated their efforts on the following basic model (along with some of its variations):

$$\frac{\partial u}{\partial t} = g(\|G_\sigma * u\|)\, \|\nabla u\|\, \mathrm{div}\left(\frac{\nabla u}{\|\nabla u\|}\right). \tag{22}$$

Although this formulation looks slightly daunting, a closer look reveals the simple geometric intuition behind it.

1. First of all, it is easy to check that introducing a new orthonormal (local) coordinate system (ξ, η) where η is oriented along the gradient ∇u while ξ is orthogonal to it, allows us to re-express the differential operator in its simplest form:

$$\|\nabla u\|\, \mathrm{div}\left(\frac{\nabla u}{\|\nabla u\|}\right) \equiv \frac{u_y^2 u_{xx} - 2u_x u_y u_{xy} + u_x^2 u_{yy}}{u_x^2 + u_y^2} = \frac{\partial^2 u}{\partial \xi^2}.$$

   From this it transpires that the evolution (22) in fact represents a degenerate diffusion where smoothing occurs along isophotes (lines of equal grey-value) and not across them. As a consequence, edges delineating salient regions will be spared.

2. The coefficient g controls the speed of the diffusion and is chosen to be a smoothly decreasing positive function such that $g(0) = 1$ and $g(p) \longrightarrow 0$ as $p \longrightarrow \infty$. As a consequence diffusion near edges is slowed down, providing extra protection.

For further elaboration of the mathematical details of this and related models (in particular, the PDE that generates the scale space of affine morphology), we refer to the original papers [1, 2], as well as to the excellent overview by DERICHE and FAUGERAS [6]. More in-depth discussions with a number of applications can also be found in WEICKERT [29].

5. SIDE's: Stabilized Inverse Diffusion Equations

5.1. Motivation and definition

The seminal work by Perona and Malik resulted in a flurry activity with surprisingly good results (see e.g. [27] which gives an overview of some of the more recent developments in the field.) Nevertheless, the nagging questions about the theoretical underpinnings of these experimental results remained. Most of the mathematical work that aimed to provide these evolution equations with sound foundations, used some form of regularisation that shied away from the reversed heat equation generated by the divergence formulation of the original Perona-Malik equation.

A recent and promising development that is much closer in spirit to the original PM-scheme, and that is striking in its simplicity, was proposed by Pollak, Willsky and Krim [22]. They got their inspiration from the study of dynamical systems in control theory. Their contribution uses a semi-discrete version of the PDE (continuous in time, discrete in space) which results in a system of ordinary differential equations (ODE's) with a discontinuous driving force. For the sake of notational convenience we will expound their theory for a 1-dimensional signal but extensions to higher dimensions are straightforward.

Starting from a signal $u(x,t)$, where as usual t is the time/scale parameter, we obtain the semi-discrete version by discretizing the space-variable to get a n-dimensional time-dependent vector $(u_1(t), u_2(t), \ldots, u_n(t))$ which, by abuse of notation, we will also denote by $u(t)$. The scale-space for the original signal is now created by the dynamics of the vector u in its n-dimensional state-space where it is driven by a (still to be specified) force field $\mathcal{F}$, giving rise to the dynamical system:

$$\frac{du}{dt} = \mathcal{F}(u), \qquad \text{and} \qquad u(0) = u_0.$$

Pollak and his co-authors extended this mechanical analogy even further by propounding that each of the elements u_k in the state-vector behaves like a small particle of mass m_k that is connected to its immediate neighbours only. More precisely, they assumed that the dynamics for each u_k is governed by an ODE of the form:

$$\frac{du_k}{dt} = \frac{1}{m_k}(F(u_{k+1} - u_k) - F(u_k - u_{k-1})) \tag{23}$$

where the functional form of the coupling force F still needs to be determined. Two observations can be made:

1. If we take $F(v) = v$, (23) collapses to $\dot{u}_k = u_{k+1} - 2u_k + u_{k-1}$, which means that we end up with the semi-discrete form of the standard diffusion equation. Similarly, taking $F(v) = -v$ yields the "inverted" diffusion equation.

2. More generally, recognising that the arguments of F are discrete incre-

ments, it is obvious that this is the discretisation of a PDE of the form:

$$\frac{\partial u}{\partial t} = \frac{1}{m}\frac{\partial}{\partial x}\left(F(\frac{\partial u}{\partial x})\right). \tag{24}$$

If F is differentiable then (24) can be rewritten as (putting $m(x) \equiv 1$)

$$\frac{\partial u}{\partial t} = F'(u_x)\frac{\partial^2 u}{\partial x^2}, \tag{25}$$

which explains why the authors call F a *diffusion force* whenever it is monotonously *increasing,* while they coin the term *inverse diffusion force* for a monotonously *decreasing* F.

It is obvious that taking $F(u_x) = u_x c(u_x)$ where c is one of the Perona-Malik coefficients defined in (20), reduces (24) to the the original model proposed by Perona and Malik.

From this second remark it transpires that (23) is a generalisation very close in spirit to the original Perona-Malik equation. Notice that for both force terms in (20), F is increasing (diffusion) for $|u_x| < K$ while decreasing (inverse diffusion) for $|u_x| > K$. However, what makes this scheme different from the original one is the specific form selected for F. The authors argue that for the force term to be interesting for its intended purposes, it should incorporate the following characteristics (see also Fig.4):

1. Since we want to enhance edges, the system should mainly behave as an inverse diffusion, implying that F should be decreasing over as large a domain as possible (for the PM-diffusion in (20), this occurs when $|u_x| > K$);

2. However, to get useful results stability of the evolution is of paramount importance, which implies that the dynamics need to incorporate some protection against enhancement at local extrema (which amounts to a maximum principle). More in detail we see that if for a arbitrary value u_k we denote the left and right increment as

   $$v_R = u_{k+1} - u_k \qquad \text{and} \qquad v_L = u_k - u_{k-1}$$

   then a local (strict) minimum is characterised by $v_L < 0$ and $v_R > 0$. Non-enhancement of the minimum then implies that the RHS of (23) satisfies $F(v_R) - F(v_L) \geq 0$, or again

   $$v_L < 0 < v_R \Rightarrow F(v_L) \leq F(v_R). \tag{26}$$

   An identical conclusion can be drawn for the case of a local maximum.

To optimally combine these two conditions, Pollak *et.al* propose to take a continuous inverse diffusion force F_{id} (i.e. F_{id} is monotonically decreasing) such that

$$F_{id}(0) = 0 \qquad \text{and} \qquad \sup_{v \in I\!R} |F_{id}(v)| \leq C, \qquad \text{for some constant C .}$$

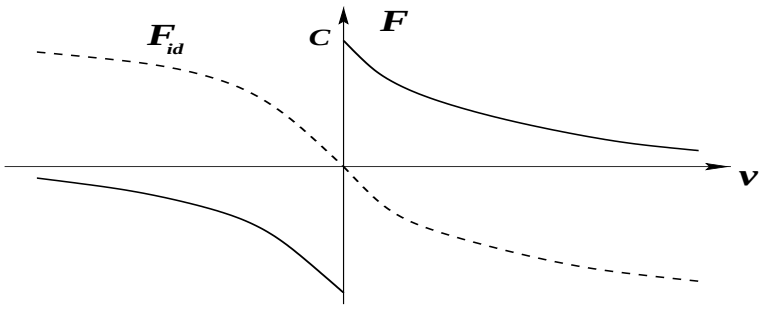

FIGURE 4. Proposed force term F and its relation to the inverse diffusion force F_{id}.

In its turn, the force term is then defined to be (see also Fig.4)

$$F(v) = F_{id}(v) + C\,\mathrm{sign}(v). \tag{27}$$

Notice that away from local extrema the dynamics now takes the form

$$m_k \frac{du_k}{dt} = F_{id}(v_R) - F_{id}(v_L),$$

emphasising the inverse diffusion character of ODE, whereas in a local minimum one has

$$m_k \frac{du_k}{dt} = F_{id}(v_R) - F_{id}(v_L) + 2C;$$

The first two terms in the RHS are dominated by the third, forcing the value u_k to increase (a similar remark holds at a local maximum). For this reason, the authors coined the name *stabilised inverse diffusion equation* (SIDE) for the evolution governed by the combination of (23) and (27).

However, the dynamics is still ill-defined at the point of discontinuity $v = 0$ of F. Here again, the choice is guided by a mechanical analogy: as soon as $u_k = u_{k+1}$ the two particles at those positions are linked together and for the further evolution are considered as one particle with a mass $m_k + m_{k+1}$. The higher mass imparts a greater stability to this newly formed region as is it will evolve more sluggishly. In terms of the original image, this means that image regions start coalescing as grey-values at neighbouring pixels become locked. From the point of view of image processing this make sense: large regions formed by coalescing contiguous pixels with similar grey-values should remain relatively stable in an image, whereas isolated pixels harbouring aberrant values are most probably due noise and should feel the full brunt of the evolution.

5.2. Properties of SIDE's

The introduction of discontinuities means that the dynamics of the system is changed abruptly as it hits lower-dimensional submanifolds characterised by

(sets of) equations of the form $u_k = u_{k+1}$; once it hits one of these submanifolds, it is forced to "slide" along its surface.

Comparing the form of the forcing term F, one notices that in a sense SIDE's are form of PM-diffusion for $K \longrightarrow 0$. However, SIDE's are decidedly well-behaved as borne out by the following properties (for more details we refer to [22]).

PROPOSITION 3.

1. *Unique solution and finite evolution time: For any $u_0 \in \mathbb{R}^n$, a SIDE started at u_0 has a unique solution and a unique equilibrium point $\overline{u}$ (with $\overline{u}_1 = \ldots = \overline{u}_1 = \sum_i u_{0,i}/n$) which it reaches in finite time.*

2. *Maximum principle: Local maxima (minima) are decreased (increased) by SIDE's, whence*

$$|u_k(t)| < max_p|u_p(0)| \qquad for \;\; t > 0.$$

   In terms of image processing, this corresponds to Koenderink's original causality requirement (see Subsection 2.2).

3. *Well-posedness* **for one spatial dimension***: The solution of a SIDE depends continuously on the initial condition, in the sense that for any $\epsilon > 0$, there is a $\delta > 0$ such that $\|u_0^{(1)} - u_0^{(2)}\| < \delta$ implies that $\|u^{(1)}(t) - u^{(2)}(t)\| < \epsilon$ for $t > 0$.*

4. *Stability of hitting times* **for one spatial dimension** *: If $u(t)$ is a typical solution (i.e. it never hits two different sliding hyperplanes simultaneously), then all solutions with initial data sufficiently close to $u(0)$ will have the same ordering of hitting times for the sliding hyperplanes. In terms of the underlying scale-space, this means that the structure of the scale-space is stable under sufficiently small perturbations.*

It is important to realise that the last two properties cannot be proven when $u(x,t)$ depends on more than one spatial dimension. Nevertheless, numerical experiments indicate that deviations are slight and segmentation results don't seem to diverge perceptually. From the point of view of multi-scale analysis this is interesting as it suggest a natural way to construct a scale-space based on the hitting times. Indeed, at each hitting time two neighbouring regions coalesce, thus reducing the complexity of the image while defining an intrinsic hierarchy on the different regions. This evolution continues, merging ever larger and more salient parts of the image until (in a finite number of steps) it converges to a uniformly grey image. Examples of this can be seen in Figs. 5 and 6.

6. CONCLUSION AND OUTLOOK

In this paper we have tried to explain the rationale behind the concept of scale-space in computer vision and outlined the evolution of the ideas from

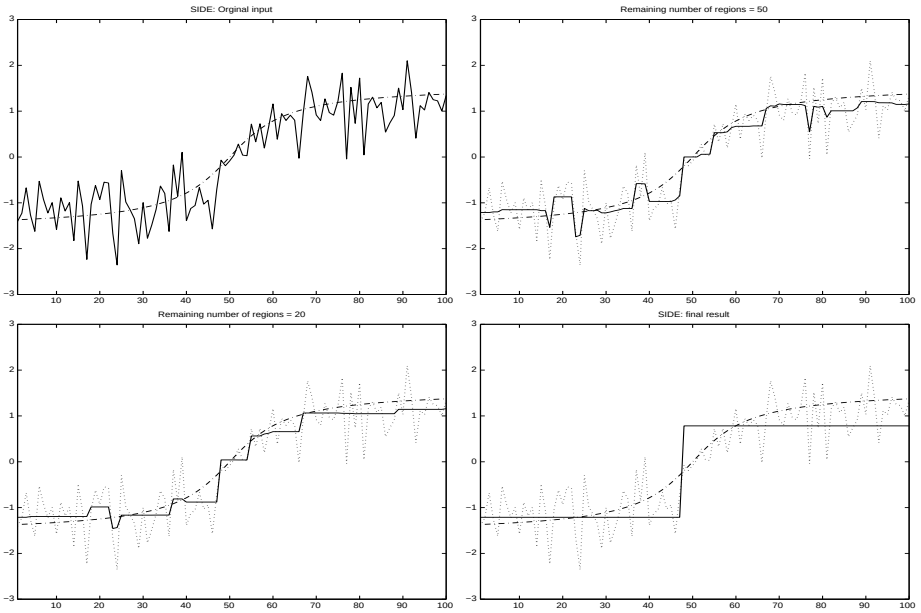

FIGURE 5. *Top left:* The original step-edge is corrupted by both blur and noise. *Bottom right:* Resulting scale-space signal (solid) when only two regions remain. *Top to bottom, left to right:* Some intermediate results (solid) during the scale-space processing.

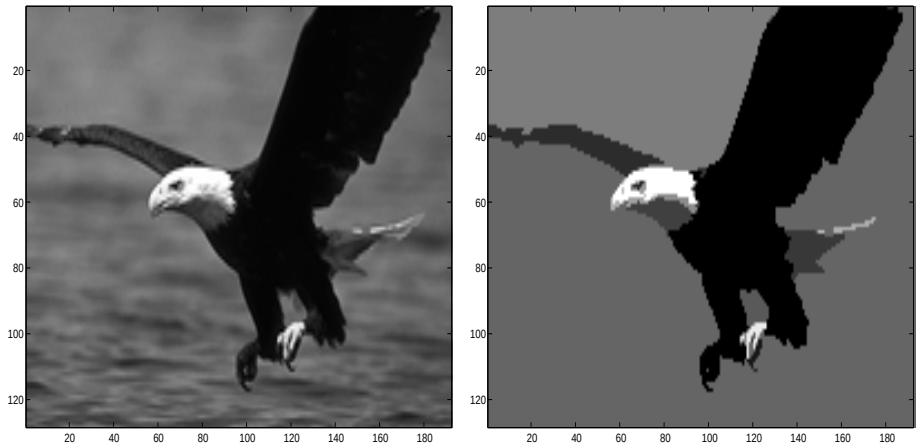

FIGURE 6. *Left:* Original grey-value input image. *Right:* The edge-preserving smoothing done by the SIDE has drastically reduced the number of regions while keeping the salient edges crisp. Segmentation of this processed image is now straightforward.

their axiomatic and rather rigid beginnings to the more pragmatic but very versatile PDE-approach. Given the limited space available it is only fair to say that we have barely scratched the surface, but we hope that we have given the reader some feel for the challenging problems that computer vision can suggest in areas such as non-linear operator theory and PDE's. Before signing off we would like to mention a few avenues of research in this area which we think hold much promise for future investigations. The interested reader is also invited to have a look at some of the more recent publications listed in the bibliography that reflect a wide and healthy variety of research activity (we especially recommend [4, 26, 13, 6, 28, 29]!).

- Experimental evidence shows that neither the bottom-up (working from pixel-level to interpretation) nor the top-down (checking interpretation by returning to the pixel-level) approach are *on their own* sufficiently performant. Combining evidence from both sides of the spectrum through some sort of feedback and input of prior information or believes, is almost always necessary to arrive at reliable results in an acceptable time. Given its dependence on local operations, the PDE-approach clearly lives near the pixel-level of the processing-domain. We are convinced that considerable progress is to be expected from the study of *integro-differential equations* of the form
$$\frac{\partial u}{\partial t} = L(u) + \int F(u)\ dx$$
(where L is one of diffusion operators discussed above), as this sort of generalised evolution equation combines both local interactions (via the differential operator L) and global information (via the integral operator).

- In the same vein, we expect that considerable progress can be made if we extend the SIDE-philosophy to include not one but several functions u that evolve simultaneously. In this way different image features (such as measures for grey-value, edge-ness, texture, etc....) can be coupled to interact in a non-linear way to counterbalance or reinforce one another. This idea has already been partially explored in e.g. PROESMANS *et.al.* [23] and RICHARDSON & MITTER [24], with very promising results, and these models stand to gain from the additional stability provided by SIDE's.

- An exciting area we haven't touched upon is the connection between differential equations and minimisation of energy functionals. In a number of cases, solutions of evolution equations can be interpreted as paths of steepest descent for an appropriate functional (see e.g. the work of Nordström [19] or WEICKERT [29]). Conversely, some problems are most naturally formulated as energy minimisations (e.g. segmentation) and quite frequently anisotropic and geometry-driven diffusion can be used to efficiently produce an excellent first approximation of the global minimum that then can be refined using more computationally expensive methods.

– Finally, it needs to be said that little is known about how consistent and unbiased all these diffusion schemes are with respect to recovering the "ground-truth". Put differently, if we have a known signal and we corrupt it with different sorts of noise, will these evolution equations *on average* converge to the correct underlying signal (recall the principle of *immediate localisation* propounded by Perona and Malik in section 4)? And if so, how quickly? Answers to these questions are very important if we want to use this methodology in sensitive application areas such as medicine or forensics.

Based on the vigorous research activity that can be witnessed in this particular area of computer vision, we are convinced that it will provide us with challenging and stimulating mathematical problems for many years to come.

Acknowledgements. The author would like to thank G. Frederix and G. Potoms for help with some of the implementations. The author also thanks two anonymous referees for a number of encouraging and helpful comments.

References

1. L. Alvarez, F. Guichard, P.L. Lions, J.M. Morel (1993). Axioms and Fundamental Equations of Image Processing. *Archive for Rational Mechanics and Analysis* **123**, No. 3, 199–257.
2. L. Alvarez, P.L. Lions, J.M. Morel (1992). Image selective smoothing and edge detection by nonlinear diffusion. *SIAM J. Numer. Anal.* **29**, No. 3, 845–866, June.
3. J. Babaud, A.P. Witkin, M. Baudin, R.O. Duda (1986). Uniqueness of the Gaussian Kernel for Scale-Space Filtering. *IEEE Trans. Pattern Anal. Machine Intell.* **8** (1), 26–33.
4. M.-O. Berger, R. Deriche, I. Herlin, J. Jaffré, J.-M. Morel (Eds.) (1996). *ICAOS'96: Images, Wavelets and PDEs.* Lecture Notes in Control and Information Sciences, **219**, Springer, London.
5. F. Catté, P.L. Lions, J.M. Morel, T. Coll (1992). Image selective smoothing and edge deteciton by nonlinear diffusion. *SIAM J. Numerical Analysis* **29** (1), 182–193.
6. R. Deriche, O. Faugeras (1996). Les EDP en Traitement des Images et Vision par Ordinateur. *Traitement du Signal* **13**.
7. L.M.J. Florack, B.M. ter Haar Romeny, J.J. Koenderink, M. Viergever (1992). Scale and the differential structure of images. *Image and Vision Computing.* **10**, 376–388. July/August.
8. L.M.J. Florack (1997). *Image Structure.* Kluwer, Dordrecht.
9. E. Hille, R.S. Phillips (1957). *Functional Analysis and Semi-groups.* American Mathematical Society Colloquium Publications, **31**.
10. D.H. Hubel (1988). *Eye, brain, and vision.* Scientific American Library No. 22.

11. R.A. Hummel (1986). Representations based on zero-crossings in scale space. *Proc. IEEE Conf. on Computer Vision and Pattern Recogn.* 204–209.
12. T. Iijima. Basic theory on normalization of pattern. *Bulletin of the Electrotechnical Laboratory* **26**, 368–388. (In Japanese)
13. (1998). *Special Issue on Partial Differential Equations and Geometry-Driven Diffusion in Image Processing and Analysis.* IEEE Trans. Image Proc. **7**, No. 3, G. Sapiro, J.-M. Morel, A. Tannenbaum (eds.).
14. J.J. Koenderink (1984). The structure of images. *Biol. Cybern.* 50, 363–370.
15. C.Y. Li, X. Pei, Y.X. Zhow, H.C. von Mitzlaff (1991). Role of the extensive area outside the X-cell receptive field in brightness information transmission. *Vision Res.* **31**, No.9, 1529–1540.
16. C.Y. Li, Y.X. Zhow, X. Pei, F.T. Qiu, C.Q. Tang, X.Z. Xu (1992). Extensive disinhibitory region beyond the classical receptive field of cat retinal ganglion cells. *Vision Res.* **32**, No.2, 219–228.
17. T. Lindeberg (1990). Scale-space for discrete signals. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **12**, No. 3, 234–254.
18. T. Lindeberg (1994). *Scale-Space Theory in Computer Vision.* Kluwer Academic Publishers.
19. N. Nordström (1990). Biased anisotropic diffusion — a unified regularization and diffusion approach to edge detection. *Image and Vision Computing* **8** (11) 318–327.
20. E.J. Pauwels, L.J. Van Gool, P. Fiddelaers, T. Moons (1995). An extended class of scale-invariant and recursive scale-space filters. *IEEE Trans. Pattern Anal. and Machine Intelligence* **17**, No. 7, 691–70, July.
21. P. Perona, J. Malik Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. Pattern Analysis and Machine Intelligence* **12** (7), 629–639, July.
22. I. Pollak, A.S. Willsky, H. Krim (1996). Image segmentation and edge enhancement with stabilized inverse diffusion equations. *Tech Report LIDS-P-2368* Laboratory for Information and Decision Systems, MIT, Oct 96.
23. M. Proesmans, E.J.Pauwels, L. Van Gool (1994). Coupled geometry-driven diffusion equations for low-level vision. B. ter Haar Romeny (Ed.): *Geometry-Driven Diffusion in Computer Vision.* Kluwer Academic Publishers.
24. T. Richardson, S. Mitter (1994). Approximation, computation, and distortion in the variational formulation. B. ter Haar Romeny (Ed.): *Geometry-Driven Diffusion in Computer Vision.* Kluwer Academic Publishers.
25. W. Rudin (1981). *Real and Complex Analysis.* McGraw-Hill, New York.
26. J. Sporring, M. Nielsen, L. Florack, P. Johansen (Eds.) (1997). *Gaussian Scale-Space Theory.* Kluwer, Dordrecht.
27. B.M. ter Haar Romeny (Ed.) (1994). Geometry-driven diffusion in com-

puter vision. *Computational Imaging and Vision* **1**, Kluwer Academic Publishers.
28. B.M. ter Haar Romeny, L.M.J. Florack, J.J. Koenderink, M. Viergever (Eds.) (1997). *Proc First Intern. Conf. on Scale-Space Theory.* Springer Lecture Notes in Computer Science, **1252**.
29. J. Weickert (1998). *Anisotropic Diffusion in Image Processing.* ECMI Series, Teubner-Verlag, Stuttgart, Germany.
30. J. Weickert, S. Ishikawa, A. Imiya (1997). On the history of Gaussian scale-space axiomatics. J. Sporring, M. Nielsen, L. Florack, P. Johansen (Eds.): *Gaussian Scale-Space Theory.* Kluwer, Dordrecht.
31. A.P. Witkin (1983). Scale space filtering. *Proc. IJCAI* Karlsruhe, W. Germany, 1019–1023.
32. A.L. Yuille, T.A. Poggio (1986). Scaling theorems for zero-Crossings. *IEEE Trans. Pattern Anal. Machine Intell.* **8**, 15–25,
33. K. Yosida (1968). *Functional Analysis,* Springer.
34. S. Zeki (1993). *A Vision of the Brain.* Blackwell Scientific Publications.