Foveal scale-space and the linear increase of receptive field size as a function of eccentricity

Tony Lindeberg[1] and Luc Florack[2]

[1]Computational Vision and Active Perception Laboratory (CVAP),
Department of Numerical Analysis and Computing Science,
Royal Institute of Technology, S-100 44 Stockholm, Sweden.
Email: tony@bion.kth.se

[2]3D Computer Vision Research Group,
Utrecht University Hospital, Heidelberglaan 100,
NL-3584 CX Utrecht, The Netherlands.
Email: luc@cv.ruu.nl

Technical report ISRN KTH NA/P–94/27–SE

Abstract

This paper addresses the formulation of a foveal scale-space and its relation to the scaling property of receptive field sizes with eccentricity. It is shown how the notion of a fovea can be incorporated into conventional scale-space theory leading to a foveal log-polar scale-space. Natural assumptions about uniform treatment of structures over scales and finite processing capacity imply a linear increase of minimum receptive field size as a function of eccentricity. These assumptions are similar to the ones used for deriving linear scale-space theory and the Gaussian receptive field model for an idealized visual front-end.

Keywords: fovea, retina, receptive field, scale-space, multi-scale representation, log-polar mapping, visual front-end, resolution, computer vision.

Contents

1 Introduction

When designing a computer vision system, a crucial decision has to be taken concerning the design of its sensory stage. Since we are obviously constrained by some finite processing capacity, one could decide to cover a large field of view with a coarse-scale detector. Alternatively, the resolving power could be improved by using a finer-scale detector at the expense of a smaller field of view. For a vision system aimed at solving a large number of visual tasks, both coarse-scale context information and high-resolution details may be needed. Including both capabilities as limiting cases naturally gives rise to a foveal design, which is the typical implementation in many mammalian vision systems.

A characteristic property of such a distribution is that the field of view is large and that the resolving power decreases with *eccentricity*, that is, the distance from the center of the visual field. For example, for human vision, psychophysical experiments by (van de Grind *et al.* 1986) indicate that velocity discrimination thresholds increase *linearly* with eccentricity. This scaling behaviour is usually explained as a basic "early vision" property, caused by a general increase of receptive field[1] sizes towards the periphery. More generally, the following constraints are known from psychophysics, neuroanatomy and electrophysiology (as summarized by (Koenderink and van Doorn 1978)):[2]

- The diameter of the smallest receptive field is proportional to eccentricity (Koenderink *et al.* 1978a, 1978b, 1978c).
- At any eccentricity all diameters greater than the corresponding smallest unit are present (Wilson 1970; Matin and Kornheiser 1976; Koenderink *et al.* 1978a, 1978b, 1978c).
- Mean receptive field size increases linearly with eccentricity (Fischer and May 1970).
- The transformation from the visual field to the cortex is logarithmic (Schwartz 1977) and the visual cortex seems rather homogeneous (Hubel and Wiesel 1974).
- At any retinal location many receptive field sizes are present but smaller fields are located more centrally (Hubel and Wiesel 1974; Fischer 1973).
- The relative overlap of receptive fields is independent of eccentricity (Fischer 1973).

A consistent explanation of all these characteristics has been given by (Koenderink and van Doorn 1978) on the basis of a "stack" of receptive fields laid out in a "sunflower-heart" distribution (van Doorn *et al.* 1972).

[1] Although physiological terms such as "receptive field", "retina", etc., are used in this paper, the visual field is not required to be two-dimensional (except in those cases where restrictions to the two-dimensional case are explicitly made). For example, one may think of a foveal vision system for three-dimensional medical images, such as computed tomography (CT) or magnetic resonance imaging (MRI) data, etc.

[2] Concerning more recent work, see also (Rovamo and Virsu 1979; Braccini *et al.* 1982; Wässle *et al.* 1990; Bijl 1991).

The purpose of this paper is to provide further theoretical support for this model and relate it to (linear) *scale-space theory*, which has been developed to handle image structures at different scales in a uniform way. We shall start from a basic assumption that underlies this theory, namely *scale-invariance* (or self-similarity), and combine that with the assumption about the existence of a *fovea*, that is, a preferred location where image structures are to be resolved with maximum resolution. Given these assumptions, we shall derive basic characteristics of a *foveal scale-space* and show how it relates to the logarithmic transformation of the visual field. In particular, we show that the linear increase of minimum receptive field size with eccentricity follows as a direct consequence.

In addition to the biological evidence, the resulting model is closely related to foveated sensors, which are highly useful for reducing the amount of data to be processed in computer vision systems (see *e.g.* (Porat and Zeevi 1988) for a review). One example of such a sensor is the "Retinal Vision System (RVS)" developed by (Sandini and Tagliasco 1980; Tistarelli and Sandini 1992). It is designed on the basis of a linear increase of the support regions of the photoreceptors, motivated by the empirical log-polar cortical remapping. Here, we shall show how such a design can be naturally motivated from similar assumptions as serve as axioms for determining the shape of the receptive field profiles in the theory of scale-space representation. Data acquired by such a foveated sensor serve as natural input for a foveal scale-space.

2 Background: Scale-space theory

A basic motivation for using a multi-scale approach when modelling early visual operations is the fact that image structure is intrinsically multi-scale; "objects" manifest themselves as meaningful entities only over certain ranges of scale.[3] Therefore, a fundamental property of a visual front-end system must be the ability to represent and process structures at different scales. In particular, in situations when no *a priori* information is available about what to expect in a scene, all scales should be treated in a similar manner. From these considerations scale-space theory has been developed as a systematic framework for dealing with image structures at different scales (Witkin 1983; Koenderink 1984b; Babaud *et al.* 1986; Yuille and Poggio 1986; Koenderink and van Doorn 1987, 1992; Lindeberg 1990, 1994b, 1994a; Florack *et al.* 1992, 1994a; Florack 1993). This theory basically states that the natural way to embed a given D-dimensional input signal $f\colon \mathbb{R}^D \to \mathbb{R}$ into a multi-scale representation is by convolving it with Gaussian kernels[4] (see figure 1)

$$g(x;\ \sigma^2) = \frac{1}{(2\pi\sigma^2)^{D/2}}\, e^{-(x_1^2+\cdots+x_D^2)/2\sigma^2} \tag{1}$$

and their derivatives

$$g_{x^\alpha}(x;\ \sigma^2) = \partial_{x_1^{\alpha_1}\ldots x_D^{\alpha_D}}\, g(x;\ \sigma^2) \tag{2}$$

[3] In this paper the term "resolution" is used to indicate the inverse of scale.

[4] Here, $x = (x_1, \ldots, x_D) \in \mathbb{R}^D$ is standard vector notation for an D-dimensional variable, and $\alpha = (\alpha_1, \ldots, \alpha_D) \in \mathbb{Z}^D$ constitutes so-called multi-index notation with $x^\alpha = x_1^{\alpha_1} \ldots x_D^{\alpha_D}$.

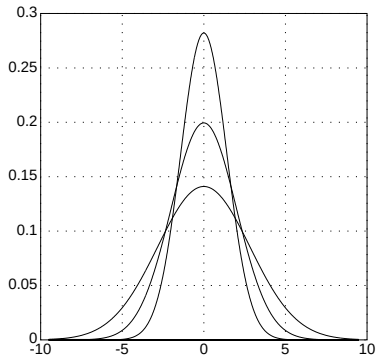

Figure 1: One-dimensional Gaussian kernels with different standard deviation $\sigma = \sqrt{t}$. The value of the scale-space representation of a signal at a certain point is the weighted average of the signal using the Gaussian weight function.

of various widths σ. The corresponding representation $L\colon \mathbb{R}^D \times \mathbb{R}_+ \to \mathbb{R}$

$$L(x;\ t) = (g(\cdot;\ t) * f)(x) = \int_{u \in \mathbb{R}^D} g(u;\ t_0)\, f(x-u)\, du \tag{3}$$

is called the scale-space representation of f. The convolution kernels $g_{x^\alpha}(x;\ t)$ give rise to the notion of *receptive fields* (of different orders α and at different scales $t = \sigma^2$), which sample the derivatives of the scale-space representation

$$L_{x^\alpha} = \partial_{x^\alpha} L \tag{4}$$

of f at any point $P = (x;\ t) \in \mathbb{R}^D \times \mathbb{R}_+$ in scale-space,

$$L_{x^\alpha}(x;\ t) = \int_{u \in \mathbb{R}^D} g_{x^\alpha}(u;\ t)\, f(x-u)\, du. \tag{5}$$

The output from such operations can then in turn be used as a basis for expressing a large number of early visual operations, such as feature detectors and image descriptors for deriving primitive shape cues. A particularly convenient framework for formalizing such processes is in terms of multi-scale differential geometric invariants and singularities of these (Koenderink and van Doorn 1987; Florack *et al.* 1993, 1994b; Lindeberg 1993a, 1994b; Johansen 1994).

Uniqueness of the Gaussian kernel. It has been shown that under plausible conditions on the first stages of visual processing the Gaussian kernel and its derivatives constitute a unique choice (see the references above). The conditions that specify uniqueness are basically linearity, shift invariance (homogeneity), rotational invariance (isotropy) combined with different ways of formalizing the notion that new structures should not be created by the smoothing operation and all scales should be treated in a uniform manner.

3 Foveal scale-space

Clearly, an actual realization of a visual front end based on (continuous) scale-space theory is limited by a finite processing capacity. Therefore, we have to select a finite number N of sample points $P^{(i)} = (x^{(i)};\ t^{(i)}) \in \mathbb{R}^D \times \mathbb{R}_+$ at which the scale-space representation is to be explicitly computed. These points correspond to (possibly overlapping) receptive fields with center locations $x^{(i)}$ and at scales $t^{(i)}$.

3.1 Assumptions

If we want to preserve uniform treatment of structures at different scales (scale invariance) then we have to take into account the following:

- For each fixed level of scale, $t^{(j)}$, the number $n^{(j)}$ of sample points $P^{(i)}$ corresponding to that level should be the same, *i.e.*, $n^{(j)} = n\,\forall j$.
- The discretization of the scale parameter should not introduce any bias towards any scale(s).

Of course, based on a finite number of samples, one cannot maintain spatial homogeneity, except possibly within a confined region. Therefore, it is natural to relax the strict homogeneity requirement of linear scale-space theory (where it applies to the overall spatial domain) by:

- For any level of scale $t^{(j)}$, the distribution of center locations $x^{(i)}$ should be homogeneous within a connected neighbourhood of a preferred point $O^{(j)}$.

Scale invariance implies that the preferred points $O^{(j)}$ should coincide for all scales. This gives rise to unique foveal point $F = O^{(j)}\,\forall j$. Because of this privileged point, it is natural to add one more symmetry requirement, namely:

- The foveal system should be rotationally symmetric with respect to F.

All of these assumptions are more or less self-explanatory except the second one. Concerning this, it can be shown that for continuous signals the only self-similar parametrization of scale is given by

$$\tau(\sigma) = A \log \frac{\sigma}{\sigma_0} \tag{6}$$

where $A \in \mathbb{R}$ and $\sigma_0 \in \mathbb{R}_+$ are arbitrary constants (Koenderink 1984b; Lindeberg 1993b; Florack *et al.* 1992). It is natural to use an equidistant sampling of τ. This corresponds to a constant ratio $\sigma^{(j+1)}/\sigma^{(j)}$ between adjacent scale levels when measured in terms of σ. (Without loss of generality, we may take $A = 1$.)

Layout of receptive fields. If we distribute the receptive fields according to the abovementioned assumptions, we obtain a foveal system as sketched in figure 2.

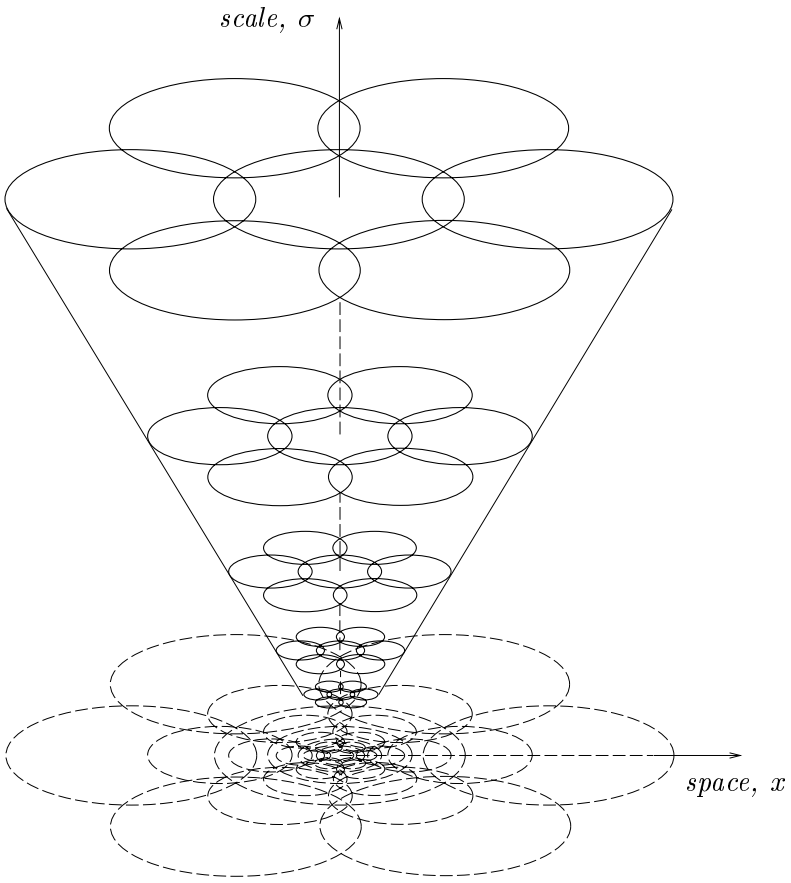

Figure 2: Schematic illustration of the (cone-like) distribution of receptive fields across scales as it results from the assumptions in section 3.1. Here, the relative overlap of neighbouring fixed-scale receptive fields has been chosen arbitrarily.

Interpretation in terms of normalized coordinates. A natural coordinate system to use when dealing with a scale-space representation is one in which length units are taken relative to scale. In such dimensionless coordinates

$$\xi = x/\sigma, \tag{7}$$

the cone in figure 2 simplifies to a cylinder (see figure 3), and both the scales of the receptive fields and the radius of the connected foveal neighbourhood occupied by the receptive fields at a certain scale will be constant across scales.

Note the similarity between this representation and the central region of the log-Cartesian fovea proposed by (Crowley *et al.* 1990; Crowley 1991); see also (Koenderink 1984a).

3.2 Linear increase of receptive field size

To simplify the treatment, let us consider the case where a continuous approximation is valid, and the distribution of receptive fields over space and scales can be treated as dense.

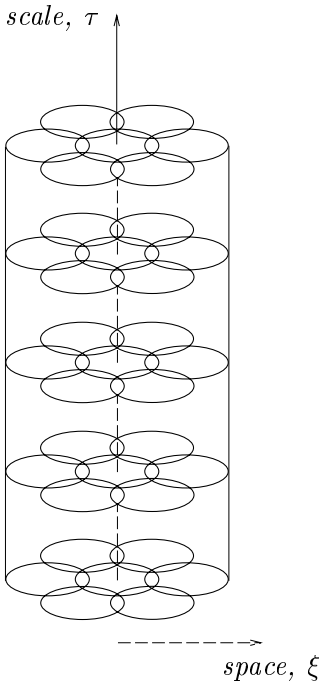

Figure 3: When expressed in terms of rescaled dimensionless coordinates, $\xi = x/\sigma$, the distribution of figure 2 becomes a cylinder, the axial symmetry of which reflects the scale invariance. Moreover, a self-similar scale sampling corresponds to uniform scale steps in effective scale.

A basic property of receptive fields is that a characteristic width r of the receptive field is proportional to the scale parameter σ, *i.e.*, $r \sim \sigma$. Given the constraints above, it follows that the receptive fields at scale σ will occupy a region of radius

$$R(\sigma) = \frac{\sigma}{\sigma_0}\varepsilon_0, \tag{8}$$

where ε_0 is the radius at a certain level of scale σ_0. The value of ε_0 is determined by the number n of receptive fields per scale level and their relative overlap (which constitute obvious freedoms of design).[5]

Let now x be the location vector for any given base point relative to the foveal center F. Obviously, at a distance $\varepsilon = |x|$ from the center of the fovea, the *minimum* scale, $\sigma_{min}(\varepsilon)$, that admits a full covering of a connected foveal neighbourhood $\{\xi \in \mathbb{R}^D : |\xi| \leq \varepsilon\}$ by $\sigma_{min}(\varepsilon)$-scaled receptive fields is given by

$$\sigma_{min}(\varepsilon) = \sigma_0 \hat{\varepsilon} \tag{9}$$

where $\hat{\varepsilon}$ denotes the relative eccentricity related to a certain ε_0

$$\hat{\varepsilon} = \frac{\varepsilon}{\varepsilon_0}. \tag{10}$$

[5] For different dimensions D, the ratio ε_0/σ_0 is related to the number of receptive fields per level n by $\varepsilon_0/\sigma_0 \sim n^{1/D}$.

Hence, in an idealized case, the width of a *minimum size receptive field* increases linearly with eccentricity ε. In an actual realization of a foveal system there is always a limiting minimum scale corresponding to the maximum resolution of the system as well as a limiting upper bound σ_{max}. If we take the minimum scale to be σ_0, then (9) can only hold when $\varepsilon \geq \varepsilon_0$, *i.e.*, outside a *foveal region*. Within this region, one can imagine a constant minimum receptive field size, $\sigma_{min}(\varepsilon) = \sigma_0 < \sigma_{max}$ (possibly combined with a smooth transition between these two qualitatively different scaling behaviours).

Note that this analysis holds independently of any specific sampling of scale, although self-similarity across scales requires an exponential sampling (see above). Moreover, the linear scaling behaviour is independent of spatial dimension D.

Figure 4 illustrates the result of sampling data in this way using minimum size receptive fields. Notice how fine-scale details are successively suppressed towards the periphery.

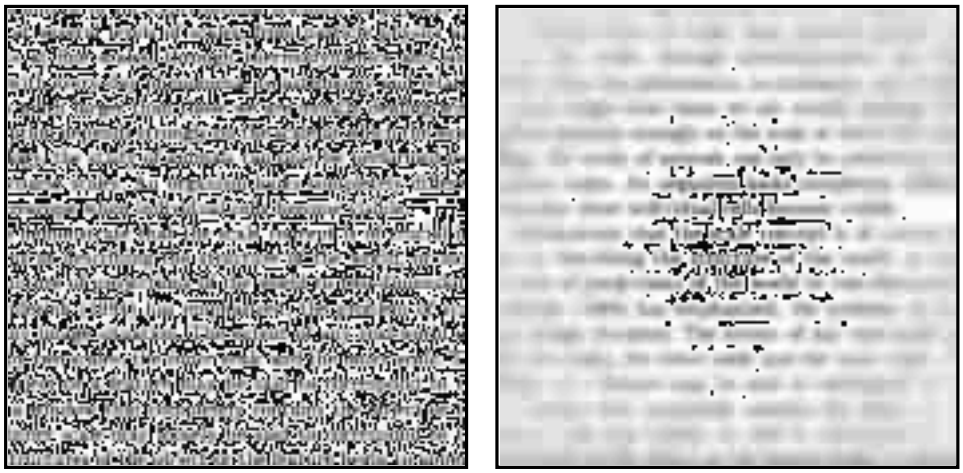

Figure 4: Left: A piece of text shown at a constant (high) resolution without smoothing ($\sigma = 0$). (The image size is 512×512 pixels.) Right: The result of registering these data with receptive fields whose scale values increase linearly with eccentricity. (At each corner the scale value is $\sigma = 8\sqrt{2}$ expressed in units of pixel length.)

3.3 Relation to the log-polar mapping

Schematically, the retinal distribution of minimum size receptive fields can (in the two-dimensional case) be illustrated by the well-known "sunflower model" as shown in figure 5(a). This distribution can be mapped onto a transformed domain where the receptive fields are laid out in a uniform way and are normalized to equal size (see figure 5(b)) by

$$\begin{aligned} x &= \lambda\, e^{\rho} \cos\theta, \\ y &= \lambda\, e^{\rho} \sin\theta, \end{aligned} \tag{11}$$

where $\lambda \in \mathbb{R}_+$ is an arbitrary constant. This is the well-known *log-polar mapping* (Schwartz 1977; Koenderink *et al.* 1978c) which transforms the visual field into a homogeneous domain such that rotations and rescalings relative to the foveal center correspond to mere translations. (Henceforth, we shall use the subscript LP to refer to the log-polar case and C to refer to the Cartesian case.)

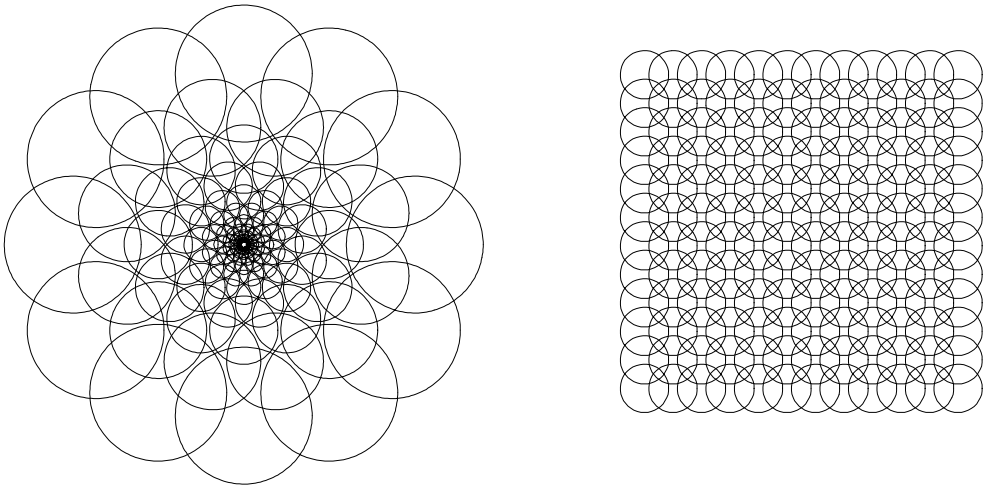

Figure 5: Left: Schematic illustration of the sunflower model for a visual front-end. Right: The log-polar transformation maps this distribution into a homogeneous one.

In terms of the distribution in figure 2, this log-polar transformation corresponds to mapping *the outer boundary* $\partial\Gamma$ of the cone Γ to a rectangular region with periodic boundary conditions (a strip on a cylinder). This set of data can be seen as the *finest level of scale* in a foveal multi-scale representation. Figure 6 shows the result of applying this log-polar transform to the data in figure 4(b).

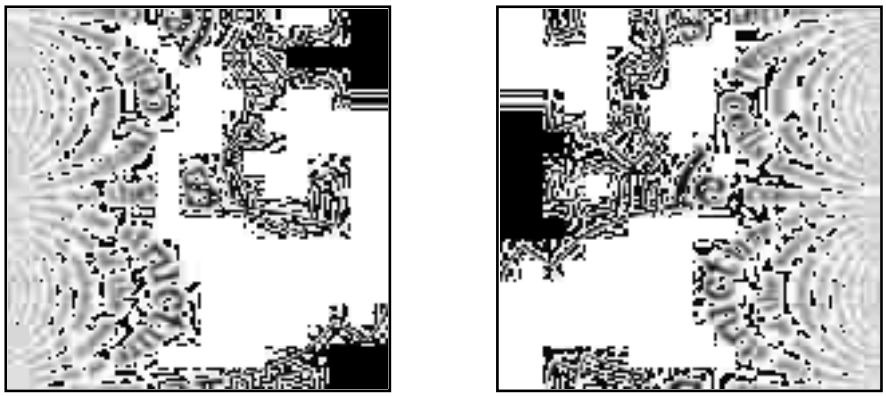

Figure 6: The result of mapping the data in figure 4(b) according to the log-polar transform (11). (The two images show the same data. The only difference is that the periodic boundaries have been differently treated to make the text easy easier to read.)

3.4 Multi-scale log-polar mapping

To describe the genuine multi-scale behaviour of the foveal system, we shall now consider the interior of the cone Γ as a parameterized family of surfaces corresponding to scaled versions of the outer boundary.

Consider a certain point $(x, y;\ t = \sigma^2)$ in the scale-space representation of a two-dimensional signal. In terms of differential entities, a natural choice for an infinitesimal line element in the Cartesian domain is

$$dl_C^2 = dx^2 + dy^2, \tag{12}$$

whereas in the log-polar domain a natural line element is

$$dl_{LP}^2 = d\rho^2 + d\theta^2. \tag{13}$$

These line elements are related by

$$dl_C = \varepsilon\, dl_{LP}, \tag{14}$$

where

$$\varepsilon^2 = x^2 + y^2. \tag{15}$$

The ratio dl_{LP}/dl_C corresponds, up to an irrelevant proportionality constant, to the "cortical magnification factor". In the case the log-polar mapping holds exactly, it is apparently equal to inverse eccentricity.

In a multi-scale representation, it is natural to measure distances relative to scale. With $\sigma_C = \sigma$ being the scale parameter in the Cartesian domain, and σ_{LP} being a corresponding scale-parameter in the log-polar domain, we can thus define normalized infinitesimal line elements by

$$\begin{aligned} d\tilde{l}_C &= \frac{dl_C}{\sigma_C}, \\ d\tilde{l}_{LP} &= \frac{dl_{LP}}{\sigma_{LP}}. \end{aligned} \tag{16}$$

By requiring these normalized line elements to be comparable ($d\tilde{l}_C = d\tilde{l}_{LP}$), it follows that the scale parameters must be related by

$$\sigma_C = \varepsilon\, \sigma_{LP}. \tag{17}$$

In other words, a fixed value of the scale parameter in the log-polar domain corresponds to a linear increase with eccentricity of the scale parameter in the Cartesian domain. Whereas σ_C parameterizes a family of horizontal slices in scale-space, σ_{LP} parameterizes a family of cones with the apex corresponding to the foveal singularity $(x, y\ \ t) = (0, 0;\ 0)$. Geometrically, σ_{LP} equals the slope of the cone it represents (see figure 7).

The combined mapping $(x, y;\ \sigma_C) \mapsto (\rho, \theta;\ \sigma_{LP})$ given by (11), (15) and (17) is the natural extension of the log-polar transform to a multi-scale representation.

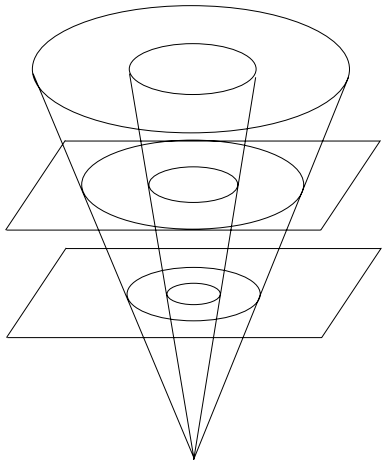

Figure 7: The Cartesian scale parameter $\sigma_C = \sigma$ parameterizes a family of horizontal slices in scale-space whereas the the log-polar scale parameter $\sigma_{LP} = \sigma/\varepsilon$ parameterizes a family of cones with the apex corresponding to the foveal singularity F.

3.5 Diffusion equation

The linear scale-space representation of a D-dimensional signal f_C defined from (3) can equivalently be defined as the solution of the diffusion equation in $\mathbb{R}^D$. In two dimensions we have

$$\partial_{t_C} L_C = \frac{1}{2}(\partial_{xx} L_C + \partial_{yy} L_C) \tag{18}$$

with initial condition $L_C(x, y;\ t_{C,0}) = f_C(x, y)$ where $t_{C,0}$ represents the sampling scale. A special property of the multi-scale log-polar transform is that it maps this equation to a similar one when expressed in terms of log-polar coordinates

$$\partial_{t_{LP}} L_{LP} = \frac{1}{2}(\partial_{\rho\rho} L_{LP} + \partial_{\theta\theta} L_{LP}). \tag{19}$$

Here, the initial condition is $L_{LP}(\rho, \theta;\ t_{LP,0}) = f_{LP}(\rho, \theta)$ where the initial data f_{LP} should correspond to data acquired with receptive fields that increase linearly with eccentricity (corresponding to the receptors marked by bold blocks in figure 8). In addition, L_{LP} is required to be 2π-periodic with respect to θ.

Disregarding the singularity, the diffusion operation and the multi-scale log-polar transform commute, so diffusion (and Gaussian convolution) can equivalently be done in either domain. Note that a log-polar scale increment Δt_{LP} corresponds to a Cartesian scale increment $\Delta t_C = \varepsilon^2 \Delta t_{LP}$ which tends to zero at the singularity.

When solving the linear diffusion equation, one may either iterate over the evolution parameter t according to some standard scheme or convolve the input data with the appropriate Greens function. In the Cartesian case, this is the

well-known normalized Gaussian, whereas in the log-polar case, the Green's function is a Gaussian function in the ρ-direction multiplied by an infinite sum of Gaussian functions with offsets $2n\pi$ ($n \in \mathbb{Z}$) in the θ-direction

$$g_{LP}(\rho, \theta;\ \sigma_{LP}^2) = g(\rho;\ \sigma_{LP}^2) \sum_{n=-\infty}^{\infty} g(\theta - 2n\pi;\ \sigma_{LP}^2). \tag{20}$$

(Here, g here denotes the one-dimensional Gaussian kernel according to (1).)

It turns out that the log-polar transform can also be derived in an alternative way by considering spatial transformations that preserve angular relationships, so-called *conformal mappings*. It can be shown (Florack 1993) that for two-dimensional signals the log-polar transform arises as the only non-trivial and physically plausible instance of such a mapping that is rotationally symmetric with respect to a certain point and preserves the (local) *form* of the linear diffusion equation (*i.e.* transforms (18) to (19)).

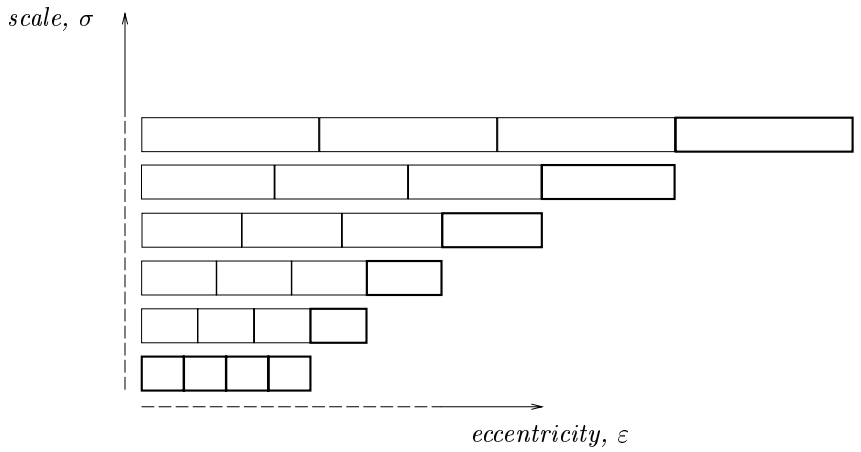

Figure 8: Schematic illustration of the linear increase of minimum receptive field size in the one-dimensional case. The bold blocks symbolize the measurement data corresponding to the finest scale $t_{LP,0}$ in the log-polar domain, that is, the initial data to the log-polar diffusion equation (19). To reduce aliasing problems, it is natural to implement these receptive fields as weighted local averages of outputs from sensors (photoreceptors) of higher resolution distributed with a higher density. This can be implemented in terms of a dense sunflower-heart sensor array. The implementation of the receptive fields corresponding to the non-bold blocks allows for an obvious freedom in design; they can be constructed either from finer-scale units (using cascade smoothing) or in the same way as the bold units (from direct connections to the photoreceptors).

3.6 Scale distribution of receptive fields at a fixed point

The linear scaling behaviour applies to the *minimum* size receptive fields with *varying* eccentricity. One may pose the question of what is the distribution over scales of the receptive field *ensemble* at any *fixed* eccentricity. Clearly, one may expect larger receptive fields to have a relatively larger probability at higher

eccentricity, but it is also clear that this probability depends upon the scale sampling, *i.e.*, upon how the receptive fields at any given point are distributed over scales.

Introduce $p_\sigma(\sigma;\varepsilon)\,d\sigma$ as the probability of finding a receptive field in the scale interval $[\sigma, \sigma + d\sigma]$ at eccentricity ε. Assume that the scale levels are parameterized by a dimensionless *effective scale parameter*, τ, such that the scale sampling of $\sigma = \sigma(\tau)$ corresponds to a uniform and equidistant sampling in τ. (Recall that scale invariance leads to (6).) At any point x at eccentricity $\varepsilon = |x|$ there are receptive fields in the scale range $[\sigma_{min}(\varepsilon), \sigma_{max}]$. Let $p_\tau(\tau;\varepsilon)$ be defined by

$$p_\tau(\tau;\varepsilon)\,d\tau = p_\sigma(\sigma;\varepsilon)\,d\sigma, \tag{21}$$

and let $T(\varepsilon) = \tau_{max} - \tau_{min}(\varepsilon)$ denote the available scale range at eccentricity ε (where τ_{max} and $\tau_{min}(\varepsilon)$ are the parameter values associated with σ_{max} and $\sigma_{min}(\varepsilon)$, respectively). In terms of the effective scale parameter, τ, we have

$$p_\tau(\tau;\ \varepsilon) = \frac{1}{T(\varepsilon)} \quad (\varepsilon \geq \varepsilon_0, \tau \geq \tau_{min}(\varepsilon)), \tag{22}$$

and in terms of the conventional scale parameter, σ,

$$p_\sigma(\sigma;\ \varepsilon) = \frac{\tau'(\sigma)}{T(\varepsilon)} \quad (\varepsilon \geq \varepsilon_0, \sigma \geq \sigma_{min}(\varepsilon)), \tag{23}$$

where $\tau'(\sigma)$ denotes the derivative of the mapping $\tau(\sigma)$ (the inverse of $\sigma(\tau)$). Within the foveal region, (22) and (23) apply by replacing $T(\varepsilon)$ by $T(\varepsilon_0)$.

Exponential scale sampling. In the case where the effective scale is given by (6), the probability density $p_\sigma(\sigma;\ \varepsilon)$ at any eccentricity decreases with σ as $1/\sigma$. The explicit expression is

$$p_\sigma(\sigma;\ \varepsilon) = \frac{1}{\sigma}\frac{1}{\log\left(\mu \min(1, 1/\hat{\varepsilon})\right)} \quad (\varepsilon \geq \varepsilon_0, \sigma \geq \sigma_{min}(\varepsilon)) \tag{24}$$

where μ is the ratio between the maximum scale and the minimum scale

$$\mu = \frac{\sigma_{max}}{\sigma_0}. \tag{25}$$

Note that the expression (24) is always well-defined, since the constraint $\sigma_{max} > \sigma_{min}(\varepsilon)$ implies $\max(1, \hat{\varepsilon}) < \mu$.

Average receptive field size at a given location. We can use the relations (22) and (23) for calculating ensemble averages in the different domains by

$$\begin{aligned} \bar{\tau}(\varepsilon) &= \int_{\tau=\tau_{min}(\varepsilon)}^{\tau max} \tau\, p_\tau(\tau;\ \varepsilon)\,d\tau, \\ \tilde{\sigma}(\varepsilon) &= \int_{\sigma=\sigma_{min}(\varepsilon)}^{\sigma max} \sigma\, p_\sigma(\sigma;\ \varepsilon)\,d\sigma. \end{aligned} \tag{26}$$

Straightforward calculations show that

$$\begin{aligned}\bar{\tau}(\varepsilon) &= \frac{1}{2}(\tau_{max} + \tau_{min}(\varepsilon)) = \frac{1}{2}\log\left(\mu \max(1, \hat{\varepsilon})\right), \\ \tilde{\sigma}(\varepsilon) &= \frac{\sigma_0(\mu - \max(1, \hat{\varepsilon}))}{\log\left(\mu \min(1, 1/\hat{\varepsilon})\right)}.\end{aligned} \tag{27}$$

For the purpose of comparison, let us define corresponding descriptors in the transformed domains by $\bar{\sigma}(\varepsilon) = \sigma_0 \exp \bar{\tau}(\varepsilon)$ and $\tilde{\tau}(\varepsilon) = \log \tilde{\sigma}(\varepsilon)/\sigma_0$. This gives

$$\begin{aligned}\bar{\sigma}(\varepsilon) &= \sqrt{\sigma_{max}\sigma_{min}(\varepsilon)} = \sqrt{\sigma_0 \sigma_{max}}\sqrt{\max(1, \hat{\varepsilon})}, \\ \tilde{\tau}(\varepsilon) &= \log \frac{\mu - \max(1, \hat{\varepsilon})}{\log\left(\mu \min(1, 1/\hat{\varepsilon})\right)}.\end{aligned} \tag{28}$$

Note that there is a conceptual difference between defining the ensemble average in the different domains, *i.e.*, in general we have $\bar{\sigma} \neq \tilde{\sigma}$ or, equivalently, $\bar{\tau} \neq \tilde{\tau}$ (see figure 9 for an illustration). Whereas $\tilde{\sigma}$ denotes the algebraic average, under the assumption of self-similarity $\bar{\sigma}$ corresponds to the geometric average and, in fact, also to the median.

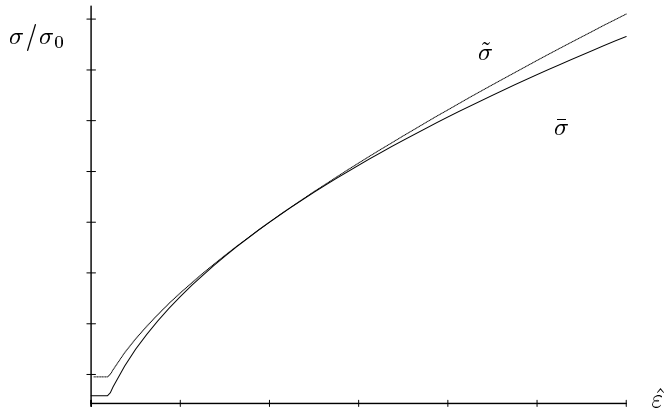

Figure 9: The variation with eccentricity of the ensemble averages of the receptive field sizes $\bar{\sigma}$ and $\tilde{\sigma}$ computed based on the effective scale τ and the linear scale parameter σ respectively.

In the limit case when the radius of the foveal region ε_0 tends to zero (corresponding to $\hat{\varepsilon} \to \infty$ outside the foveal singularity) the asymptotic behaviour of these entities is given by

$$\begin{aligned}\bar{\sigma}(\varepsilon) &= \sqrt{\sigma_0 \sigma_{max}}\sqrt{\hat{\varepsilon}}, \\ \tilde{\sigma}(\varepsilon) &= \sigma_0 \frac{\hat{\varepsilon}}{\log \hat{\varepsilon}/\mu}.\end{aligned} \tag{29}$$

so the ensemble average $\tilde{\sigma}$ increases faster with eccentricity.

Relations to biological vision. At first sight, the behaviour of $\bar{\sigma}(\varepsilon)$ and $\tilde{\sigma}(\varepsilon)$ may seem to conflict the results by (Fischer and May 1970) stating that the

average receptive field sizes increase linearly with ε. However, their averaging relates to a *different* type of ensemble. The entities $\bar{\sigma}$ and $\tilde{\sigma}$ express the average over the receptive fields that are *simultaneously* stimulated by a point stimulus at a certain visual location, hence *prior* to the log-polar mapping. Thus, these averages may have psychophysical relevance in situations when information over scales is combined.

This is to be distinguished from the ensemble of receptive field *centers* (corresponding to axons in a biological system), which have a $1/\sigma^2$-decreasing spatial density as a function of scale. This type of ensemble is relevant for electrophysiological cell recordings, where probing is typically performed in the homogeneous log-polar domain. For this ensemble, the average receptive field size does indeed increase linearly with eccentricity; see (Koenderink and van Doorn 1978).

3.7 Smooth transition from foveal to peripheral region

Due to the limitation on the minimum receptive field size, there is an inevitable deviation from the ideal model, which has a singularity at the foveal center. Previously, this problem was handled in an *ad hoc* way by letting the minimum receptive field size be constant within a finite foveal neighbourhood. This leads to a discontinuous transition from this foveal region to the periphery. Although it may not be a crucial problem in practical implementations, one might want to avoid such a discontinuity and require a smooth transition from a peripheral log-polar behaviour to a more regular (Cartesian) behaviour near the fovea.

A straightforward way to obtain smooth transition is as follows: The family of cones in figure 7 can be described by

$$C_{\sigma_{LP}}: \quad x^2 + y^2 = \varepsilon_0^2 \, (\frac{\sigma}{\sigma_0})^2 (\frac{\sigma_{LP,0}}{\sigma_{LP}})^2, \tag{30}$$

where σ_{LP} is the parameter and $\sigma_{LP,0}$ is the parameter value corresponding to the outer boundary. To avoid the singularity, we can replace the cones by a family of hyperboloids

$$H_{\sigma_{MLP}}: \quad x^2 + y^2 = \varepsilon_0^2 \left((\frac{\sigma}{\sigma_0})^2 (\frac{\sigma_{MLP,0}}{\sigma_{MLP}})^2 - 1 \right) \tag{31}$$

parameterized by a modified log-polar scale parameter $t_{MLP} = \sigma_{MLP}^2$ given by

$$t_C = \sigma_0^2 \, (1 + \hat{\varepsilon}^2) \, t_{MLP}. \tag{32}$$

Each $H_{\sigma_{MLP}}$ asymptotically approaches $C_{\sigma_{MLP}}$ with increasing eccentricity $\varepsilon = (x^2 + y^2)^{1/2}$. For each σ_{MLP} the hyperboloid has a smooth transition from

$$\sigma_C(0) = \frac{\sigma_{MLP}}{\sigma_{MLP,0}} \sigma_0 \tag{33}$$

at zero eccentricity to the asymptotic behaviour

$$\sigma_C(\varepsilon) = \frac{\sigma_{MLP}}{\sigma_{MLP,0}} \, \sigma_0 \hat{\varepsilon} \tag{34}$$

at high eccentricities. Note, however, that in contrast to the iso-surfaces of σ_C and σ_{LP} (see figure 7) an iso-surface of σ_{MLP} is not a developable surface (it has non-zero Gaussian curvature). Hence, in contrast to the log-polar transformation, which is a conformal mapping, it is impossible to map an iso-surface of σ_{MLP} onto a planar domain without introducing local angular distortions. (Such distortions can, however, be reduced by dividing the surface into several parts as done. *e.g.*, in the division between the left and right fields of view in the human vision system.)

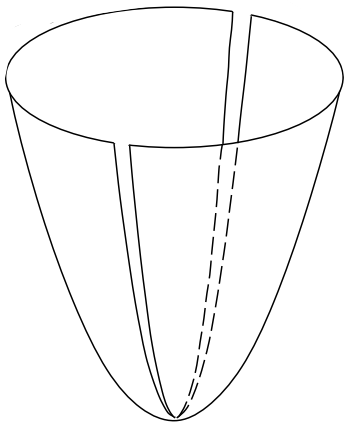

Figure 10: Parameterizing the scale-space representation by a family of hyperboloids constitutes one way of obtaining iso-surfaces of the (modified) scale parameter that correspond to an approximate Cartesian behaviour at the fovea and an approximate log-polar behaviour at the periphery. The smooth transition between these qualitatively different behaviours, however, gives rise to an intrinsically curved iso-surface which cannot be mapped onto a planar surface without introducing angular distortions.

In terms of the diffusion equation, the modified log-polar transformation corresponds to

$$\partial_{t_{MLP}} L_{MLP} = \frac{1}{2}\, \sigma_0^2 (1 + \hat{\varepsilon}^2)\, \nabla^2 L_{MLP}. \tag{35}$$

At low eccentricities, this behaviour can obviously be well modelled by the Cartesian diffusion equation (18) and at high eccentricities it approaches the log-polar diffusion equation (19) which can be written

$$\partial_{t_{LP}} L_{LP} = \frac{1}{2}\, \sigma_0^2 \hat{\varepsilon}^2\, \nabla^2 L_{MLP}. \tag{36}$$

Given receptive fields that register $\nabla^2 L$ at different scales it is straightforward to use (35) for constructing the scale-space by summing up the contributions from different scales. Assuming that $\lim_{\sigma_C \to \infty} L(x;\ \sigma_C^2) = 0$ we have

$$\begin{aligned} L(x;\ t) &= -(L(x;\ \infty) - L(x;\ t)) && (37)\\ &= -\int_{t'=t}^{\infty} \partial_t L(x;\ t')dt' && (38)\\ &= -\frac{1}{2}\int_{t'=t}^{\infty} \nabla^2 L(x;\ t')dt'. && (39) \end{aligned}$$

4 Summary

We have presented a model for how a multi-scale image representation can be constructed for a foveal visual system. This model is an adaptation of conventional linear scale-space theory and provides a direct link to the well-known log-polar mapping. In fact, log-polar sampled image data serve as natural input data to such a foveal scale-space.

An immediate consequence of this model is that there is a linear relationship between minimum receptive field size and eccentricity. Whereas there is a natural choice of exponential scale sampling based on scale invariance, the linear relationship holds independently of any specific scale sampling scheme.

The model is qualitatively consistent with many results from neuroscience and psychophysics concerning visual front-end characteristics. It may have implications both for the construction of artificial vision systems as well as for understanding qualitative characteristics of biological systems. It should be stressed though that the intention has not been to model biological vision as such, but to provide an idealized theoretical model based on first principles. (For an extensive review concerning biological species; see (Hughes 1977).)

5 Discussion

This theory has been developed for (zero-order) Gaussian smoothing functions. However, it equally applies to any filter profile that has an intrinsic scale, in particular, the family of Gaussian derivatives (Koenderink and van Doorn 1990).

Although each filter has a non-infinitesimal extent, it is natural to scale the filters uniformly according to the eccentricity at their *central points*.[6] Of course, one could also imagine stretching the filters continuously which would yield non-symmetric filter profiles (as done in the temporal domain by (Koenderink 1988)). In the spatial domain, however, such an approach would be unnatural, since it violates the translational invariance within each scale layer.

Acknowledgments

We would like to thank Jan Koenderink, Andrea van Doorn and Bart ter Haar Romeny for their valuable comments. This work is a result of a collaboration between Royal Institute of Technology and Utrecht Biophysics Research Institute within the Esprit-BRA project InSight and the Esprit-NSF collaboration Diffusion. The support from the Swedish National Board for Industrial and Technical Development, NUTEK, the Swedish Research Council for Engineering Sciences, TFR, the SPIN project 3D Computer Vision, its participating industrial partners, and the Netherlands Ministry of Economic Affairs is gratefully acknowledged.

The second author gratefully acknowledges a fellowship from the European Research Consortium for Informatics and Mathematics (ERCIM) financed by the Commission of the European Communities.

[6] This means that the shape of the receptive field profiles are not affected by the log-polar transform and corresponds to interpreting the operators as living in the "tangent bundle".

References

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 Intell.*, 8(1):26–33, 1986.

P. Bijl. *Aspects of Visual Contrast Detection.* PhD thesis. , University of Utrecht, University of Utrecht, Dept. of Med. Phys., Princetonplein 5, Utrecht, the Netherlands, May 1991.

C. Braccini, G. Gambarella, G. Sandini, and V. Tagliasco. A model of the early stages of the human visual system: Functional and topological transformations performed in the visual field. *Biological Cybernetics*, 44:47–58, 1982.

J. L. Crowley, A. Chehikian, and F. Veillon. Definition of image primitives. VAP Tech. Rep. IR.A.3.1, LIFIA (IMAG) - I. N. P., Grenoble, France, 1990.

J. L. Crowley. Towards continuously operating integrated vision systems for robotics applications. In P. Johansen and S. Olsen, editors, *Proc. 7th Scandinavian Conf. on Image Analysis*, pages 494–505, Aalborg, Denmark, Aug 13–16 1991.

B. Fischer and H. U. May. Invarianzen in der Katzenretina: Gesetzmäßige Beziehungen zwischen Empfindlichkeit, Größe und Lage receptiver Felder von Ganglienzellen. *Experimental Brain Research*, 11:448–464, 1970.

B. Fischer. Overlap of receptive field centers and representation of the visual field in the cat's optic tract. *Vision Research*, 13:2113–2120, 1973.

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

L. M. J. Florack, B. M. ter Haar Romeny, J. J. Koenderink, and M. A. Viergever. Cartesian differential invariants in scale-space. *J. of Mathematical Imaging and Vision*, 3(4):327–348, 1993.

L. M. J. Florack, B. M. ter Haar Romeny, J. J. Koenderink, and M. A. Viergever. Linear scale-space. *J. of Mathematical Imaging and Vision*, 1994. (To appear).

L. M. J. Florack, B. M. ter Haar Romeny, J. J. Koenderink, and M. A. Viergever. General intensity transformations and differential invariants. *J. of Mathematical Imaging and Vision*, 1994. In press.

L. M. J. Florack. *The Syntactical Structure of Scalar Images.* PhD thesis. , Dept. Med. Phys. Physics, Univ. Utrecht, NL-3508 Utrecht, Netherlands, 1993.

D. H. Hubel and T. N. Wiesel. Uniformity of monkey striate cortex: A parallel relationship between field size, scatter, and magnification factor. *J. of Computational Neurology*, 158:295–305, 1974.

A. Hughes. The topography of vision in mammals of contrasting life style: Comparative optics and retinal organization. In *Handbook of Sensory Physiology*, volume VII/5, pages 613–756. Springer Verlag, New York, 1977.

P. Johansen. On the classification of toppoints in scale space. *J. of Mathematical Imaging and Vision*, 4:57–67, 1994.

J. J. Koenderink and A. J. van Doorn. Visual detection of spatial contrast; influence of location in the visual field, target extent and illuminance level. *Biological Cybernetics*, 30:157–167, 1978.

J. J. Koenderink and A. J. van Doorn. Representation of local geometry in the visual system. *Biological Cybernetics*, 55:367–375, 1987.

J. J. Koenderink and A. J. van Doorn. Receptive field families. *Biological Cybernetics*, 63:291–298, 1990.

J. J. Koenderink and A. J. van Doorn. Generic neighborhood operators. *IEEE Trans. Pattern Analysis and Machine Intell.*, 14(6):597–605, Jun. 1992.

J. J. Koenderink, M. A. Bouman, A. E. Bueno de Mesquita, and S. Slappendel. Perimetry of contrast detection thresholds of moving spatial sine wave patterns. I. The near peripheral visual field (eccentricity 0–8°). *J. of the Optical Society of America*, 68:845–849, 1978.

J. J. Koenderink, M. A. Bouman, A. E. Bueno de Mesquita, and S. Slappendel. Perimetry of contrast detection thresholds of moving spatial sine wave patterns. II. The far peripheral visual field (eccentricity 0–50°). *J. of the Optical Society of America*, 68:850–854, 1978.

J. J. Koenderink, M. A. Bouman, A. E. Bueno de Mesquita, and S. Slappendel. Perimetry of contrast detection thresholds of moving spatial sine wave patterns. III. The target extent as a sensitivity controlling parameter. *J. of the Optical Society of America*, 68:854–860, 1978.

J. J. Koenderink. The concept of local sign. In A. J. van Doorn, W. A. van de Grind, and J. J. Koenderink, editors, *Limits in Perception*, pages 495–547. VNU Science Press, Utrecht, 1984.

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

J. J. Koenderink. Scale-time. *Biological Cybernetics*, 58:159–162, 1988.

T. Lindeberg. Scale-space for discrete signals. *IEEE Trans. Pattern Analysis and Machine Intell.*, 12(3):234–254, Mar. 1990.

T. Lindeberg. Discrete derivative approximations with scale-space properties: A basis for low-level feature extraction. *J. of Mathematical Imaging and Vision*, 3(4):349–376, 1993.

T. Lindeberg. Effective scale: A natural unit for measuring scale-space lifetime. *IEEE Trans. Pattern Analysis and Machine Intell.*, 15(10):1068–1074, Oct. 1993.

T. Lindeberg. Scale-space theory: A basic tool for analysing structures at different scales. *Journal of Applied Statistics*, 21(2):223–261, 1994. Special issue on 'Statistics and Images' (In press).

T. Lindeberg. *Scale-Space Theory in Computer Vision.* The Kluwer International Series in Engineering and Computer Science. Kluwer Academic Publishers, Dordrecht, Netherlands, 1994.

L. Matin and A. Kornheiser. Linked spatial integration, size discrimination and increment threshold with change in background diameter. *Vision Research*, 16:847–860, 1976.

M. Porat and Y. Y. Zeevi. The generalized Gabor scheme of image representation in biological and machine vision. *IEEE Trans. Pattern Analysis and Machine Intell.*, 10(4):452–468, 1988.

J. Rovamo and V. Virsu. An estimation and application of the human cortical magnification factor. *Experimental Brain Research*, 37:495–510, 1979.

G. Sandini and V. Tagliasco. An antropomorphic retina-like structure for scene analysis. *Computer Graphics and Image Processing*, 14(3):365–372, 1980.

E. L. Schwartz. Spatial mapping in the primate sensory projection: Analytic structure and relevance to perception. *Biological Cybernetics*, 25:181–194, 1977.

M. Tistarelli and G. Sandini. Dynamic aspects in active vision. *CVGIP: Image Understanding*, 56(1):108–129, Jul 1992.

W. A. van de Grind, J. J. Koenderink, and A. J. van Doorn. The distribution of human motion detector properties in the monocular visual field. *Vision Research*, 26(5):797–810, 1986.

A. J. van Doorn, J. J. Koenderink, and M. A. Bouman. The influence of the retinal inhomogeneity on the perception of spatial patterns. *Kybernetik*, 10:223–230, 1972.

H. Wässle, U. Grünert, J. Röhrenbeck, and B. Boycott. Retinal ganglion cell density and cortical magnification factor in the primate. *Vision Research*, 30:1897–1911, 1990.

M. E. Wilson. Invariant features of spatial summation with changing locus in the visual field. *J. of Physiology*, 207:611–622, 1970.

A. P. Witkin. Scale-space filtering. In *Proc. 8th Int. Joint Conf. Art. Intell.*, pages 1019–1022, Karlsruhe, West Germany, Aug. 1983.

A. L. Yuille and T. A. Poggio. Scaling theorems for zero-crossings. *IEEE Trans. Pattern Analysis and Machine Intell.*, 8:15–25, 1986.