Biol. Cybern. 55, 367–375 (1987)

Biological
Cybernetics

Representation of Local Geometry in the Visual System

J. J. Koenderink and A. J. van Doorn

Department of Medical and Physiological Physics, Physics Laboratory, State University Utrecht, NL-3508 TA Utrecht, The Netherlands

Abstract. It is shown that a convolution with certain reasonable receptive field (RF) profiles yields the exact partial derivatives of the retinal illuminance blurred to a specified degree. Arbitrary concatenations of such RF profiles yield again similar ones of higher order and for a greater degree of blurring.

By replacing the illuminance with its third order jet extension we obtain position dependent geometries. It is shown how such a representation can function as the substrate for "point processors" computing geometrical features such as edge curvature. We obtain a clear dichotomy between local and multilocal visual routines. The terms of the truncated Taylor series representing the jets are partial derivatives whose corresponding RF profiles closely mimic the well known units in the primary visual cortex. Hence this description provides a novel means to understand and classify these units.

Taking the receptive field outputs as the basic input data one may devise visual routines that compute geometric features on the basis of standard differential geometry exploiting the equivalence with the local jets (partial derivatives with respect to the space coordinates).

1 Introduction

The available luminance distribution (as a function of position of the vantage point, direction in space, and sample apertures) has to be represented in the sensorium such as to enable the organism to tune its behaviour to the optical structure of the physical environment. The sensorium is a physiological substrate that embodies the optical structure of the environment in "machine readable form." This notion catches several important features, e.g. that of *selection* (that which did not become represented does not exist for the organism), of *segregation of quality* (such that different aspects of structure may be directly "addressed"), and of *uniform format* (e.g. spectral composition, temporal or geometrical aspects of the retinal illuminance are each represented as spatiotemporally distributed electrochemical activity). The sensorium is an input device for certain *processors* that address the sensorium roughly as follows: a dataword is issued that determines *where* to read (like page and rough position for an atlas – e.g. "map 137, square E8") and *what* to read (e.g. "find average yearly rainfall," etc.). The latter aspect (the "what") can be understood as the specification of a *logical format* that actually defines the *meaning* of the dataword that is returned. (Remember that the sensorium uses a uniform *physical format*). Thus a reading from the sensorium is not a small "glyph" but is a piece of information for the processor (not a physical structure but a word in the processor's vocabulary) that is defined jointly by the available physical structure in the sensorium and the logical format issued by the processor.

This model neatly banishes the homunculus and enables us to handle the jump between logical levels (i.e. from the physical to the semantic domain). The sensorium is merely a part of the physical world that happens to be in the right physical format such that it can be addressed by the processor. It has little to do with perception as such, except that it presents a bottleneck: what is not represented is nonexistent for the processor. But the sensorium is no inner screen and its activity pattern has no intrinsic meaning at all. A spatial segregation of different aspects of physical structure is mandatory if simple formats are desired, but an eventual convergence to any "final map" is not an issue.

Exactly *what* need be represented in the sensorium depends on the sophistication of the processors (and thus on the complexity of the sensorimotor behaviour of the organism as a whole). In order to leave all options open essentially *all* physical structure will have to be represented and in such a form that quite different processors can set up formats to read their data. In this paper we describe such a representation for two dimensional scalar functions (say the momentary retinal illuminance pattern – we abstract from complications due to spectral composition, temporal structure, etc. for the moment). This representation will be shown to have some theoretical merits, but even more interestingly it will be shown to provide a means to place many of the well known receptive field (RF in the sequel) types described in electrophysiology in a simple unifying scheme and to provide a novel way to understand their relevance in terms of the local geometrical structure of the retinal illuminance.

2 Fuzzy Derivatives

For many visual routines it is important to describe the illuminance on several levels of resolution simultaneously. Many algorithms can be made to work dependably only in this way (Koenderink 1986). Thus one wants to replace the illuminance with a one parameter family of luminances, each member "blurred" to a specific degree. It has been shown that this can be done in a unique way, namely through the diffusion equation. (Other methods generate spurious structure.) One may establish "projections" between members, such that coarse detail may be traced to its "cause" at higher resolution (Koenderink 1984).

Local geometry depends on the differential structure of the illuminance at given points, thus the partial derivatives (or rather certain invariant combinations of them) of the illuminance are important. Hence we want to study the partial derivatives with respect to the spatial parameters throughout the multiresolution representation. It is especially important to express the derivatives of the blurred illuminances in terms of the original illuminance (i.e. the illuminance at highest resolution). The relations are simple: the derivatives of the blurred illuminances are equal to the convolution of the original image with certain RF profiles that may aptly be called "fuzzy derivatives." These profiles are (Fig. 1a and b):

$$\varphi_n(x;t)=\frac{\partial^n}{\partial x^n}\frac{e^{-\frac{x^2}{4t}}}{\sqrt{4\pi t}}$$

$$=\left(\frac{-1}{2\sqrt{t}}\right)^n H_n\left(\frac{x}{2\sqrt{t}}\right)\frac{e^{-\frac{x^2}{4t}}}{\sqrt{4\pi t}}.$$

We have the simple property ($\otimes$ denotes convolution)

$$\frac{\partial^n}{\partial x^n}[f(x)\otimes\varphi_0(x;t)]=\frac{\partial^n f}{\partial x^n}\otimes\varphi_0=f\otimes\varphi_n.$$

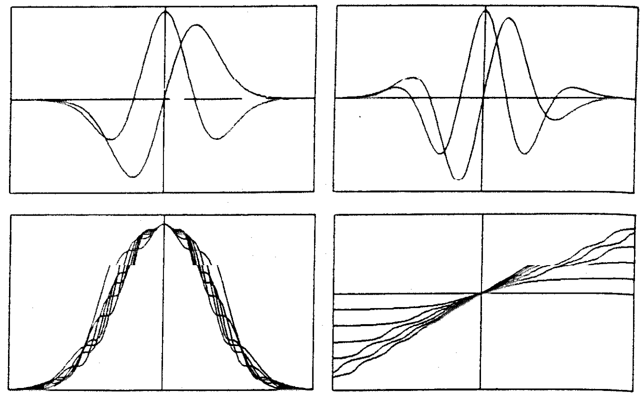

Fig. 1. *Upper left:* The RF profiles of the orders 1 and 2 ($m=0$). In this and the next figure the even components have been scaled with the frequency. *Upper right:* The RF profiles of the orders 3 and 4 ($m=1$). *Lower left:* The amplitude as a function of position for $m=0$ to 5. All orders yield curves that are close to the asymptotic gaussian approximation. *Lower right:* The phases as a function of position for $m=0$ to 5. Note that the phases move between the values $\pm m\pi$. Thus there are exactly m cycles for the orders $2m+1$, $2m+2$

Thus the n-th derivative of the blurred function equals both the blurred n-th derivative of the function and the convolution of the function with φ_n that is the n-th derivative of the blurring kernel. One may concatenate such RF's and obtain higher order derivatives at lower resolution because of a "concatenation theorem:"

$$\varphi_n(x;t)\otimes\varphi_m(x;s)=\varphi_{n+m}(x;t+s).$$

(These theorems are trivial in the Fourier domain, v.i.) Specifically, if one blurs the RF profiles they grow in size without change of shape like in Hartmann's scheme (1982).

Derivatives of the RF profiles satisfy the simple rules:

$$\frac{\partial}{\partial x}\varphi_n(x;t)=\varphi_{n+1}(x;t),$$

$$\frac{\partial}{\partial t}\varphi_n(x;t)=\varphi_{n+2}(x;t)$$

from which one sees immediately that $\frac{\partial^2}{\partial x^2}=\frac{\partial}{\partial t}$; i.e. the RF profiles are solutions of the diffusion equation.

If we set $a_k(t)=\int_{-\infty}^{+\infty} f(x)\varphi_k(x;t)dx$, then we have obviously

$$f\otimes\varphi_0=\sum_{k=0}^{\infty}\frac{a_k}{k!}x^k$$

(which is just the Taylor expansion of the blurred function), but we also have

$$f(x)\cdot\varphi_0(x;t)=\sum_{k=0}^{\infty}\frac{a_k}{k!}2^{k-1}t^{k-\frac{1}{2}}\varphi_k(x;t)$$

(because of the orthonormality of the parabolic cylinder functions).

The latter formula gives a "deblurring" of the blurred function; windowed by $\varphi_0(x; t)$.

The RF profiles look like "Gabor functions" (Gabor 1946), and in fact the asymptotic form for a very high order of differentiation is just a sine or cosine (for odd and even orders, respectively) with spatial (circular) frequency $\omega_k = \sqrt{\frac{1}{t}\left[\frac{n+1}{2}\right]}$ and modulated with a Gaussian envelope of halfwidth $4\sqrt{t}$. Their representation in the spatial frequency domain is an especially simple band-pass type:

$$\Phi_n(\omega; t) = (i\omega)^n e^{-\omega^2 t}$$

from which

$$\langle\omega\rangle = \frac{\Gamma\left(\frac{n+2}{2}\right)}{\sqrt{t}\cdot\Gamma\left(\frac{n+1}{2}\right)}$$

and

$$\langle\omega^2\rangle = \frac{n+1}{2t}.$$

The term "fuzzy derivative" is seen to be very apt in this representation because the n-th spatial derivative corresponds to a multiplication with $(i\omega)^n$. Thus the $\Phi_n(\omega; t)$ are "band limited derivatives." The bandwidth is approximately constant and asymptotically equal to $\Delta\omega = \frac{1}{2\sqrt{t}}$. The transfer functions have no zeros at finite frequencies and are thus especially "nice." We may equally think of the RF's as representing fuzzy spatial derivatives as of tuned spatial frequency filters (Fig. 2).

The "width" of the RF's (defined as the square root of the second moment of the square of the RF profile; the first moment vanishes) depends on the order of differentiation and amounts to:

$$\Delta x = (\langle x^2\rangle - \langle x\rangle^2)^{1/2} \sim 2\sqrt{t}.$$

Thus the width is approximately independent of the order. The "uncertainty relation" for the order n gives the product of the spatial extent and the spectral width as approximately unity, that is near the theoretical limit of one half.

It is often convenient to introduce another notation and set

$$s_m(x; t) = \omega_m \varphi_{2m+1}(x; t),$$

$$c_m(x; t) = \varphi_{2m+2}(x; t)$$

with

$$\omega_m = \sqrt{\frac{m+1}{t}}.$$

The functions s_m, c_m behave like local versions of the sine and cosine with frequency ω_m (except for a scale factor). Thus we introduce an amplitude A and a phase angle η as follows

$$A = \sqrt{s_m^2 + c_m^2},$$

$$\eta = \mathrm{arctg}(s_m/c_m).$$

For a local disturbance $\delta(x)$ the amplitude behaves like $e^{-x^2/8t}$ (times the scale factor), whereas the phase varies between $-m\pi$ and $+m\pi$, with a slope proportional to ω_m at the origin. Thus amplitude ratios are independent of position whereas the phase angles are proportional with position (Fig. 1b and c).

Two dimensional RF profiles can be obtained through multiplication and addition in the same way as the partial derivatives, e.g. the second partial derivative with respect to x is (note that we multiply by the zeroth derivative with respect to y!)

$$\varphi_{20}(x, y; t) = \varphi_2(x; t)\varphi_0(y; t) = \frac{1}{2t}\left(\frac{x^2}{2t} - 1\right)\frac{e^{-\frac{x^2+y^2}{4t}}}{4\pi t}.$$

By way of an example: the Laplacean $\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$ is $\varphi_{20}(x, y; t) + \varphi_{02}(x, y; t)$

$$= \frac{1}{\varrho}\frac{\partial}{\partial\varrho}\varrho\frac{\partial}{\partial\varrho}\frac{e^{-\varrho^2/4t}}{4\pi t} = \Delta\varrho\frac{e^{-\frac{\varrho^2}{4t}}}{4\pi t}.$$

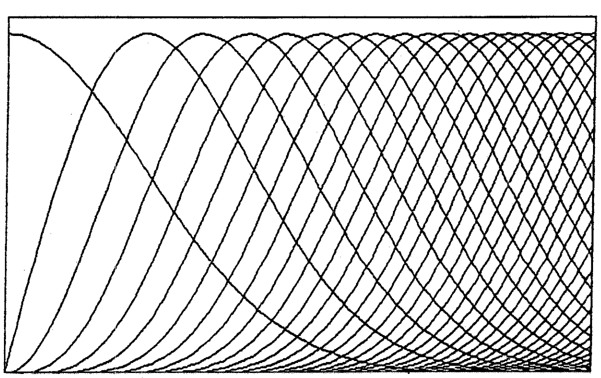

Fig. 2. Modulus of the normalized spectral transfer functions of the RF's plotted on linear scales

3 Local Jets

The retinal image can be modelled as a function $L: \mathbb{R}^2 \rightarrow R$, where $L(x, y; t)$ denotes illuminance. The *local jet* (Poston and Stewart 1978) of order N of the illuminance $L(x, y; t)$ at the point $p=(X, Y)$ is denoted j_p^k and is the equivalence class of smooth functions with contact of the order k at p. The jet may be represented by the truncated Taylor series

$$L(x+\xi, y+\eta; t)=\sum_{n=0}^{\infty}\frac{1}{n!}\left(\xi\frac{\partial}{\partial x}+\eta\frac{\partial}{\partial y}\right)^n L$$
$$=\sum_{n=0}^{k}\sum_{i=0}^{n}\binom{n}{i}a_{1\ldots 2}\frac{\xi^{n-i}\eta^i}{n!}.$$

Note that the coefficients $a_{1\ldots 2}$ can all be found simply as the activities of RF's centered at (X, Y):

$$a_{1\ldots 2}=\frac{\partial^{n+m}}{\partial x^n \partial y^m}L(x, y; t)$$
$$=L(x, y; 0)\otimes\varphi_{nm}(x, y; t)$$

(the coefficient $a_{1\ldots 2}$ has n subscripts "1" and m subscripts "2").

Instead of $L(x, y; t)$ we consider its k-jet extension $j^k L:\mathbb{R}^2\to J_k(\mathbb{R}^2\to\mathbb{R})$ [with $J_k(\mathbb{R}^2\to\mathbb{R})$ the space of k-jets for functions $\mathbb{R}^2\to\mathbb{R}$]. The jets characterize the local geometry in the neighbourhood of the point. The order of the jet determines the amount of geometry represented, e.g. a second order jet allows us to detect "line orientation" but not "line curvature" (a third order jet would also enable us to find the latter), etc. Note that this representation is redundant in the sense that a zeroth order jet at any position obviously contains the complete illuminance! But then geometrical features become *multilocal* objects, i.e. in order to compute boundary curvature the processor would have to look at different positions simultaneously, whereas in the case of third order jets it could establish a format (v.i.) that provided the information by addressing a *single location.* Routines accessing a single location may aptly be called *point processors*, those accessing multiple locations *array processors*. The difference is crucial in the sense that point processors need no geometrical expertise at all, whereas array processors do (e.g. they have to know the environment or neighbours of a given location).

The order of the jets in the representation determines the "features" (the geometrical properties) that can be computed by a point processor. Not all available features need play a role in the effectively used visual routines, of course.

One may avoid redundancy by reducing the sample density according to the number of degrees of freedom in the jet. Thus a k-th order jet permits reduction by a factor $F=(k+1)(k+2)/2$, and the average distance between samples may be taken as $\sqrt{F}$ times the acuity.

In practice it will be advantageous to use not the raw Taylor series as such, but to group coefficients according to their properties under transformations (e.g. rotations, homotheties). For example, the coefficients a_1, a_2 are better understood as a single entity (the gradient) and this can be done in the sensorium by setting up a continuum (set of cells in practice) of "gradient detectors" ordered with respect to orientation (μ), with RF profiles

$$\varphi_{10}(x\cos\mu-y\sin\mu, x\sin\mu+y\cos\mu; t).$$

The second order terms can most conveniently be written

$$H=\tfrac{1}{2}(a_{11}+a_{22})\quad\text{(one half the trace of the Hessian)},$$
$$K=a_{11}a_{22}-a_{12}^2\quad\text{(the determinant of the Hessian)},$$
$$a_{11}x^2+2a_{12}xy+a_{22}y^2$$
$$=\varrho^2\left[H+\sqrt{H^2-K}\cos\left(2\varphi-\operatorname{arctg}\frac{2a_{12}}{a_{11}-a_{22}}\right)\right]$$

such as to reveal the rotation invariant (the Laplacean) and the transformation properties of the remnant (a shear). The Laplacean can be represented with a single RF (v.s.) whereas the shear can again be mapped on a continuum (orientation). Alternatively, one may represent the second order terms conveniently and completely through a continuum of "line detectors" with profiles

$$\varphi_{20}(x\cos\mu-y\sin\mu, x\sin\mu+y\cos\mu; t).$$

The extrema of activity yield the orientation of the "line," the sum of the extremal activities H, their product K, and the ratio the degree of "blobness" or "lineness" (or "elongatedness").

Similar considerations apply to the third and higher order terms of the jet.

4 Example: Curvature

We are now ready to consider specific visual processors using differential geometry and implementing the resulting expressions from analysis through the simple artifice of substituting RF's for the partial derivatives.

Consider the structure of a boundary (or edge) curvature routine. This processor returns the value of the curvature of an isophote encircling a "blob," i.e. the curvature of the curve defined as $L(x, y)=L^*$. Twice implicit differentiation of this expression yields the equations:

$$a_1+a_2y'=0,$$

$$a_{11}+2a_{12}y'+a_{22}(y')^2+a_2y''=0.$$

Let us now pick the $x-y$ axes in such a way that $a_1=0$ (and consequently $y'=0$), this is done by consulting the first order terms and making the gradient direction the

y-axis. Then we immediately find that $y'' = -a_{11}/a_2$. This second derivative of y with respect to x is the curvature of the isophote. Thus we may easily define a format that looks at the second order jet and returns an ordered pair of numbers (a_{11}, a_2) as a description of the local boundary (both sharpness and curvature). Note that this could be done at any place where the gradient doesn't vanish, not merely at "the" boundary. The "edge curvature" makes most sense on curves defined through $a_{22}=0$ (steepest gradient) but is well defined over a strip along this curve.

It is instructive to consider yet another point processor for curvature which may be called "line curvature" (in contradistinction to edge curvature). We define a local line piece as the occurrence of a local jet of the form (in a suitable coordinate system defined by the extremal output of the line detectors)

$$L(x+\xi, y+\eta)$$
$$= a_0 + a_1\xi + \tfrac{1}{2}(a_{11}\xi^2 + a_{22}\eta^2) + \ldots$$

with $|a_{11}| \ll |a_{22}|$ and $a_2=0$. [This is a "cylinder" in (x, y, L) space with axis along the x-direction.] The curvature of this line is the rate of change of its orientation along the line direction. This is determined by the third order terms, one easily shows that the curvature is $-a_{112}/(a_{11}-a_{22}) \approx a_{112}/a_{22}$. Thus we may easily define a format for a "line curvature routine."

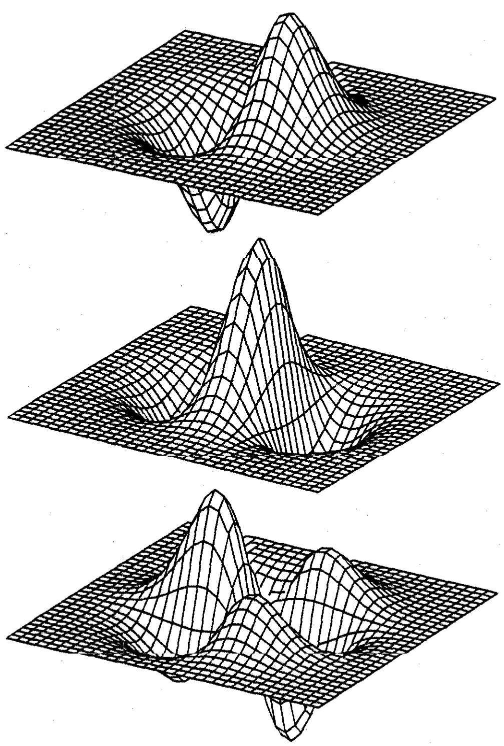

Fig. 3. The φ_i, φ_{ii}, and φ_{iij} RF's. These RF's can be combined to compute boundary curvature or line curvature. The RF φ_{ii} plays the role of curvature sensitive element for boundary curvature detection (the RF φ_i detects the edge) but of line detector for line curvature (then the RF φ_{iij} is the curvature sensitive element). Thus the RF's should not be considered as "feature detectors" with fixed functions

This example is instructive because it shows clearly that one should not think of the terms of the Taylor series as "feature detectors" (Sekuler and Blake 1985) at all, even if e.g. φ_{20} behaves in many respects as a "line detector." The "features" are defined by the context and the format, e.g. in the routine for line curvature φ_{20} features as a line detector but in the routine for boundary curvature it appears as a curvature sensitive device! The local geometry is represented in the jet as a whole and features are best defined as functions on the jet (Fig. 3).

In a similar way as for the curvature we may find expressions for such significant entities as "ridges" (in a coordinate system where $a_1=0$ a ridge point satisfies $a_{111}a_2 - a_{12}a_{11}=0$) or "steepest edges" (in a similar coordinate system the condition is $a_{22}=0$; $a_2 \neq 0$), etc.

5 Geometrical Interpretation of the Jets up to the Fourth Order

The jets are useful to categorize local structure. The categories will depend on the relative magnitude of the terms (coarse quantizations of the jets). Examples are "flat" versus "illuminance gradient," "blobness" versus "elongatedness," etc. Very useful categories can also be obtained from the *sign* of certain differential invariants (e.g. K is positive for blob-like, negative for transition regions; in the blob-like regions negative values of H signal light, positive values dark blobs, etc.). In order to understand the categories we have to study the local geometry in terms of the Taylor series.

5.1 Zeroth and First Order

Obviously the zeroth order jet is nothing but the local illuminance. Nothing at all is made explicit, although in principle no structure has been discarded. Point processors can compute no geometrical entities at all.

The first order jet yields as a single geometrical entity the *gradient*, i.e. a direction and a magnitude available to point processors.

Although both zeroth and first order carry information about the physical environment (exterospecific information), this information is highly dependent on many *interospecific* factors. E.g. the zeroth order depends on the transmittance of the eye media (which may be no real problem because it is more or less constant although unknown) and on the sensitivity of the transducers (which may be very problematic because it is governed by variable peripheral structures that may not be accessible to the central processor). The first order depends on unknown but largely constant physical factors and would thus be a useful source of information. However, it appears that the human visual system *discards the first order* at a very peripheral stage. (Although the issue has certainly not been decided definitely.)

5.2 Second Order

As has been discussed above the second order is most easily studied in a coordinate system (which may be found from the activity of "line detectors") such that (at a point of vanishing gradient)

$$L(x+\xi, y+\eta)=a_0+\tfrac{1}{2}(a_{11}\xi^2+a_{22}\eta^2)+\dots.$$

The relevant geometry (apart from orientation) then depends on a_{11}, a_{22}, whereas the orientation of the system is itself of considerable geometrical interest. It is most convenient to use instead the "deviation from flatness" $D=\sqrt{a_{11}^2+a_{22}^2}$ and the angle defined through $\sin\mu=(a_{11}+a_{22})/\sqrt{2D}$, $\cos\mu=(a_{11}-a_{22})/\sqrt{2D}$. Small values of D indicate the absence of structure, larger values increasing importance of deviations from the 1st order jet. The interesting range of values of μ is limited to a single quadrant because the other quadrants can be obtained by interchange of the $x-y$ axes and/or contrast reversal. If μ is increased from zero one starts with an "anti-umbilical" (like $L=x^2-y^2$) and then obtains a series of elongated shapes until $\mu=\frac{\pi}{2}$ for which one has a circular blob (an "umbilical" like $L=x^2+y^2$). The elongatedness is most pronounced for $\mu=\frac{\pi}{4}$ for which one has a "line" [like $L=x^2$: a cylinder in (x, y, L)-space]. Thus $\sin^2 2\mu$ is a useful "measure of elongatedness," whereas $\cos^2 2\mu$ yields a measure of "blobness" for $\cos 2\mu<0$ and of "featurelessness" if $\cos 2\mu>0$. (Note that the sum of "elongatedness," "blobness and "featurelessness" is neatly 100% in any case.)

5.3 Third and Fourth Order

In most visual routines the third order terms will only be consulted if the second order structure is important. (Otherwise the local geometry is "flat" anyway and it makes more sense to study a wider environment.) The third order then appears as a *perturbation* on the second order. This is a reason for representing the third order in terms of the canonical coordinates for the second order, rather than going for third order invariants. The most interesting case is that of an elongated second order structure $|a_{11}|\gg|a_{22}|$:

$$L(x+\xi, y+\eta)=a_0+\tfrac{1}{2}(a_{11}\xi^2+a_{22}\eta^2)$$
$$+\tfrac{1}{6}(a_{111}\xi^3+3a_{112}\xi^2\eta+3a_{122}\xi\eta^2+a_{222}\eta^3)+\dots.$$

Then the terms acquire simple geometrical meanings. These may be found by studying the changes of the second order terms due to the third order. (A similar analysis yields the geometrical interpretation of the fourth order.) We illustrate the method for the third order case:

At a point $x^*=x+\delta x$, $y^*=y+\delta y$ you have:

$$a_{11}^*=a_{11}+a_{111}\delta x+a_{112}\delta y+\dots,$$
$$a_{12}^*=a_{112}\delta x+a_{122}\delta y+\dots,$$
$$a_{22}^*=a_{22}+a_{122}\delta x+a_{222}\delta y+\dots.$$

Clearly the second order form is no longer in canonical coordinates. This can be remedied through a rotation of the local coordinate axes over a suitable angle v. For $v=-\dfrac{a_{112}\delta x+a_{122}\delta y}{a_{11}-a_{22}}$ the coefficient a_{12}^* vanishes to the first order. The other coefficients do not change in this approximation. Thus the orientation has been rotated over v due to the third order terms, whereas the width and length of the elongated blob have also changed (they are proportional with $a_{11}^{*-\frac{1}{2}}$ and $a_{22}^{*-\frac{1}{2}}$, respectively). The geometrical effects can be described

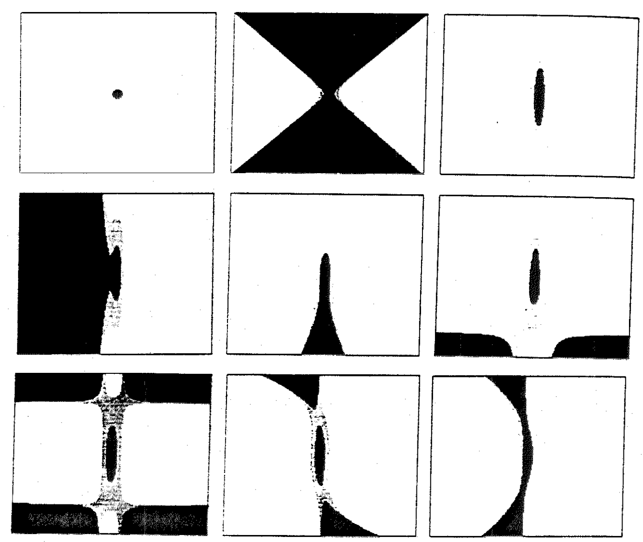

Fig. 4. *Top row – left:* only quadratic terms. A 100% blob; *middle:* only quadratic terms. A 100% featureless image (no blobness, no elongatedness, the image is congruent to its own negative – thus it is neither a dark nor a light thing); *right:* only quadratic terms. A "line," with 92% elongatedness, 8% blobness to it. The two lower rows do all have the "92% line" for the quadratic part and various amounts of third and fourth order terms.
Middle row – left: addition of (third order) edginess term; *middle:* addition of (third order) endedness term; *right:* addition of (third order) splay term.
Bottom row – left: addition of (fourth order) hourglass tendency. This can be understood as a "splay tendency;" *middle:* addition of (fourth order) curvature tendency. Because the third order curvature term is absent here one gets an inflection; *right:* addition of a (third order) curvature term

as a "curvature," a "splay," and "edginess" or "endedness" of the elongated blob (Fig. 4).

From the fourth order we can similarly derive conditions for a "curvature trend," "ridginess" versus "barness," and "spindle" versus "hourglass" tendencies.

Other routines may come into play if certain terms vanish. E.g. if the "curvature trend" vanishes the fourth order may yield evidence for the type of "vertex," etc. Some important features are completely determined by the vanishing of certain terms, e.g. the "zero crossings" ($H=0$).

5.4 Higher Order Jets

The RF profiles are not orthogonal functions; although even and odd types are obviously mutually orthogonal, the overlap integrals for RF's of similar types do not vanish and correlation between the orders n and $n+2$ amounts to $-\sqrt{(2n+1)/(2n+3)}$. Thus the higher derivatives are highly correlated, irrespective the amount of blurring. In view of this fact it makes sense to limit the order of the jets to about three or four. [The correlation is easily understood from the frequency representation: the finite spectral resolving power requires the center frequencies to have a minimum spacing in order to avoid overlap. This immediately yields the results that the orders should differ by at least $\Delta n=\sqrt{2(n+1)}$.]

It is, of course, possible to form linear combinations of the RF profiles that are *mutually uncorrelated*. The resulting profiles are the "parabolic cylinder functions" (Abramowitz and Stegun 1964). Whereas they inherit many of the nice properties of the RF profiles considered here (and to boot form a complete orthonormal set), they also have a few drawbacks. Most of the "weight" in these RF's is at the outer limit of the fields (which themselves grow in size with the square root of the order) and are unlike physiological RF's. The Fourier transforms of these RF's have the *same* functional form as their spatial forms and thus contain many zeros. Although the parabolic cylinder functions are attractive from a mathematical viewpoint they appear to be unlikely candidates for physiologically relevant models. (Even from the former point of view this approach has the drawback that it is hardly local because the RF sizes grow without bound when the order increases.)

As mentioned before the zeroth order may be dropped without appreciable loss. The first order is useful, but in view of the fact that given an analytic curve and an analytic second principal form (second order part of the Taylor series) there exists a surface through the curve with that principal form (Cartan 1943), one appears to loose little by dropping it, too. Thus the limitation to the second and third order terms in organic vision makes sense from diverse points of view. These can be handled with only three types of RF ($\varphi_{ii}, \varphi_{iii}, \varphi_{iij}$).

6 Multilocal Routines

Some visual routines are inherently multilocal, e.g. *finding things* (finding corresponding loci in left and right retinal patterns, finding repetition of pattern, line continuation, parallel line elements, etc.). Such routines often take the form of a correlation or matching. This may be done for patterns that mutually differ by combinations of a translation, a rotation, a dilation, and a shear. It is possible to implement correlation through comparison of the jets rather than the local illuminance pattern itself. This has the advantage that it is easy to make the comparison on the basis "features" at no extra cost, moreover the comparison can easily be stratified. (I.e. one first compares broad features and exits on "reject" but initiates the comparison of finer features on an "accept.") If the comparison is on the basis of *amplitudes* only, then the exact position of the stimulus with respect to the RF centers is immaterial.

In the processing of spatiotemporal illuminance these methods may be used to implement routines for the extraction of image flow. Such flow algorithms would be multilocal. On the other hand, pure "point processors" are also feasible: the change of phase angles yields the velocity, or – equivalently – one may compute $\dot{a}_k/a_{k+1}$ which is numerically equal to the velocity, independent of the order.

Other examples of multilocal entities are the loci of zero crossings of certain differential invariants (e.g. of the Laplacean). It does not suffice that a value is very low (i.e. below the threshold of some processor) because that would be true for a featureless region, too. It is necessary that the value be low and in addition assumes values of both polarities in a neighbourhood of the point. But then a "zero crossing detector" has to address a point with its nearest neighbours. This assumes a certain *topological* expertise.

For yet different processors one requires even more sophisticated expertise. E.g. a "vertex detector" has to compare the curvature at three successive points along a curve. Thus the processor must be able to address a jet at a certain distance in the direction as specified by another jet. Here a rudimentary expertise in *projective* geometry is required (a "connection"). A *connection* is merely the specification of a constant field of orientations, e.g. a set of pointers between neighbouring jet representations linking "the same" orientations with each other. In the visual system the connection could well be gauged through translations induced by eyemovements. (This mechanism could even provide a *metrical* connection.)

On the other hand, a few significant routines use hardly any geometrical expertise (e.g. those that merely use statistical measures like texture routines) or even none at all. An example of the latter are routines for "numerosity," the very concept of number forcing such routines to be invariant under arbitrary permutations of the arguments (or "formal parameters").

7 Discussion

The theory can be used to put several apparently unrelated facts from electrophysiology and visual psychophysics into a single uniform framework. It may also serve to put some widely held preconceptions into a new perspective.

The sensorium is essentially the input medium for various central processors that in their turn form the knowledge base for perceptual programs. It is a corner of the physical environment that is in the right physical format to be "machine readable" (like a papertape for a tape reader). In itself the physical structure of the sensorium is perceptually meaningless (in the same sense as the holes in a papertape are irrelevant to the program), although the actual result of a read operation has semantic content due to the logical format issued by the calling routine (as a certain sequence of holes may have very different meaning depending on the format issued by the program using the papertape-reader). The sensorium does not "compute perceptions" at all but merely embodies the structure of the environment in a convenient form.

Within a local jet the directionally or orientationally polarized RF's (like a_{ii} or a_{iij}, etc.) are represented by a "continuum" of similar RF profiles that merely differ in orientation. This is, of course, just the empirical definition of the cortical (orientation)-"hypercolumn" (Sekuler and Blake 1985). For many visual routines it is necessary that different orders of RF's of the *same orientation* can be recognized as such, thus a grouping in which such RF's are contained in a sub-record seems indicated. However, it is also necessary to be able to address different RF's of orientations differing by 90 degrees. Whether this is an actual physiological possibility is uncertain.

The modules (like "cortical columns" in the physiological domain or "records" of raw data in the syntactic domain) of the sensorium are local approximations (Nth order jets) of the retinal illuminance that can be addressed as *a single datum* by the point processors. The records are not iconic or glyph-like but more like N-tuples of numbers satisfying certain invariances under coordinate transformations. They have semantic content in terms of certain visual routines. There is an essential difference between point and array processors. The latter need a geometrical "connection" provided by the calling routine (a "local sign"), the former do not, they are pure "point" processors. Thus certain local structures (e.g. in textured patterns) may be equivalent stimuli for point processors, but discriminable for array processors, etc. Many intermediate forms are conceivable.

The records are not "features" in the classical sense, but general (not interpreted) representations. The sensorium does not "commit" itself as it would if it were to use a feature based representation. Neither are the fields of the records (e.g. the output of a certain "line detector") features: their meaning is as yet undecided (merely syntactically determined) and is only fixed when the record is read and used by a certain processor. The meaning is then established relative to the use of the processor.

The jets up to the fifth order contain many RF's that are immediately reminescent of well known physiological structures. [E.g. H: a Kuffler unit, a_{ii}: a line detector, a_{iij}: an "end stopped" cell, certain combinations mentioned in the text: curvature detectors or corner detectors (Hubel and Wiesel 1977). Routines using only amplitude information would be similar to the "complex cells" in many respects, e.g. they would lack clear on/off regions and be insensitive to precise stimulus position.] In fact the historical sequence in which the diverse types of RF in the visual system have been discovered follows roughly the order of the corresponding fuzzy partial derivative. Thus the Taylor series appears to yield a simple scheme to classify RF's, and perhaps it could be used to guide future search (by actually "predicting" new RF's). This latter possibility seems viable once the processors for important geometrical objects have been identified. Moreover, the present model is well suited to put a partial order (computational complexity) on elementary visual processors.

The spectral description of the jets shows them to consist of a set of staggered spatial bandpass filters with tuning widths that seem to admit them as "spatial frequency channels" of the type proposed by Campbell and Robson (1968). In fact, a "spatial frequency description" would be equally apt as the spatial one. The interesting point is that the present representation sustains two useful views at the same time: "viewed from the outside" the RF's represent partial derivatives of the blurred retinal image, "viewed from the inside" the RF's provide a Fourier decomposition of a windowed part of the retinal image at full resolution. Both viewpoints have their obvious uses although the spectral representation has been overly popular lately. That looking at a retinal illuminance distribution through a receptive field profile (or even through several layers of them!) is *equivalent* to looking at a certain partial derivative of a blurred pattern is a new insight that immediately leads to useful interpretations in terms of differential geometry.

Acknowledgement. The work described in this paper was made possible by a NATO grant. I thank Whitman Richards for stimulating discussions.

References

Abramowitz M, Stegun IA (1964) Handbook of mathematical functions. Dover, New York

Arnold VI, Gusein-Zade SM, Varchenko AN (1985) Singularities of differential maps. Birkhäuser, Boston

Campbell DT, Robson JG (1968) Application of Fourier analysis to the visibility of gratings. J Physiol 197:551–566

Cartan E (1943) Les surfaces qui admettent une seconde forme fondamentale donnée. Bull Sci Math 67:8–32

Gabor D (1946) Theory of communication. J IEE 93:429–459

Hartmann G (1982) Recursive features of circular receptive fields. Biol Cybern 43:199–208

Hubel D, Wiesel T (1977) Functional architecture of Macaque monkey visual cortex. Proc R Soc (London) B 198:1–59

Koenderink JJ, Doorn AJ van (1984) The structure of images. Biol Cybern 50:363–370

Koenderink JJ, Doorn AJ van (1986) Dynamic shape. Biol Cybern 53:383–396

Poston T, Stewart I (1978) Catastrophy theory and its applications. Pitman, London

Sekuler R, Blake R (1985) Perception. A.A. Knopf, New York

Received: July 2, 1986

Dr. Jan Koenderink
Rijksuniversiteit Utrecht
Fysisch Laboratorium
Princetonplein 5
Postbus 80.000
NL-3508 TA Utrecht
The Netherlands