ESANN'1993 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 7-8-9 April 1993, D-Facto public., ISBN 2-9600049-0-6, pp. 129-136

Motion Processing in the Retina: About a Velocity Matched Filter

Jeanny Hérault & William Beaudot

Laboratoire de Traitement d'Images et Reconnaissance de Formes,
Institut National Polytechnique de Grenoble, 46 Avenue Félix Viallet,
38031 Grenoble Cedex, France

Abstract. This paper presents an original approach to visual motion analysis. We propose an analog model of the vertebrate retina which performs a regularisation of the retinal image and we show, from a signal processing viewpoint, that a high-pass temporal filter and a spatial directional filtering, after this retinal processing, can act naturally as a matched filter, insensitive to spatiotemporal noise, to detect motion in an image sequences composed of translational moving objects. The motion detection and the directional selectivity are achieved for local and global motion .

1. Introduction

Motion analysis in an image sequence is not still fully resolved: the approaches inspired by biological studies such as the Reichardt model or the motion-energy model are not unanimously approved [1], and the more "theoretical" ones, such as those based on the regularisation theory [2] or on the Markov random field theory [3], often require some heavy computational efforts without ensuring perfect results and sufficient noise immunity. However these last two approaches come down to a problem of minimisation of an energy function [4], [5], [6] and might be solved with the use of analog resistive networks [2], [6] in order to reduce the computational effort. Even though analog neural networks are used to analyse the visual motion, their relations with some biological data are often very far in spite of a great interest in the biological mechanisms involved in this process [7]. In the present paper, we propose a mixed approach for the motion detection which consists initially of the inspiration of neurobiological data, especially about the neural structure of the vertebrate retina, and then followed by the study of its motion processing in a signal processing viewpoint. In the second section, an analog model of the retinal processing is presented, which might be an appropriate reply to the need of a spatiotemporal regularisation. The equations and transfer functions are then derived from the analog circuit and used in section 3 to find the type of the matched filters that can naturally appear after the first level of retinal processing: the high-pass time filtering and the directional selectivity might be the early visual process involved in the visual motion detection. In section 4, we present some experimental results on artificial image sequence composed of basic objects with and without addition of noise. Finally we will conclude on the neurophysiological plausibility of our retinal processing model and compare its advantages in contrast to the motion representation given by the classical gradient equation [5].

2. A Spatiotemporal Model of the Retinal Processing

2.1 An Enhanced Analog Model

The retina is our main biological reference in early vision for two main reasons: first it was the more studied neural structure in vision [9], and secondly, some analog realisations have been attempted [10], [8]. We present now a retinal model, derived from the model of Mead [8], but including more complete and realistic features (Fig. 1).

ESANN'1993 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 7-8-9 April 1993, D-Facto public., ISBN 2-9600049-0-6, pp. 129-136

This enhanced model consists of two resistive grids, c(k,t) for the photoreceptors layer and h(k,t) for the H-cell layer, where each node (neuron) is attached to a leaky integrator (membrane characteristics). The output layer s(k,t), that is the bipolar cells layer, is computed as the difference between the photoreceptors and H-cells layers, and provides the global output of the Outer Plexiform Layer (OPL).

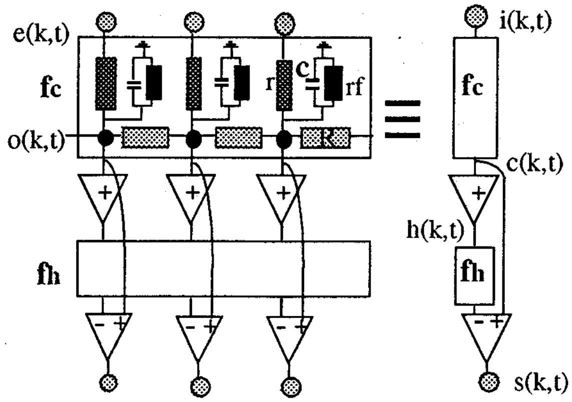


Fig. 1. 1D analog model of the Outer Plexiform Layer (OPL) of the vertebrate retina: f_c and f_h are the same kind of circuits but with different parameters. f_c stands for the coupling between cones c(k,t), f_h stands for the coupling between horizontal cells h(k,t), s(k,t) is the output of the triadic synapse (c(k,t) - h(k,t)) and provides an input to the bipolar cells.

2.2 The Spatiotemporal Transfer Function

The OPL transfer function can be derived from the Kirchhoff's current law at each node of the circuit. We obtain for each resistive square mesh with input e and output o:

$$o(\mathbf{k},t)/r_f + C.\partial o(\mathbf{k},t)/\partial t + [\textstyle\sum_{i \in \mathcal{V}_4} o(\mathbf{k},t) - o(\mathbf{k}\text{-}\mathbf{i},t)\ ]/R = [e(\mathbf{k},t) - o(\mathbf{k},t)]/r \qquad (2.1)$$

where **k** denotes the spatial position (k_x,k_y) of a node, the spatial sampling periods Δx and Δy being equal to 1, $\mathcal{V}_4$ defines the spatial 4-neighbourhood in the grid, e(**k**,t) is the input signal and o(**k**,t) the output signal of the grid, r_f, R, r and C are shown in Fig. 1. By applying the Z-transform and the Fourier transform with respect to **k** and t respectively, a frequency representation of (2.1) is derived:

$$(\beta + j2\pi f_t\tau + 4\alpha - \alpha.[z_x + z_x^{-1} + z_y + z_y^{-1}] + 1).O(f_x,f_y,f_t) = E(f_x,f_y,f_t)$$

where α, β and τ are given by r/R, r/r_f, and r.C respectively and $z_i = \exp(j2\pi f_i)$. f_x and f_y are the spatial frequencies normalised to $1/\Delta x = 1$, f_t is the temporal one. We obtain the (discrete-in-space continuous in time) transfer function of a resistive grid:

$$\mathcal{R}(f_x,f_y,f_t) = O(f_x,f_y,f_t)\ /\ E(f_x,f_y,f_t) = (1+\beta+4\alpha-2\alpha[\cos(2\pi f_x)+\cos(2\pi f_y)]+j2\pi f_t\tau\ )^{-1}$$

The model includes two resistive grids, so two transfer functions: $\mathcal{R}_c(f_x,f_y,f_t)$ for the photoreceptors layer and $\mathcal{R}_h(f_x,f_y,f_t)$ for the H-cells layer. Obviously each layer can be characterised by different values for α, β and τ. The output signal is the difference between the outputs of these two layers, then the total transfer function is written as:

$$G(f_x,f_y,f_t) = \mathcal{R}_c(f_x,f_y,f_t)\ .\ [1 - \mathcal{R}_h(f_x,f_y,f_t)\ ] = \qquad (2.2)$$

$$\frac{\beta_h+2\alpha_h[2-\cos(2\pi f_x)-\cos(2\pi f_y)]+j2\pi f_t\tau_h}{[1+\beta_c+2\alpha_c[2-\cos(2\pi f_x)-\cos(2\pi f_y)]+j2\pi f_t\tau_c].[1+\beta_h+2\alpha_h[2-\cos(2\pi f_x)-\cos(2\pi f_y)]+j2\pi f_t\tau_h]}$$

ESANN'1993 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 7-8-9 April 1993, D-Facto public., ISBN 2-9600049-0-6, pp. 129-136

The purely spatial transfer function (for long lasting steady images) is given at fig. 3a.

2.3 The Signal Processing in the Retinal Model

Let us denote g(x,y,t) the impulse response of this retinal filter (which has (2.2) as transfer function). It is written, where $*$ is the 3D-spatiotemporal convolution:

$$g(x,y,t) = R_c(x,y,t) * [\ \delta(x,y,t) - R_h(x,y,t)\ ]$$

Rc and Rh are the impulse response of each resistive grid. If i(x,y,t) designates the brightness in the input image, the output of the retinal OPL model is expressed by:

$$s(x,y,t) = g(x,y,t) * i(x,y,t)$$

Both the filters, Rc and Rh, are spatiotemporal low-pass filters, then our retinal model computes a difference of spatiotemporal low-pass filters (that is a spatiotemporal band-pass filter). Therefore it is a generalisation of the well-known lateral inhibition to the time domain. This spatiotemporal inhibition leads to some interesting properties as the predictive coding, the redundancy reduction [11] which express the enhancement of the spatiotemporal events (that is edges and motion). Figure 2 shows the result of this processing on an image sequence.

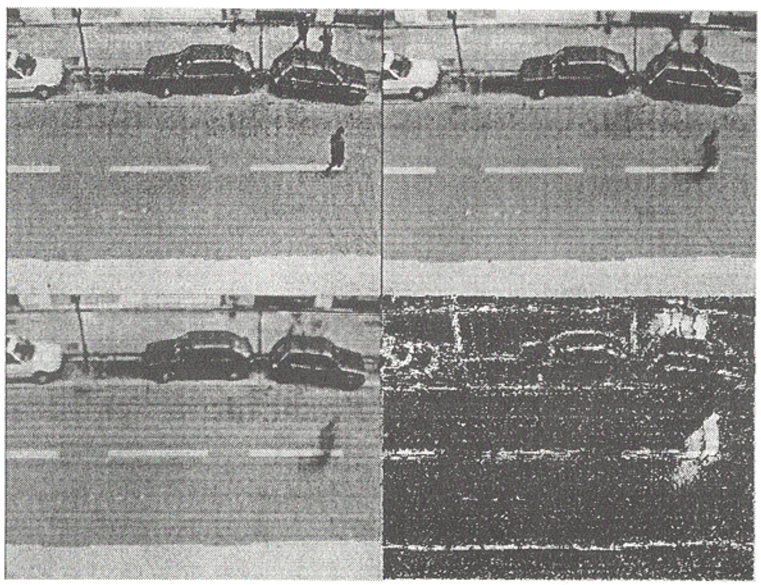


Fig. 2. Result of the simulation of the retinal processing on a real image sequence: at left-top an input image of the sequence (256x256x8bits) where some people walk in a street, at right-top the image of the photoreceptor layer, at left-bottom the image of the horizontal layer and at right-bottom the image of the bipolar layer. This last image is rough: no thresholding has been made. One can see at this early level, that moving parts are highlighted. Further processing is needed to remove the remaining static objects.

We wish to emphasise here a fundamental property of the retinal filtering which must be noted: its temporal and its spatial filtering components are not separable. Therefore, this system should behave in a specific way with respect to the images where the time and space variables are bound, that is moving patterns.

ESANN'1993 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 7-8-9 April 1993, D-Facto public., ISBN 2-9600049-0-6, pp. 129-136

3. Motion Analysis and the Retinal Outer Plexiform Layer

3.1 The OPL Response to a Translational Moving Pattern

Let us consider a 2D pattern i(x,y) with a Fourier transform given by $I(f_x,f_y)$. If this pattern has a translational velocity $\mathbf{V} = (v_x,v_y)$, the spatiotemporal pattern can be written $i(x-v_x.t,y-v_y.t)$ and its Fourier transform $\mathcal{I}(f_x,f_y,f_t)$, after the spatial sampling, can be written as:

$$\mathcal{I}(f_x,f_y,f_t) = [\, I(f_x,f_y) \,.\, \delta(f_t+v_x.f_x+v_y.f_y) \,] * \delta_{1,1}(f_x,f_y) \qquad (3.1)$$

where $\delta(.)$ is the Dirac distribution, $\delta_{1,1}(.,.)$ is the 2-D Dirac brush and means that the spectrum becomes periodic due to the sampling. The resulting aliasing is in fact strongly reduced by the crystalline lens filtering and by an irregular sampling [12]. So, for simplicity, let us drop the term $\delta_{1,1}(.,.)$ and let us interest in the main period.
If this spatiotemporal pattern is applied as input to the retinal filter, we obtain a retinal output s(x,y,t) that is expressed in the frequency domain by:

$$S(f_x,f_y,f_t) = G(f_x,f_y,f_t) \,.\, I(f_x,f_y) \,.\, \delta(f_t+v_x.f_x+v_y.f_y) \qquad (3.2)$$

The term $\delta(f_t+v_x.f_x+v_y.f_y)$ means that the output is only defined for $f_t = -v_x.f_x-v_y.f_y$. Thus, (3.2) can be rewritten as:

$$S(f_x,f_y,f_t) = G(f_x,f_y,-v_x.f_x-v_y.f_y) \,.\, I(f_x,f_y) \,.\, \delta(f_t+v_x.f_x+v_y.f_y) \qquad (3.3)$$

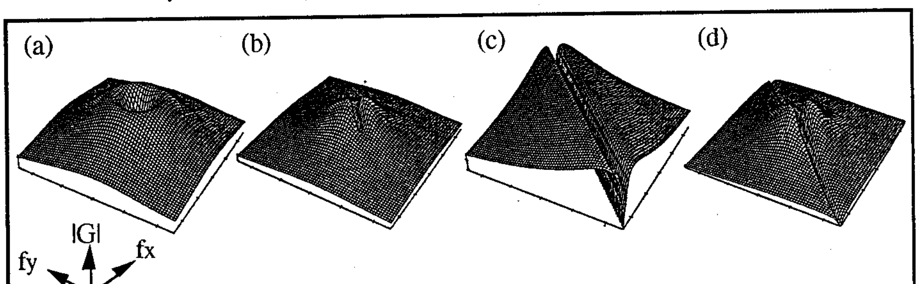


Fig. 3. Frequency responses for (a) the spatial OPL filter $G(f_x,f_y,0)$ to a static pattern, (b) the OPL filter $G(f_x,f_y,-v_x.f_x-v_y.f_y)$ for a moving image, (c) the spatial aspect of a pure temporal filter $\Psi(f_t)$ under same motion and (d) the result of $G.\Psi$ for the moving image.

Therefore, $G(f_x,f_y,-v_x.f_x-v_y.f_y)$ expresses a spatial retinal transfer function related to a given velocity. We have drawn in fig. 3b this spatial filter for a given velocity. This curve shows a preferential direction (according to the basin between the two peaks), that is perpendicular to the direction of the moving object. That means simply that the OPL filter enhances the spectral components of the moving image, in the direction of movement, leaving unchanged the spectral components perpendicular to the direction of movement. For a fixed object, no orientation is privileged (Fig. 3a).

3.2 Matched Filter to The Retinal OPL Response

Let us consider now a more realistic (noisy) input image: a moving pattern i(x,y,t) with an additive independent noise n(x,y,t). The response of the retinal OPL to this new pattern can be written as:

$$s_n(x,y,t) = g(x,y,t) * [\, i(x-v_x.t,y-v_y.t) + n(x,y,t) \,]$$

$$s_n(x,y,t) = s(x,y,t) + g(x,y,t) * n(x,y,t) = s(x,y,t) + n'(x,y,t)$$

ESANN'1993 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 7-8-9 April 1993, D-Facto public., ISBN 2-9600049-0-6, pp. 129-136

Let us search a filter $\psi(x,y,t)$ matched to the signal of interest $s(x,y,t)$:

$$\psi(x,y,t) * s_n(x,y,t) = \psi(x,y,t) * s(x,y,t) + \psi(x,y,t) * n'(x,y,t)$$

In order to maximise the signal-to-noise ratio at time t and at the position (x,y), the matched filter must satisfy [13]: $\psi(x,y,t) = s(-x,-y,-t)$, that is in the frequency domain:

$$\Psi(f_x,f_y,f_t) = S^*(f_x,f_y,f_t) = \text{Complex Conjugate of } S(f_x,f_y,f_t)$$

$$\Psi(f_x,f_y,f_t) = G^*(f_x,f_y,-v_x.f_x-v_y.f_y) . I^*(f_x,f_y) . \delta(f_t+v_x.f_x+v_y.f_y)$$

Such a filter would be only matched with a given image $I(f_x,f_y)$, moving in a given direction with a given velocity (v_x,v_y): it is obviously unrealistic. But one can imagine a filter matched only to the velocity, that is respecting only the form of G where f_t intervenes:

$$\Psi(f_x,f_y,f_t) \approx G^*(0,0,f_t) = G(0,0,-f_t)$$

In other words, using (2.2), we obtain:

$$\Psi(f_t) = (a - j2\pi f_t.\tau_1) / [\ (1+a-j2\pi f_t.\tau_1).(1+b-j2\pi f_t.\tau_2)\ ]$$

If we suppose that a is small and if we keep a causal form for spatial low-pass components, we obtain $\Psi(f_t)$ a time derivative filter convolved with a time low-pass filter:

$$\Psi(f_t) = (-\ j2\pi f_t.\tau_1) / [\ (1+j2\pi f_t.\tau_1).(1+b+j2\pi f_t.\tau_2)\ ]$$

Because of motion, we use $f_t = -v_x.f_x-v_y.f_y$. Hence, this temporal filter is equivalent to a purely spatial filter matched to a given velocity at least for low spatial frequencies. Fig. 3c shows the magnitude of this spatial filter. This shape simply expresses that the time derivative of the retinal filter is naturally matched to the velocity vector of the moving objects. Notice that the low-pass component helps keeping a low level of noise.

3.3 The Time Derivative: A Matched Filter for Motion Detection

We have stressed previously the fact that a kind of time derivative filtering can act as a matched filtering after the retinal processing. Let us consider now any filter $\psi(t)$ of the time derivative type. Let us apply it to the retinal output $s(x,y,t)$. Thus, we obtain a new output $s'(x,y,t)$, that is, if $\psi(t)$ is the ideal time derivative:

$$s'(x,y,t) = \partial s(x,y,t)/\partial t = \partial[g(x,y,t) * i(x,y,t)]/\partial t$$

which can be written in the frequency domain, using (3.3):

$$S'(f_x,f_y,f_t) = j2\pi f_t . G(f_x,f_y,-v_x.f_x-v_yf_y) . I(f_x,f_y) . \delta(f_t+v_x.f_x+v_y.f_y) \quad (3.4)$$

This last expression, using the identity $f_t = -v_x.f_x-v_y.f_y$, becomes:

$$S'(f_x,f_y,f_t) = j2\pi . (\ -v_x.f_x - v_y.f_y\ ) . S(f_x,f_y,f_t) \quad (3.5)$$

Therefore, the time derivative filter is equivalent to a space derivative filter if the input image is a translational moving object. Fig. 3d shows the result of the convolution of the retinal filter and the time derivative filter. It is very easy to verify that the motion detection is performed: let us consider a fixed object, that is with a null velocity (v_x=0 & v_y=0). The expression (3.3), which denotes the frequency response of the retinal OPL filter, becomes:

$$S(f_x,f_y,f_t) = G(f_x,f_y,0) . I(f_x,f_y) . \delta(f_t) \neq 0$$

ESANN'1993 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 7-8-9 April 1993, D-Facto public., ISBN 2-9600049-0-6, pp. 129-136

Thus, the retinal OPL filter is not sufficient for the motion detection (see at right-bottom of fig. 2). On the contrary, the expression (3.5), which denoted the frequency response of the time derivative filter applied to the retinal filter, gives:

$$S'(f_x,f_y,f_t) = j2\pi \,.\, 0 \,.\, G(f_x,f_y,0) \,.\, I(f_x,f_y) \,.\, \delta(f_t) = 0$$

Applying a time derivative filtering on the output of the retinal filtering is then a natural way to detect moving objects and more particularly their motion-oriented components (Compare figures 3b and 3d).

3.4 The Directional Space Derivative: A Matched Filter to the Motion Direction

We have shown previously with (3.4) and (3.5) that the time derivative filtering is equivalent to a spatial derivative filtering. (3.5) can be rewritten as:

$$S'(f_x,f_y,f_t) = \mathcal{H}(f_x,f_y,v_x,v_y) \,.\, S(f_x,f_y,f_t) \qquad (3.6)$$
$$\text{with} \quad \mathcal{H}(f_x,f_y,v_x,v_y) = H(v_x.f_x + v_y.f_y) \text{ and } H(\varphi) = -j2\pi.\varphi$$

$\mathcal{H}(f_x,f_y,v_x,v_y)$ denotes the spectrum of the spatial filter equivalent to the time derivative. However, a filter with an impulse response written as $f(x,y,a,b) = f(a.x+b.y) \,.\, \delta(a.y-b.x)$ has a Fourier transform written as $\mathcal{F}(f_x,f_y,a,b) = \rho^{-2}.\, F([a.f_x+b.f_y] \,.\, \rho^{-2})$ with $\rho^2 = a^2 + b^2$. Fig. 4 shows such a filter: it is a spatial directional filter along the line given by the equation a.y-b.x = 0.

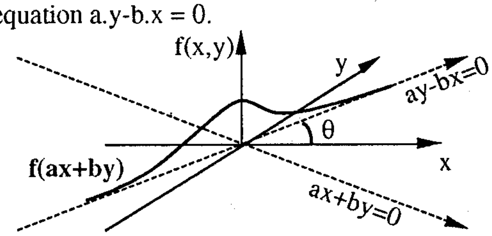


Fig. 4. Form of the filter given by $f(x,y) = f(a.x+b.y) \,.\, \delta(a.y-b.x)$

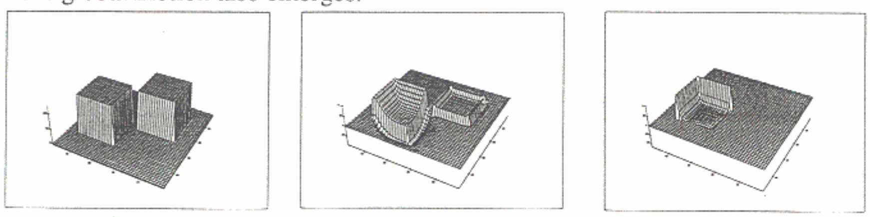
In expression (3.6), $\mathcal{F}$, F, a and b can be identified with $\mathcal{H}$, H, v_x and v_y respectively. Thus the temporal derivative, which is a filtering matched to the motion, is equivalent to the directional spatial derivative of the image along the line given by the equation $v_x.y-v_y.x = 0$, that is in the direction of the velocity vector. Therefore, this directional filter (a directional gradient) is also a filter matched to the motion, moreover, because it is purely spatial, it can be designed to match exactly the anti-causal form of $\Psi(f_t)$ in section 3.2. A set of directional filters covering the whole 2D space can then be naturally designed. The more matched filter to the direction of the motion will have the strongest response. This set of filters can be worked out after the time high-pass filter (Y-pathway) or even directly on the OPL output (X-pathway).

4. Results

The following results (more qualitative than quantitative) do not intend to show that the retina is capable of performing the best motion detection. Instead we want to emphasise that it is a neural structure that must inspire the designers of smart visual sensors in order to realise some early visual stages in a quick and simple way.
Figure 5a shows an image sequence at a given time. It consists of two squares with the same light intensity, one moving (the left one) and the other being motionless (the right one). Figure 5b shows the result of the OPL filtering (the bipolar image) on this

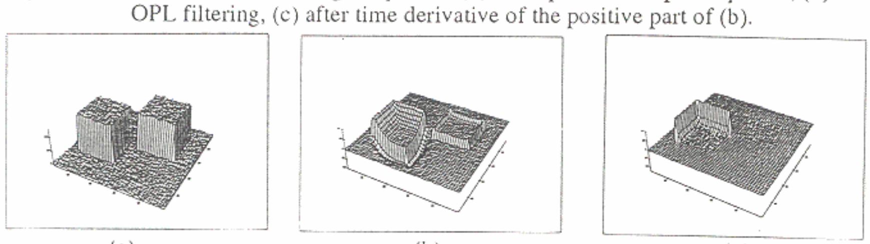
ESANN'1993 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 7-8-9 April 1993, D-Facto public., ISBN 2-9600049-0-6, pp. 129-136

sequence. The fixed object goes through a purely spatial band-pass filter (see fig. 3a): the edges are enhanced and the homogeneous areas (low spatial frequencies) are removed. The moving object goes through a different filter: the spatial components in the direction of the motion are accentuated, while the more low-pass ones decrease with the adaptation. Moreover, a negative trace, that disappears with time, follows the moving object. Figure 5c is the result of a time derivative filtering on the positive part of 5b: the fixed object is removed and the moving object remains. Figures 6a, 6b and 6c show the result on a noisy version of the same sequence.
In figures 7a, 7b and 7c (respectively 8a, 8b and 8c) a spatial gradient was performed on the OPL output (bipolar image), on its positive part and on the ON-IPL output (time derivative of this positive part): the local gradient is represented by a vector and the result is suspiciously like the optical flow. The local motion does not only appears but the global motion also emerges.



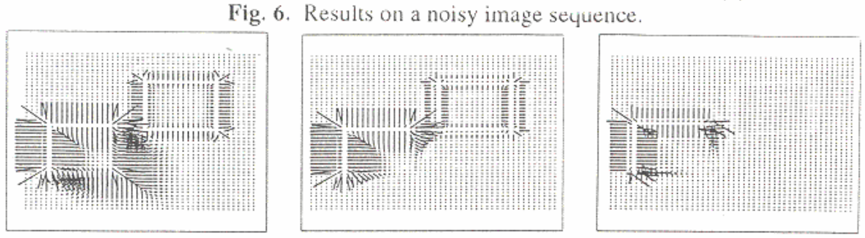
(a) (b) (c)

Fig. 5. Results on a noiseless image sequence: (a) a sample of the input sequence, (b) after OPL filtering, (c) after time derivative of the positive part of (b).



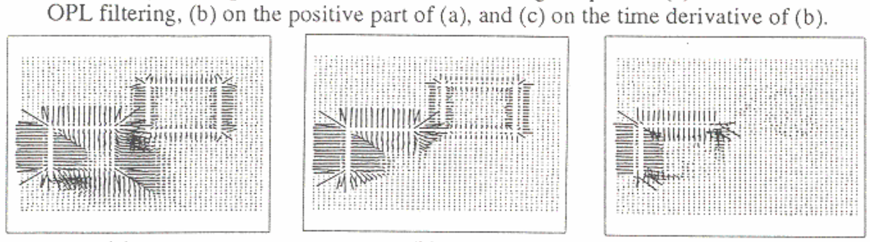
(a) (b) (c)

Fig. 6. Results on a noisy image sequence.



(a) (b) (c)

Fig. 7. Spatial gradient performed for the noiseless image sequence: (a) on the result of the OPL filtering, (b) on the positive part of (a), and (c) on the time derivative of (b).



(a) (b) (c)

Fig. 8. Spatial gradient performed for the noisy image sequence.

ESANN'1993 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 7-8-9 April 1993, D-Facto public., ISBN 2-9600049-0-6, pp. 129-136

5. Discussion

A bionic-inspired solution has been presented for the motion detection and the directional selectivity. A retinal model is used to regularise the input image. Time and directional derivative filters, which are naturally matched to the motion, detect the moving objects after the retinal regularisation. All the results fit with neurobiological evidences. Our retinal filtering might be a model for the processing that occurs in the Outer Plexiform Layer of the vertebrate retina. The time derivative filtering might model the processing of some amacrin cells in the Inner Plexiform Layer [14] while the directional selectivity might occur in area 17 of the visual cortex. Furthermore, the output of the filter g can be split up into two sustained channels X-ON and X-OFF to distinguish dark and bright objects, just as the output of the time derivative filtering into two transient channels Y-ON and Y-OFF to distinguish dark and bright moving objects. Our work results from the use of the frequency representation of the motion $f_t+v_x.f_x+v_y.f_y = 0$. It is a frequency generalisation of the well-known Horn & Schunck equation, $\partial i(x,y,t)/\partial t = - \mathbf{V} \cdot \nabla \mathbf{i}(x,y,t)$, that is the classical gradient equation [5] which leads mainly to an unreliable estimation of the local velocity in the direction of the spatial gradient $\nabla \mathbf{i}$ (aperture problem) due to its very high sensitivity to spatiotemporal noise. Efficient numerical simulations have been worked out to confirm these theoretical results. VLSI implementation of this processing and application to the global motion estimation are yet to be considered.

References

1. R. Emerson, J. Bergen, E. Adelson: Directionally selective complex cells and the computation of motion energy in cat visual cortex. Vision Res. 32, 203-218 (1992)
2. T. Poggio, V. Torre, C. Koch: Computational vision and regularization theory. Nature, 317, 314-319 (1985)
3. J. Marroquin, S. Mitter, T. Poggio: Probabilistic solution of ill-posed problems in computational vision. J. Am. Stat. Assoc., 82:397, 76-89 (1987)
4. A. Yuille: Energy Functions for Early Vision and Analog Networks. Biol. Cybern. 61, 115-124 (1989)
5. B. Horn, B. Schunck: Determining Optical Flow. Artif. Intel. 17, 185-203 (1981)
6. J. Hutchinson, C. Koch, J. Luo, C. Mead: Computing Motion Using Analog and Binary Resistive Networks. IEEE Computer, 52-63 (1988)
7. E. Hildreth, C. Koch: The analysis of visual motion: From computational theory to neuronal mechanisms. Ann. Rev. Neurosci., 10:477-533 (1987)
8. C. Mead, M. Mahowald: A silicon model for early visual processing. Neural Networks, 1:1, 91-97 (1988)
9. J. Dowling: Neurons and Networks: An Introduction to Neuroscience. The Belknap Press of Harvard University Press, Cambridge, Massachusetts (1992)
10. R. Siminoff: An analogue model of the luminosity-channel in the vertebrate cone retina: 1. hardware and responses to square wave voltages. Biol. Cybern. 46, 101-110 (1983)
11. Srinivasan MV, Laughlin SB, Dubs A: Predictive coding: a fresh view of inhibition in the retina. Proc. R. Soc. London B216, pp 427-459 (1982)
12. H. Wässle, B. Boycott: Functional architecture of the mammalian retina. Physiological Reviews, 71:2, 447-480 (1991)
13. F. de Coulon: Théorie et traitements des signaux. Dunod (1984)
14. J. Richter, S. Ullman: A model for the temporal organization of X- and Y-type receptive fields in the primate retina. Biol. Cybern. 43, 127-145 (1982)