An Efficient Neuromorphic Analog Network for Motion Estimation

Antonio B. Torralba and Jeanny Hérault

Abstract— **Optical flow estimation is a critical mechanism for autonomous mobile robots as it provides a range of useful information. As real-time processing is mandatory in this case, an efficient solution is the use of specific VLSI analog circuits. This paper presents a simple and regular architecture based on analog circuits which implements the entire processing line from photoreceptor to accurate and reliable optical flow estimation. The algorithm we propose, is an energy-based method using a novel wideband velocity-tuned filter which proves to be an efficient alternative to the well known Gabor filters. Our approach shows that a high level of accuracy can be obtained from a small number of loosely tuned filters. It exhibits similar or improved performance to that of other existing algorithms, but with a much lower complexity.**

Keywords— **Optical flow estimation, neuromorphic systems, velocity-tuned filters, aperture problem, VLSI motion chips**

I. Introduction

Motion estimation refers to the computation of velocity vectors (optical flow) at each pixel. In a mobile robot, motion perception can provide a range of useful information, such as egomotion, time to collision, detection of moving objects, 3D structure of the enviroment, ... However, in order to be adequately estimated, these items may require accurate measurements of optical flow.

A powerful solution for real-time processing is the realization of specific VLSI circuits. Motion estimation algorithms on silicon require a compromise between the number of pixels in the input image and the complexity of each processing unit. Up to now, only simple motion algorithms have been implemented using analog circuits, see [18] for a review.

Energy-based algorithms [1], [27], [8] are known to be robust in the face of noise and aliasing, they give reliable measurements of velocity and they allow an easy treatment of the aperture problem. However, due to the complexity of implementing a battery of spatiotemporal filters, current VLSI motion chips use gradient-based algorithms [24], [6] or correlation-based algorithms [12], [7], [9] as they can be implemented within very compact circuits.

In this paper we present a new energy-based algorithm that significantly minimizes the complexity of these kinds of methods. The reduction of complexity is due to: 1) the use of a new wideband velocity-tuned filter (VTF) simpler than the narrow band spatiotemporal Gabor filters usually used in energy-based algorithms. 2) A simple circuit for energy and velocity estimation. This paper focuses on the theorical aspects of the approach in order to give the basis for low complexity energy-based algorithms.

A. B. Torralba and J. Hérault are with the Laboratoire des Images et des Signaux (LIS), Institut National Polytechnique de Grenoble, 46 avenue Félix Viallet, 38031 Grenoble Cedex, France. E-mail: {torralba, herault}@tirf.inpg.fr

The paper is organized as follows: section II presents the model of motion and the basic theory of VTF's. Section III introduces a simple analog network that implements a wideband VTF using four neighbor interactions. Section IV describes how this network is used for motion estimation, highlighting the aperture problem and providing a complete scheme for motion estimation, well tailored for CNN analog circuits. Finally, section V shows that the results compare favourably to those of other more complex energy-based algorithms.

II. Translational motion and the theory of velocity-tuned filters

A basic model of motion assumes that the brightness signal translates with constant velocity and direction. In such a case we can write: $e(\mathbf{x},t) = e(\mathbf{x}-\mathbf{v}t)$, where e is the brightness function, $\mathbf{x} = (x,\, y)^T$ are the spatial variables and t the temporal variable, $\mathbf{v} = (v_x,\, v_y)^T$ is the velocity vector and T means transpose. By successively applying the Fourier transform to the spatial and temporal variables, we obtain:

$$E(\mathbf{f}_s\,,\, f_t) = E(\mathbf{f}_s)\,\delta(f_t + \mathbf{v}^T\mathbf{f}_s) \tag{1}$$

where $\mathbf{f}_s = (f_x,\ f_y)^T$ represents the spatial frequency vector, f_t the temporal frequency, $E(\mathbf{f}_s)$ the spatial Fourier transform of the static brightness pattern $e(\mathbf{x})$ and $\delta(.)$ is the Dirac delta distribution. The power of the signal lies on a plane passing through the origin [27] with the equation $f_t + \mathbf{v}^T\mathbf{f}_s = 0$.

Energy-based methods use a set of filters sampling the frequency domain in order to detect the orientation of the energy plane. Different filter types have been proposed in the literature: a) spatiotemporal frequency tuned filters such as Gabor filters [8], [20], [23], b) velocity-tuned filters [5], [26], c) space-velocity separable filters [22], [21]. In this paper, we will focus on the velocity-tuned filters since they can yield simpler architectures than the other two approaches.

The output of a spatiotemporal filter $H(\mathbf{f}_s\,,\, f_t)$ to a moving pattern can be writen as:

$$S(\mathbf{f}_s\,,\, f_t) = H(\mathbf{f}_s\,, -\mathbf{v}^T\mathbf{f}_s)\,E(\mathbf{f}_s)\,\delta(f_t + \mathbf{v}^T\mathbf{f}_s) \tag{2}$$

meaning that the input is filtered by an equivalent spatial filter with transfer function $H(\mathbf{f}_s\,, -\mathbf{v}^T\mathbf{f}_s)$.

In order to estimate the local mean power of the filter output, we consider that the integration window is sufficiently wide the that local mean power approximates to

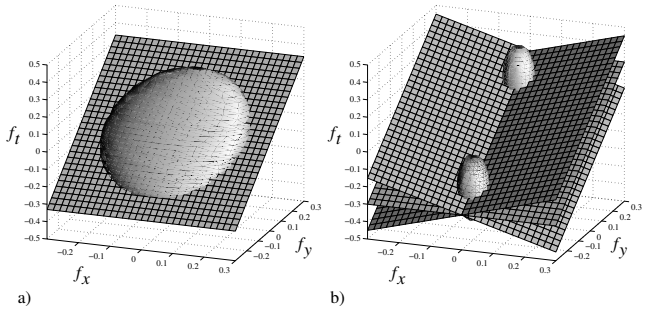


Fig. 1. -3dB section of the spatiotemporal frequency magnitude of: a) velocity-tuned filter, the grid indicates the orientation of the energy plane matched to the velocity-tuned filter; b) frequency tuned filter, several planes maximises of the output energy.

the total mean power. For a window of infinite size, the output power is:

$$P = \int s(\mathbf{x},t)\,d\mathbf{x} = \int \Gamma_e(\mathbf{f}_s)\,\left|H(\mathbf{f}_s,-\mathbf{v}^T\mathbf{f}_s)\right|^2\,d\mathbf{f}_s \qquad (3)$$

where Γ_e and Γ_s are the spatial power density spectrum of the input and output signals. Total mean power P does not depend on time because temporal and spatial frequencies are linked through motion.

This output power is a function of input velocity: $P(\mathbf{v})$. We define a *velocity-tuned filter* (VTF) for velocity $\mathbf{v}_o$ as a filter with mean ouput power $P(\mathbf{v})$ having a unique maximum for some input velocity $\mathbf{v}=\mathbf{v}_o$, independently of the spectral content of the input signal. Thus, the magnitude of a VTF $H_{\mathbf{v}_o}(\mathbf{f}_s,\ f_t)$ must have a unique maximum at $f_t=-\mathbf{v}_o^T\mathbf{f}_s$, for any spatial frequency $\mathbf{f}_s$.

The transfer function of a VTF for velocity $\mathbf{v}_o$ may be written as:

$$H_{\mathbf{v}_o}(\mathbf{f}_s,\ f_t) = H_{\mathbf{0}}(\mathbf{f}_s,\ f_t+\mathbf{v}_o^T\mathbf{f}_s) \qquad (4)$$

where $H_{\mathbf{0}}(\mathbf{f}_s,\ f_t)$ is a VTF for to null velocity. This is a direct consequence of the definition of VTF and equation (3). Separable Gabor filters do not verify this property as they are frequency-tuned and not velocity-tuned. Figure 1 shows the difference between a wideband VTF and a narrow band frequency-tuned Gabor filter.

III. Velocity-tuned analog network

We are interested in filters that can be easily implemented as analog circuits. Therefore, the filter must be of low order to reduce connectivity. This is the case for the analog RC network of figure 2.a, often used in vision chips [13], [17].

The RC network is a low-pass spatiotemporal filter. As shown in figure 2.a, each output node is connected to the input via a resistor r, to its four neighbors via resistors R and to the ground via a capacitor C. Nodes are indexed by the discrete spatial variables n and m. By applying the Kirchoff currents' law at the output node (n,m), we obtain:

$$e_{n,m}(t) = s_{n,m}(t) + \gamma[4\,s_{n,m}(t) - s_{n-1,m}(t) - s_{n+1,m}(t) - s_{n,m-1}(t) - s_{n,m+1}(t)] + \tau\,ds_{n,m}(t)/dt \qquad (5)$$

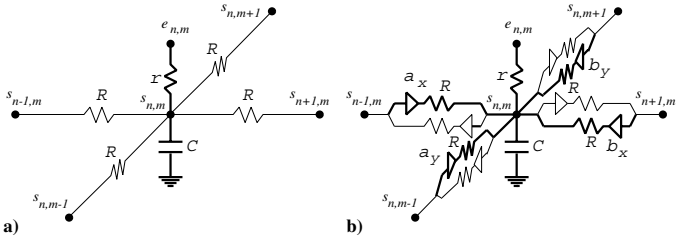


Fig. 2. a) RC network. b) Velocity-tuned analog network. Asymmetrical interactions are responsible for the velocity tuning.

where $\gamma = r/R$ and $\tau = rC$. Equation (5) is discrete in space but continuous in time. It can be considered as an approximation of the continuous equation:

$$e(x,y,t) = s(x,y,t) - \gamma\ \Delta s(x,y,t) + \tau\ \partial s(x,y,t)/\partial t \qquad (6)$$

where Δ is the spatial Laplacian operator. Applying the Fourier transform to equation (6) gives the transfer function of the analog network (valid for low spatial frequencies):

$$H_{\mathbf{0}}(\mathbf{f}_s,\ f_t) = \frac{1}{1+4\pi^2\gamma\left|\mathbf{f}_s\right|^2 + j2\pi\tau f_t} \qquad (7)$$

Due to its low-pass spatiotemporal characteristic, the output energy will be at a maximum for static inputs. This filter verifies the definition of a filter tuned to null velocity. By applying equation (4), it is then possible to steer it to an arbitrary velocity $\mathbf{v}_o$:

$$H_{\mathbf{v}_o}(\mathbf{f}_s,\ f_t) = \frac{1}{1+4\pi^2\gamma\left|\mathbf{f}_s\right|^2 + j2\pi\tau(f_t+\mathbf{v}_o^T\mathbf{f}_s)} \qquad (8)$$

This low-pass function (fig. 1.a) is oriented in the spatiotemporal frequency space. From the inverse Fourier transform of equation (8), we derive the following differential equation:

$$e(x,y,t) = s(x,y,t) - \gamma\ \Delta s(x,y,t) + \tau\left(\mathbf{v}_o^T\nabla s(x,y,t) + \partial s(x,y,t)/\partial t\right) \qquad (9)$$

where ∇ is the spatial gradient operator. Implementation of this filter as an analog network demands that we return to a discrete approximation of the spatial derivatives, the temporal derivative being implemented by means of the capacitor. For the spatial derivatives, we use the following approximations: $\partial s/\partial x \simeq [s_{n+1,m}-s_{n-1,m}]/2$ and $\partial^2 s/\partial x^2 \simeq s_{n+1,m} - 2\ s_{n,m} + s_{n-1,m}$, and the same for $\partial s/\partial y$ and $\partial^2 s/\partial y^2$. The distance between samples is one spatial unit. Replacing the approximations of derivatives into equation (10) and grouping the terms with the same indices, we obtain:

$$e_{n,m}(t) = s_{n,m}(t) + \gamma[4s_{n,m}(t) - a_x s_{n-1,m}(t) - b_x s_{n+1,m}(t) - a_y s_{n,m-1}(t) - b_y s_{n,m+1}(t)] + \tau ds_{n,m}(t)/dt \qquad (10)$$

where:

$$a_x = 1+\frac{v_{x_o}\tau}{2\gamma},\ a_y = 1+\frac{v_{y_o}\tau}{2\gamma},\ a_x+b_x = a_y+b_y = 2 \qquad (11)$$

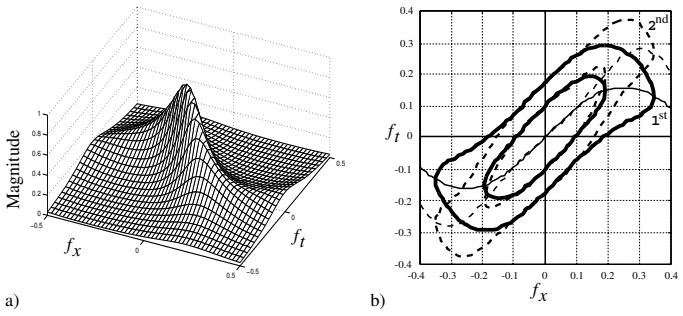


Fig. 3. a) Magnitude of the spatiotemporal transfer function of the VTF ($f_y = 0$). b) Contour diagram at -3dB and -6dB for the transfer function obtained with the first order (solid lines) and second order (broken lines) approximation of the spatial derivative.

Equation (10) is implemented by the network of figure 2.b, with only four neighbor interactions. The VTF is specified by 4 parameters: γ = spatial scale, τ = temporal scale and $\mathbf{v}_o^T = (v_{x_o}, v_{y_o})$ = tuning velocity which are a function of the circuit parameters; r, R, C, a_x, b_x, a_y and b_y.

Due to the discrete nature of the spatial derivative in equation (10), there is some distortion of the transfer function of the analog network with respect to equation (8). By applying the Fourier transform to equation (10) we obtain the transfer function:

$$H(\mathbf{f}_s, f_t) = \frac{1}{P(\mathbf{f}_s) + jQ(\mathbf{f}_s, f_t)} \tag{12}$$

where $j = \sqrt{-1}$, P and Q being the following real functions:

$$\begin{aligned} P(\mathbf{f}_s) &= 1 + \gamma[4 - (a_x + b_x)cos(2\pi f_x) \\ &\quad -(a_y + b_y)cos(2\pi f_y)] \\ Q(\mathbf{f}_s, f_t) &= 2\pi\tau f_t + \gamma(a_x - b_x)sin(2\pi f_x) \\ &\quad +\gamma(a_y - b_y)sin(2\pi f_y) \end{aligned} \tag{13}$$

The spatial frequencies f_x and f_y are given in cycles/pixel and the temporal frequency f_t in cycles/second. The function Q is responsible for the velocity tuning of the filter. For low spatial frequencies we can approximate $Q(\mathbf{f}_s, f_t)/(2\pi\tau) \simeq f_t + \mathbf{v}_o^T\mathbf{f}_s$ with $v_{x_o} = \gamma(a_x - b_x)/\tau$ and $v_{y_o} = \gamma(a_y - b_y)/\tau$, which are the two components of the tuning velocity. Using more points than in equation (10) to approximate the spatial derivatives would increase the range of spatial frequencies for which this aproximation is valid. This would also increase the complexity of the filter, each node being connected to more than four neighbors.

When $Q(\mathbf{f}_s, f_t) = 0$, the function $P(\mathbf{f}_s)$ determines the spatial frequency form of the transfer function (12). For H being a low-pass filter, a_x, b_x, a_y and b_y verify that $a_x + b_x > 0$ and $a_y + b_y > 0$. Figure 3.a shows the magnitude of the transfer function (12).

Figure 3.b shows the results obtained by approximating the spatial derivative with two and four points: if the filter is not very narrow around the plane equation $f_t + \mathbf{v}_o^T\mathbf{f}_s = 0$ (low velocity selectivity), the two approximations are almost equivalent for spatial frequencies below 0.2 cycles/pixel.

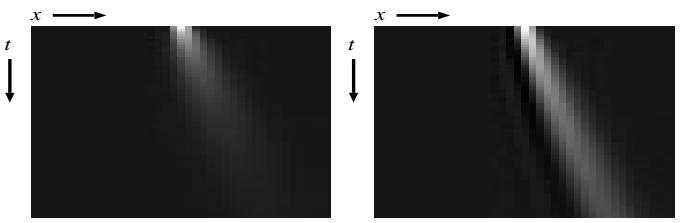


Fig. 4. Spatiotemporal impulse responses of two filters with low velocity selectivity, left, and another with high velocity selectivity, right.

A. Stability

As B. Shi [21] has shown in a more general framework, the network is stable if, for each spatial frequency $\mathbf{f}_s$, the filter exhibits temporal stability. That is, if in equation (12), we replace $j2\pi f_t$ by s (the Laplace complex variable), stability requires that the roots of the denominator $P(\mathbf{f}_s) + jQ(\mathbf{f}_s, s/(j2\pi)) = 0$ for $s \in \mathbb{C}$, all lie on the open left-half side of the complex plane. This requires $P(\mathbf{f}_s) > 0$, that is: $|a_x + b_x| + |a_y + b_y| < 4 + 1/\gamma$. This condition is already satisfied since $a_x + b_x = a_y + b_y = 2$ as given in (11).

B. Spatiotemporal impulse response

An approximation of the spatiotemporal impulse response of the velocity-tuned analog network can easily be calculated by applying the inverse Fourier transform to the transfer function (8), the low spatial frequency approximation. The result is:

$$h_{\mathbf{v}_o}(\mathbf{x}, t) \simeq A(t)e^{-|\mathbf{x} - \mathbf{v}_o t|^2/\sigma^2(t)}U(t) \tag{14}$$

where: $A(t) = e^{-t/\tau}/(4\pi\gamma t)$, $\sigma^2(t) = 4\gamma t/\tau$ and $U(t)$ is the Heaviside step. The approximation is valid for a filter with low velocity selectivity and with a spatial bandwidth lower than 0.2 cycles/pixel. The impulse response is a spatial Gaussian signal which varies causally with time and propagates in space with the tuning velocity of the filter. The amplitude of the Gaussian, $A(t)$, decreases with time at a rate controlled by the time constant τ. The spatial width of the Gaussian, $\sigma^2(t)$, increases linearly with time. At $t = 0$ the impulse response is a Dirac delta distribution.

Velocity selectivity refers to the sensitivity of the filter output to differences in the input velocity. High velocity selectivity requires a narrow shape around the plane $f_t + \mathbf{v}_o^T\mathbf{f}_s = 0$ and is obtained by increasing the value of τ while keeping constant the other filter parameters, γ and $\mathbf{v}_o$. But this increases the duration of the impulse response and when the filter is very selective, the first order approximation of the spatial derivative introduces a distorsion, giving some low energy oscillations (see figure 4). Furthermore, a longer impulse response duration would damp short duration motions.

C. Moving input: velocity-tuned filter

As shown in equation (2), the spatial response of a spatiotemporal filter to an input with constant velocity $\mathbf{v}$ can be calculated by filtering the input image with an

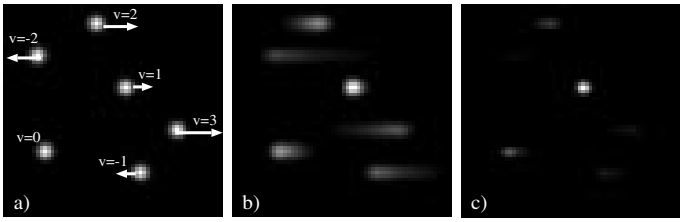


Fig. 5. Output of a velocity-tuned analog network for an input composed of several dots moving with different velocities. a) input image. b) Output of a filter tuned to velocity $v_{o_x} = 1$ and $v_{o_y} = 0$. Each dot presents motion blur, except that moving at the tuned velocity of the filter. The blurring has been exagerated for visibility by making the filter very selective (large value of τ). c) Squared output.

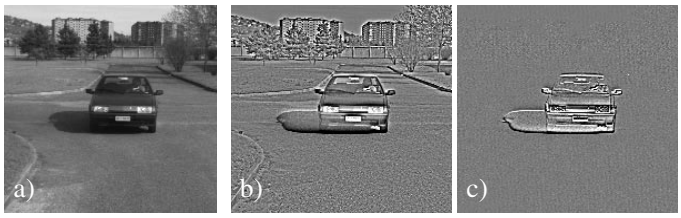


Fig. 6. Retinal prefiltering: a) input image, b) output of bipolar cells and c) temporal derivative.

equivalent spatial filter. The VTF in the presence of a moving input corresponds to the spatial filter $G_\beta(\mathbf{f}_s) = H_{\mathbf{v}_o}(\mathbf{f}_s, -\mathbf{v}^T\mathbf{f}_s)$. If we use the low spatial frequency approximation given in equation (8) we obtain:

$$G_\beta(\mathbf{f}_s) = \frac{1}{1 + 4\pi^2\gamma\,|\mathbf{f}_s|^2 + j2\pi\tau\boldsymbol{\Delta}\mathbf{v}^T\mathbf{f}_s} \qquad (15)$$

where $\boldsymbol{\Delta}\mathbf{v} = \mathbf{v}_o - \mathbf{v} = |\boldsymbol{\Delta}\mathbf{v}|\,(\cos\beta, \sin\beta)^T$ is the difference between the tuning velocity and the input velocity. This is an oriented spatial filter in the direction of $\boldsymbol{\Delta}\mathbf{v}$. For $|\boldsymbol{\Delta}\mathbf{v}| = 0$ the filter has a symmetrical response. As $|\boldsymbol{\Delta}\mathbf{v}|$ increases, the spatial filter reduces its spatial frequency bandwidth in the direction given by the angle β. Therefore, the output will be blurred in the direction β.

The mean output power is:

$$P(\boldsymbol{\Delta}\mathbf{v}) = \int \Gamma_e(\mathbf{f}_s)\;|G_\beta(\mathbf{f}_s)|^2\;d\mathbf{f}_s \qquad (16)$$

It will be at a maximum when the filter's transfer function $G_\beta(\mathbf{f}_s)$ has its greatest spatial bandwidth, i.e., when $|\boldsymbol{\Delta}\mathbf{v}|$=0. The maximum $-$3dB spatial bandwidth is $\Delta B = (9.6\,\pi^2\,\gamma)^{-1/2}$. The mean output power will also be maximized if $\boldsymbol{\Delta}\mathbf{v}^T\mathbf{f}_s = 0$ on the support of the input power spectrum $\Gamma_e(\mathbf{f}_s)$, that is, if the input spatial pattern $e(\mathbf{x})$ depends only on one spatial direction. This leads to the aperture problem, see next section.

IV. Motion estimation

Based on biological architecture where accuracy can be obtained by a small number of loosely tuned filters, we have developped a simple method based on wide-band tuning, formerly presented in [25]. As it is shown in figure 8, the algorithm is composed of four stages: a) the retinal prefiltering, b) velocity-tuned filters, c) local mean output power estimation for each filter and d) velocity estimation.

A. Prefiltering

As shown in equation (16), the output power depends on the spectral content of the input pattern. Atick and Redlich [2] show that natural images have a spectrum of the form $1/\,|\mathbf{f}_s|^\alpha$ and that retinal filtering compensates for this characteristic, by whittening the spectrum.

We use a model of the retina based on analog circuits as a preprocessing stage [4], [10] (see figure 8.a) consisting of two layers. The first, the receptor layer, computes a low-pass spatiotemporal filter, primarily to improve the signal to noise ratio. The second layer, the horizontal cells layer, computes a spatiotemporal average of the receptor output. The difference amplifiers model the bipolar cells (they compute the difference between the outputs of the receptor layer and of the horizontal cells layer). Therefore, overall the retina behaves as a spatiotemporal band-pass filter (see figure 6.b). Thus, for low frequencies, the filtering will compensate for the $1/\,|\mathbf{f}_s|^\alpha$ spectrum of the images. For high frequencies, the filtering will reduce noise.

As the proposed VTF responds to low spatiotemporal frequencies, band-pass prefiltering enhances the contrast between responses of different VTFs by cancelling the common part of the spatiotemporal transfer functions. In order that the low frequency approximation of the VTF be valid, the prefiltering must have a spatial bandwidth of $\Delta B < 0.2$ cycles/pixel which is the maximum allowed spatial bandwidth of the VTFs.

As many VLSI circuit implementations of the retina have already been proposed [14], [15], the same technologies apply to our VTF.

B. Local mean output power estimation

We are interested in local velocity estimation in order to deal with variations of the velocity field of the image. This requires an estimation of the local output power for each VTF by a local integration over a domain sufficiently wide to avoid the aperture problem. The integration window represents the "aperture" through which we look at the moving pattern. By reducing the window's size, we increase the possibility of losing pertinent information in order to estimate the full motion vector. By increasing it, we obtain smoothed velocity fields and can cancel the motion of small objects.

This spatial integration is performed by a resistive analog network (without capacitor) applied to the squared output of the VTF (figure 8.c). Local output power estimation at each pixel is given by the voltage at the corresponding node in the resistive network. The resistive network implements a low-pass spatial filter. The ratio $\gamma_i = r_i/R_i$ controls the size of the integration region. The -3dB frequency bandwidth of the integration network is $\Delta B = (9.6\,\pi^2\,\gamma_i)^{-1/2}$. Increasing γ_i increases the size of the integration region. In order to ensure that the integration region is larger than the

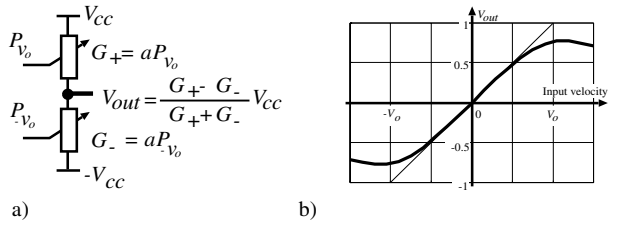


Fig. 7. Shunting inhibition is modeled by a voltage divider (a) where conductances are proportional to inputs. In the operating range, the output varies linearly with respect to velocity (b).

impulse response of the VTF, it is necessary that $\gamma_i \gg \gamma$ (γ is a parameter of the VTF defined in section III). The exact value of γ_i depends on the prominence of the aperture problem in the input images. For our simulations, $\gamma_i = 100\,\gamma$.

C. Component velocity estimation from two VTFs

In this section we propose a simple mechanism for estimating the components of the input velocity vector by combining the output powers of two loosely tuned VTFs.

We consider that after retinal prefiltering the input signal has a flat spectrum, $\Gamma_e(\mathbf{f}_s) = \Gamma$ constant. For a loosely tuned VTF, the integral of (16) can be approximated by the following expression:

$$P_{\mathbf{v}_o} = P(\mathbf{v} - \mathbf{v}_o) \simeq \frac{\Gamma}{\pi\gamma\sqrt{8}} \frac{1}{\sqrt{1 + |\mathbf{v} - \mathbf{v}_o|^2 / \Delta v_o^2}} \qquad (17)$$

being $\mathbf{v} = (v_x, v_y)^T$ the input velocity and $\mathbf{v}_o$ the velocity of tuning of the VTF. It can be verified numerically that the error remains within $\pm 5\%$. $P_{\mathbf{v}_o}$ is a function of $\mathbf{v}$ with its maximum at $\mathbf{v} = \mathbf{v}_o$. The velocity selectivity of the filter is $\Delta v_o^2 = 8\gamma/\tau^2$ and controls the shape of the function $P_{\mathbf{v}_o}$. Small values of Δv_o give a sharp maximum.

We use two filters tuned to velocities $\mathbf{v}_o = |\mathbf{v}_o|(\cos\theta, \sin\theta)^T$ and $-\mathbf{v}_o$. Velocity component in the direction θ is estimated using a voltage divider (shunting inhibition mechanism [16]) where each conductance is controlled by the output power of a VTF (figure 7.a). The upper conductance G_+ is the "excitatory" connection and is proportional to the output power of the filter tuned to $\mathbf{v}_o$. The lower conductance G_- is the "inhibitory" connection and is proportionnal to the output power of the filter tuned to $-\mathbf{v}_o$. The voltage at the output node is:

$$V_{out} = \frac{G_+ - G_-}{G_+ + G_-} V_{cc} = \frac{P_{\mathbf{v}_o} - P_{-\mathbf{v}_o}}{P_{\mathbf{v}_o} + P_{-\mathbf{v}_o}} V_{cc} \qquad (18)$$

where V_{cc} is a constant voltage and V_{out} is the output voltage. The numerator is strongly dependant on velocity. The denominator acts as a normalization term. For low velocities we can approximate V_{out} as:

$$V_{out} = \frac{\mathbf{v}^T \mathbf{v}_o}{\Delta v_o^2 + |\mathbf{v}_o|^2 + O(|\mathbf{v}|^2)} V_{cc} \qquad (19)$$

For velocities in the range $|\mathbf{v}| < |\mathbf{v}_o|$, the voltage V_{out} has a linear dependence on $\mathbf{v}^T \mathbf{v}_o / |\mathbf{v}_o| = |\mathbf{v}| \cos(\theta - \alpha) = v_\theta$,

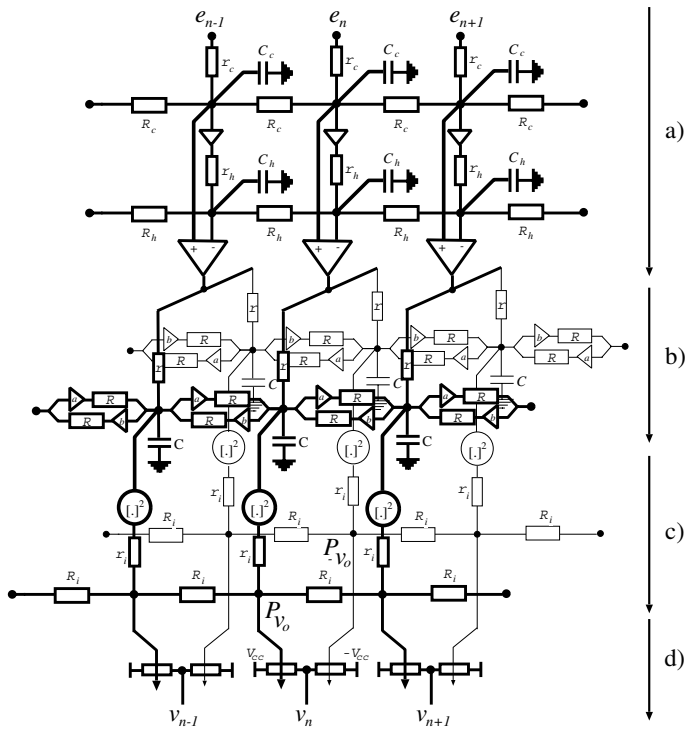


Fig. 8. Diagram of the velocity estimation algorithm for one spatial dimension from input brigthness e_n to velocity estimation v_n: a) retina prefiltering, b) two filters tuned to opposite directions v_o and $-v_o$, c) local energy integration by squaring filter outputs and filtering with a resistive analog network and d) velocity estimation with a shunting inhibition mechanism.

the component of input velocity for the direction θ. For different values of θ, estimated velocities are distributed on a circle (figure 9.a). For velocities larger than $|\mathbf{v}_o|$, V_{out} decreases to zero due to the term $O(|\mathbf{v}|^2)$ in the denominator which represents terms depending on high powers of $\mathbf{v}$. Figure 7.b shows the V_{out} characteristic with respect to $\mathbf{v}$. The non-linear characteristic will introduce some errors for velocities $|\mathbf{v}| \geq |\mathbf{v}_o|$.

Estimating two components of velocity in two orthogonal directions will be sufficient to give the input velocity vector. This can be achieved with four filters tuned to velocities $|\mathbf{v}_o|(\cos\theta, \sin\theta)^T$ with $\theta = 0, \pi/2, \pi$ and $3\pi/4$, providing a simple architecture for VLSI implementations. Figure 8 shows the complete diagram of the analog circuit for velocity estimation in one spatial dimension. Simulation results on real images (see section V) show that this method gives accurate results despite its simplicity.

In some situations, this simple method will fail to produce the correct motion vector. The flat spectrum hypothesis of the input patterns is unlikely for some natural images, even after whittening prefiltering. An extreme case is an input pattern with a spatial structure oriented in only one direction, that is $e(\mathbf{x}) \Rightarrow e(\mathbf{x}^T \mathbf{n})$, where $\mathbf{n} = (\cos\beta, \sin\beta)^T$, β being the direction of variation. In such a case, the pattern has a one dimensional spatial structure and the aperture problem will be present for all scales of analysis. Such a pattern has a spatial power spectrum of the form

$\Gamma_e(\mathbf{f}_s^T\mathbf{n})\ \delta(\mathbf{f}_s^T\mathbf{n}_\perp)$, where $\mathbf{n}_\perp = (\sin\beta, -\cos\beta)^T$. If we suppose that the brightness pattern has a flat spectrum in the direction of variation, that is $\Gamma_e(\mathbf{f}_s^T\mathbf{n}) = \Gamma$, as it will be the case after prefiltering, then, output power can be exactly calculated as:

$$P(\mathbf{v}-\mathbf{v}_o) = \frac{\Gamma\pi^2\sqrt{\gamma}}{\sqrt{2}} \frac{1}{\sqrt{1+(v_n-\mathbf{v}_o^T\mathbf{n})^2/\Delta v_o^2}} \qquad (20)$$

where $v_n = \mathbf{v}^T\mathbf{n}$ (normal velocity) is the component of velocity in the direction of variation of the pattern, β. Output power does not depend on the velocity component orthogonal to the normal velocity. By a Taylor developement in v_n we can approximate V_{out} by:

$$V_{out} = \frac{|\mathbf{v}_o|\, v_n \cos(\theta-\beta)}{\Delta v_o^2 + |\mathbf{v}_o|^2 \cos(\theta-\beta)^2 + O(v_n^2)} V_{cc} \qquad (21)$$

where θ is the direction of tuning of the filters and β is the apparent direction of motion (i.e. the direction of variation of the spatial input pattern), v_n is the component of the input velocity in the direction β, $O(v_n^2)$ represents the high order terms that can be ignored for low values of v_n. For different values of θ, we will obtain estimates distributed on an ellipse in polar coordinates passing through the origin and centered at $v_n/2\,(\cos(\beta), \sin(\beta))$. When computing the velocity, it would be necessary to integrate over large regions in order to minimize the aperture problem.

When the input consists in a pure translationnal sinusoid, it can be shown that V_{out} depends on the spatial frequency. This is an undesirable behavior as we are only interested in the velocity dependence. However, this is not a common input pattern when dealing with real images.

D. Component velocity estimation from three VTFs

Some of the limitations of the estimation performed with two filters can be overcome by using a third filter tuned to null velocity. The use of three filters allows to obtain a better linearity on the estimation, a simpler treatement of the aperture problem and eliminate the dependency on the input frequency for sine waves. Although this yields to a more complex combination of output powers of the VTFs, there is a significant improvement on accuracy, see section V. No analog circuit is proposed here. Velocity computation could be calculated by an external processor.

We propose the next expression in order to estimate velocity:

$$\widetilde{v}_\theta = \frac{|\mathbf{v}_o|}{2} \frac{P^2_{\mathbf{v}_o} - P^2_{-}\mathbf{v}_o}{P^2_{\mathbf{v}_o} + P^2_{-}\mathbf{v}_o - 2P^2_{\mathbf{v}_o}P^2_{-}\mathbf{v}_o/P^2_{\mathbf{0}}} \qquad (22)$$

$\widetilde{v}_\theta$ is an estimation of the velocity component of the input velocity $\mathbf{v}^T = |\mathbf{v}|\,(\cos\alpha, \sin\alpha)$ onto the direction θ. $P_{\pm\mathbf{v}_o}$ are as already defined, $P_{\mathbf{0}}$ is the output power of a VTF tuned to null velocity. The range of validity of equation (22) is limited by input noise and by the approximation seen in equation (17), which is more biased when input velocity $\mathbf{v}$ differs greatly from $\mathbf{v}_o$. As input velocity $|\mathbf{v}| > |\mathbf{v}_o|$, the mean output powers of the three filters decrease and measurements are affected by noise and approximation errors.

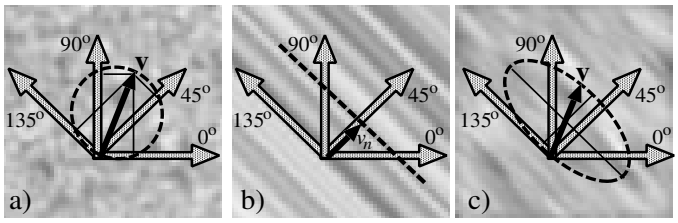


Fig. 9. Distribution of estimated components for 4 orientations ($\widetilde{v}_\theta$, for $\theta = 0,\ 45,\ 90$ and 135 degrees). The four components lie (a) on a circle for a flat spectrum, (b) on a line for a 1D pattern, (c) on an ellipse for an oriented texture.

In the case of a flat power spectrum input (17) we obtain:

$$\widetilde{v}_\theta \simeq \frac{\mathbf{v}^T\mathbf{v}_o}{|\mathbf{v}_o|} = |\mathbf{v}|\cos(\theta-\alpha) \qquad (23)$$

with α the direction of input motion. If we calculate $\widetilde{v}_\theta$ for different directions we see that they are distributed on a circle (figure 9.a). This result is similar to that obtained with two filters, equation (19), but here the denominator has desappear giving a better linearity. We use $2n+1$ filters to estimate velocity components at n orientations.

For the one dimensional pattern (20) we obtain:

$$\widetilde{v}_\theta = v_n / \cos(\theta-\beta) \qquad (24)$$

where β is the apparent direction of motion. This expression is the equation of a line in polar coordinates (figure 9.b).

An intermediate situation can be represented by an input pattern consisting in a plaid resulting from the addition of vertical and horizontal sine waves with different amplitudes: $e(x,y) = sin(2\pi f_o x) + A\, sin(2\pi f_o y)$. In the case where $A = 1$, the input pattern has no orientation and motion is perceived without ambiguity. In this case, the velocity estimations for different directions are distributed on a circle, equation (23). In the case where $A = 0$, only one sine wave is present, the pattern has a one-dimensional structure and motion is ambiguous. Velocity estimations will be distributed on a line, equation (24). In the case where $A = 0.5$, motion can also be perceived without ambiguity. As this pattern is slightly oriented, there is no reason for velocity estimations to be distributed on a circle. In fact, we found that they are distributed on an ellipse. For a moving pattern of two sine waves, the output mean power of a VTF, $\mathbf{v}_o$, is:

$$P_{\mathbf{v}_o} = \frac{1}{(1+4\pi^2\gamma f_o^2)^2 + 4\pi^2\tau^2(v_x - v_o cos(\theta))^2 f_o^2} + \frac{A^2}{(1+4\pi^2\gamma f_o^2)^2 + 4\pi^2\tau^2(v_y - v_o sin(\theta))^2 f_o^2} \qquad (25)$$

where v_x and v_y are the components of the input motion vector. We use equation (22) in order to estimate the component of motion in the direction θ. If we consider that the plaids have a frequency lower than the spatial bandwidth

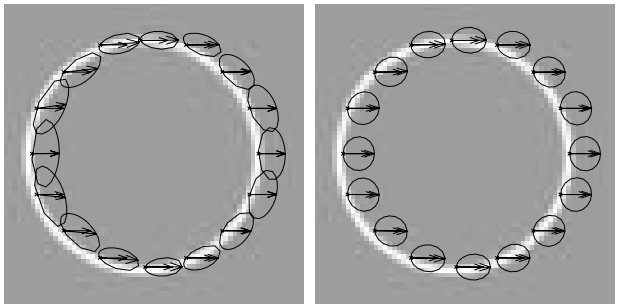


Fig. 10. Polar plots of velocity components estimated at each pixel for a moving circle. Both correct and estimated velocity are superimposed. Local mean power is estimated on a small integration region (right) and a large integration region (left).

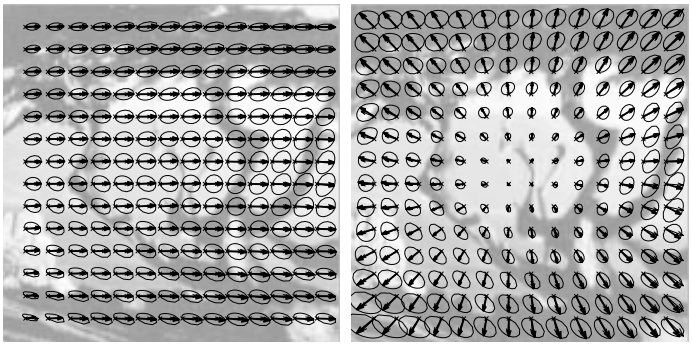


Fig. 11. Left) Translating tree (TT), right) diverging tree (DT).

of the filters ($f_o < \Delta B$), we can approximate equation (22) by a Taylor developpement in $f_o/\Delta B$. We obtain:

$$\widetilde{v}_\theta = \frac{cos(\theta)v_x + A^2 sin(\theta)v_y}{(1-A^2)cos(\theta)^2 + A^2} + O(f_o^2/\Delta B^2) \qquad (26)$$

where $O(.)$ represents the higher order terms that, for low spatial frequencies, can be ignored. The approximation is correct for a loosely tuned VTF ($|\mathbf{v}_o| \simeq \Delta v_o$). Expression (26) is the equation of an ellipse in polar coordinates with center ($v_x/2$, $v_y/2$) passing through the origin (figure 9.c). It must be noted that this expression does not depend on the input spatial frequency f_o. For $A = 1$, we obtain the equation of a circle, and for $A = 0$, this is the equation of a line. As an ellipse passing through the origin is described by four parameters, it will be necessary to estimate motion over at least four directions. This will require 9 VTFs (2 opposed velocities for each direction and one for null velocity). In figure 10, we show the estimated velocity field for a circle moving to the right. At each pixel we show a polar plot with the components estimated in each direction and the velocity vector obtained by estimating the center of the ellipse. Figure 10.a shows the results for a small integration region. We can see that ellipses have their major axis parallel to the contour at each location. In those pixels where velocity is parallel to the contour, the aperture problem is more prominent, giving some errors in the estimation. The eccentricity of the ellipse gives an indication of the significance of the aperture problem. Figure 10.b shows the results for a larger integration area. In this case, the aperture problem has been reduced as the curvature of the circle clearly appears, velocity estimations are distributed on circles.

E. Dealing with motion boundaries

The algorithm presented in this paper supposes that, at least in the integration region and during the time-size of the impulse response of the VTFs, the image has a unique constant translational motion. This means that the input energy lies on a unique plane in the frequency space. However, as at object boundaries, two different motions can co-exist, the power is distributed onto two planes: $f_t + \mathbf{v}_1^T \mathbf{f}_s = 0$ and $f_t + \mathbf{v}_2^T \mathbf{f}_s = 0$; $\mathbf{v}_1$ and $\mathbf{v}_2$ being two velocity vectors. Therefore, the algorithm will fail to produce good estimates.

As proposed by Koch et al. [11] it is possible to open switches in the integration resistive network ("line processes") in order to adapt the power integration window to the object boundaries, avoiding interactions between different objects. Though it is an interesting principle, this solution increases the complexity of the implementation.

In static camera applications, most of motion boundaries are due to the occlusion of a static background by the moving objects. Around the boundaries, power will lie on the planes: $f_t + \mathbf{v}_1^T \mathbf{f}_s = 0$ (object) and $f_t = 0$ (background); $\mathbf{v}_1$ being the velocity of the moving object. A simple solution will consist of adding a temporal derivative in the prefiltering (see figure 6.c) so as to cancel the power plane $f_t = 0$. Mean local power will be due only to the moving object and the algorithm will produce the correct estimate.

As temporal derivative provides a high-pass filtering, it also compensates for the $1/|\mathbf{f}_s|^\alpha$ spectrum decrease of input images (due to coupling between spatial and temporal frequencies in the presence of motion). Therefore, for economic VLSI implementations, the retinal filtering can be avoided. However, the first layer of the retina can improve performance since low-pass filtering reduces sensitivity to noise and aliasing.

V. Results

This section describes performances obtained with the simple architecture described in this paper. They are as accurate as results provided by more complex architectures in the framework of energy-based methods [3], [8], [22].

Table I compares the results obtained with the algorithms of Heeger [8] and Shi et al. [22] with three versions of our algorithm using two artificial sequences, "Translating tree" (TT) and "Diverging tree" (DT) [3], Figure 11. The error at each pixel is measured in degrees using the angular measure given by Barron et al. [3]. This measure combines amplitude and direction of the difference between real and estimated velocity vectors. Table I gives the mean value (m) and the standard deviation (σ) of the error, the number of filters used by each algorithm and the complexity for implementation of each filter with analog circuits (the number of layers corresponds to the number of nodes per pixel and r is the radius of the neighborhood to which

TABLE I

Comparision of performances.

	TT	DT	Filters	Complexity
VTF$^{(i)}$	$m = 8.58°$	9.62°	4	1 layer
	$\sigma = 3.24°$	6.37°		r=1
VTF$^{(ii)}$	$m = 4.71°$	6.47°	9	1 layer
	$\sigma = 1.69°$	2.56°		r=1
VTF$^{(iii)}$	$m = 2.16°$	3.09°	9	1 layer
	$\sigma = 1.26°$	1.68°		r=2
Heeger	$m = 4.52°$	4.49°	36	2 layers
	$\sigma = 2.41°$	3.10°		r=1
Shi	$m = 1.93°$	2.77°	28	2 layers
et al.	$\sigma = 1.52°$	4.55°		r=10

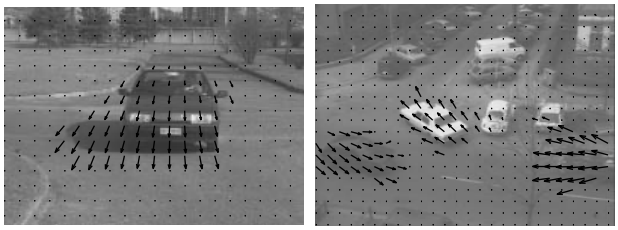


Fig. 12. Estimated optical flows with four VTFs and the shunting inhibition mechanism. Left) car running toward the camera, and right) "Hamburg taxi" sequence. Vector velocities are shown only where motion is detected.

each node is connected).

Heeger [8] uses a large set of narrow band Gabor filters. An efficient approximation to the quadrature pair of Gabor filters may be obtained using analog circuits [20]. This circuit has two layers (two ouput nodes) and connection radius of $r = 1$. The algorithm proposed by Shi et al. [22] uses a set of space-velocity separable filters. This yields a complex implementation with analog circuits (2 independent layers for implementing the quadrature pair and a connection radius of $r = 10$). Furthermore, these two methods require a complex architecture to combine the filter outputs for optical flow computation. The first version, VTF$^{(i)}$, of our algorithm has 4 filters and the shunting inhibition mechanism. The second version, VTF$^{(ii)}$ has 9 filters (4 directions), with three filters for estimating each velocity component. It shows the same performances as Heeger's algorithm. The third version, VTF$^{(iii)}$ has 9 filters, where spatial derivatives of equation (12) are approximated with four points ($r = 2$) and it exhibits performances similar to the algorithm of Shi et al.

All these algorithms are limited to velocities inferior to 3 pixels/frame. This limitation comes from the time discretization required for numerical simulations. However, for an analog circuit with continuous time, such a constraint is relaxed.

When using four filters, VTF$^{(i)}$, the major source of errors occurs in oriented textures regions and in regions with velocities around $|\mathbf{v}_o|$ because of the non linear characteristic of the shunting inhibition. However, estimated optical

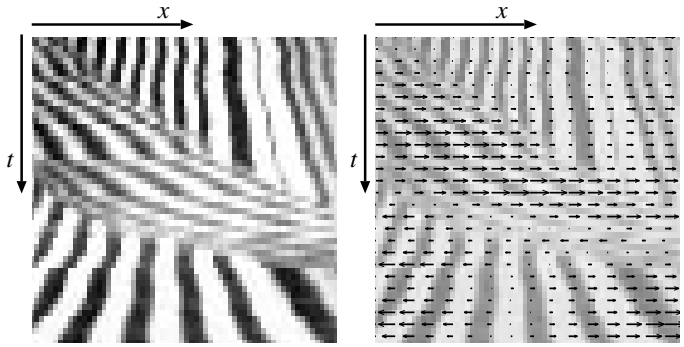


Fig. 13. Simulation of the analog system for optical flow estimation on a real sequence obtained by a linear camera of a mobile robot (each line represents a time step). The algorithm uses two velocity-tuned filters and the shunting mechanism for velocity estimation. Left: input sequence, right: velocity field.

flow allows the detection of moving objects, the estimation of time to contact and the classification of patterns of motion due to self-motion (translation, rotation). Figure 12 shows results obtained with four VTFs using the shunting inhibition mechanism. The first sequence contains a car moving toward the camera on a static background. Estimated velocity vectors are shown only where the output of the temporal derivative is larger than a predefined threshold. The second sequence is the "Hamburg taxi", with three vehicles moving in different directions. For all the sequences, the filter parameters $|\mathbf{v}_o|$ and Δv_o are set to 3 pixels/frame. Figure 13 shows the results obtained by simulation of the circuit of Figure 8 on a real sequence taken by the on-board linear camera of a mobile robot (KHEPERA ©). The robot is moving towards a wall painted with a regular pattern. At the same time, a moving object crosses perpendicularly to the trajectory of the robot, from left to right. The object is correctly detected and, as the robot approaches the wall, the pattern of the divergent optical flow is easily identified.

VI. Discussion

To summarize, the main advantages which make our approach efficient and reliable are as follows:

We use a retinal prefiltering that reduces high frequency noise, "whittens" the $1/|\mathbf{f}_s|^{\alpha}$ spectrum of natural images and enhances the contrast between the responses of different VTFs.

We use a temporal derivative that cancels the power contained in the plane $f_t = 0$. This improves the results because energy integration will not be biased at the motion boundaries between moving objects and the static background. These boundaries can be recovered by detecting the presence of motion directly from the output of the temporal derivative.

We use wideband velocity-tuned filters as motion detectors. They are loosely tuned to different velocities and provide accurate estimations. A shunting inhibition mechanism between the outputs of two filters tuned to opposite velocities allows a very economical means for motion esti-

mation.

The overall structure of the proposed algorithm is well suited for VLSI implementations: based on a neuromorphic approach, it is simple and robust and exhibits sufficient accuracy for applications involving mobile robots (estimation of patterns of optical flow, detection of moving objects, tracking, estimations of time to collision, etc.).

On going work consists on the implementation of the proposed architecture on VLSI, with emphasis on robustness and noise sensitivity. Most of the components can be build with circuits already proposed in the literature [5], [14], [15], [19]. The choice of the velocity $\mathbf{v}_o$ will depend on the application requirements and may be limited by technology. A maximum velocity of $\mathbf{v}_o = 200$ pixels/second will require resistance values around 5G Ohms and capacitor values around 2 pF for the velocity-tuned analog network. These values are compatible with 0.5 μ technologies. Larger values of $\mathbf{v}_o$ require lower resistor and capacitor values.

Acknowledgments

The authors would like to thank B. E. Shi and three anonymous reviewers for helpful discussions and D. Alleysson, G. Sicard, A. Oliva and K. Davies for their comments. Thanks also to P. Bessiere and O. Lebeltel for providing the Khepera sequence. This work has been partly supported by the French Groupement d'Interêt Scientifique "Sciences de la cognition".

References

[1] E. H. Adelson and J. R. Bergen. Spatiotemporal energy models for the perception of motion. *J. Opt. Soc. Am. A*, 2(2):284–299, February 1985.

[2] J. Atick and A. Redlich. What does the retina know about natural scenes? *Neural Computation*, 4:196–210, 1992.

[3] J. L. Barron, D. J. Fleet, and S.S. Beauchemin. Performances of optical flow techniques. *International Journal of Computer Vision*, 12(1):43–77, 1994.

[4] W. Beaudot, P. Palagi, and J. Hérault. *Realistic simulation tool for early visual processing including space, time and colour data.* Lecture notes in computer sciences, 686, "New trends in Neural Computation", Springer Verlag. 1993.

[5] T. Delbrück. Silicon retina with correlation-based, velocity-tuned pixels. *IEEE Trans. Neural Networks*, 4:529–541, 1993.

[6] R. A. Deutschmann and C. Koch. Compact real-time 2-D gradient-based analog VLSI motion sensor. In *Int. conf. on advanced focal plane arrays and electronic cameras*, 1998.

[7] R.A. Deutschmann, C.M. Higgins, and C. Koch. Real-time analog VLSI sensors for 2-D direction of motion. In W. Gerstner, A. Germound, M. Hasler, and J.D. Nicoud, editors, *Proc. International Conference on Artificial Neural Networks (ICANN97)*, volume 1327 of *Lecture Notes in Computer Science*, pages 1163–1168, Lausanne, Switzerland, October 1997. Springer Verlag.

[8] D. J. Heeger. Model for the extraction of image flow. *Journal of Optical Society of America A*, 4(8):1455–1471, August 1987.

[9] C.M. Higgins and C. Koch. Analog CMOS velocity sensors. In *Electronic Imagin'97*, San Jose, CA., February 1997.

[10] J. Hérault. A model of colour processing in the retina of vertebrates: From photoreceptors to colour opposition and colour constancy phenomena. *Neurocomputing*, 12(2-3):113–129, 1996.

[11] C. Koch, J. Luo, and C. Mead. Computing motion using analog binary resistive networks. *Computer*, pages 52–63, March 1988.

[12] J. Kramer. Compact integrated motion sensor with three-pixel interaction. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 18(4):455–460, April 1996.

[13] C. Mead. *Analog VLSI and Neural Systems.* Addison-Wesley, Reading, MA, 1989.

[14] C. A. Mead and M. A. Mahowald. A silicon model of early visual processing. *Neural Networks*, 1:91–97, 1988.

[15] A. Mhani, G. Sicard, and G. Bouvier. Analog vision chip for sensing edges contrasts and motion. *ISCAS*, pages 326–329, 1997.

[16] B. Nabet. Electronic hardware for vision modeling. In R. B. Pinter and B. Nabet, editors, *Nonlinear vision: determination of neural receptive fields, function and networks*, chapter 18, pages 463–474. CRC Press, 1992.

[17] T. Poggio, V. Torre, and C. Koch. Computational vision and regularization theory. *Nature*, 317(26):314–319, September 1985.

[18] R. Sarpeshkar, J. Kramer, G. Indiveri, and C. Koch. Analog VLSI architectures for motion processing: from fundamentals limits to system applications. *Proc. IEEE*, 84(7):969–987, July 1996.

[19] R. Sarpeshkar, R. F. Lyon, and C. Mead. A low-power wide-linear-range transconductance amplifier. *Analog Integrated Circuits and Signal Processing*, 13(1/2), 1997.

[20] B. E. Shi. Gabor-type filtering in space and time with cellular neural networks. *IEEE Trans. on Circuits and Systems-I*, 45(2):121–132, February 1998.

[21] B. E. Shi, T. Roska, and L. O. Chua. Design of linear cellular neural networks for motion sensitive filtering. *IEEE trans. on circuits and systems-II*, 40(5):320–331, May 1993.

[22] B. E. Shi, T. Roska, and L. O. Chua. Hyperacuity in cellular neural networks and the measurement of optical flow. *Int. J. of Circuit Theory and Its Applications*, 26:343–364, 1998.

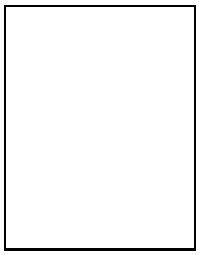
[23] A. Spinei, D. Pellerin, and J. Hérault. Spatiotemporal energy-based method for velocity estimation. *Signal processing*, 65(3):347–362, 1998.

[24] J. Tanner and C. Mead. An integrated analog optical motion sensor. In R.W. Brodersen and H.S. Moscovitz, editors, *VLSI Signal Processing*, volume 2, pages 59–87, New York, 1988. IEEE.

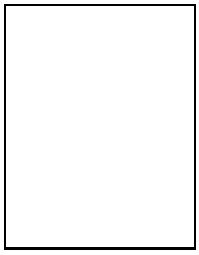
[25] A. Torralba and J. Hérault. From retinal circuits to motion processing: a neuromorphic approach to velocity estimation. In Michel Verleysen, editor, *ESANN'97*, pages 47–54, Brussels, Belgium, April 1997. D facto.

[26] A. B. Torralba and J. Hérault. Minimal complexity velocity-tuned filters with analogue neuromorphic networks: A theoretical approach for efficient design. *Neural Processing Letters*, 8(3):229–139, December 1998.

[27] A. B. Watson and A. J. Ahumada. A look at motion in the frequency domain. In J. K. Tsotsos, editor, *Motion: perception and representation*, pages 1–10, New York, 1983. Association for computing machinery.



Antonio B. Torralba received the B.S. degree in Telecommunications engineering in 1994 from the ETSETB-UPC (Barcelona, Spain) and the M.S. degree in Signal and Image Processing at the Institut National Polytechnique (Grenoble, France). Now, he is part of the LIS laboratory and he is studying for his Ph.D degree. His field of research covers human visual system including retina modeling and optical flow estimation. The models are inspired from Neuromorphic circuits and Cellular Neural Networks.



Prof. Jeanny Hérault received the M.S. degree in electronics engineering in 1966, Docteur-Ingénieur and Docteur es Sciences degrees in 1974 and 1980 from Institut National Polytechnique of Grenoble. He is Professor at Université Joseph Fourier of Grenoble. From 1985 to 1991, he has also been the head of the "Institut des Sciences et Techniques" grouping the four Engineering Schools of this university. His interests include design and hardware implementations of neural machines, with applications in models of Visual perception by means of Signal and non-linear Data Processing. He is expert of the European Comunity and reviewer for several Journals and conferences in Neural Networks and Signal Processing.