

Analog VLSI Focal-Plane Array With Dynamic Connections for the Estimation of Piecewise-Smooth Optical Flow

Alan A. Stocker, *Member, IEEE*

Abstract—An analog very large-scale integrated (aVLSI) sensor is presented that is capable of estimating optical flow while detecting and preserving motion discontinuities. The sensor's architecture is composed of two recurrently connected networks. The units in the first network (the optical-flow network) collectively estimate two-dimensional optical flow, where the strength of their nearest-neighbor coupling determines the degree of motion integration. While the coupling strengths in our previous implementations were globally set and adjusted by the operator, they are now *dynamically and locally* controlled by a second on-chip network (the motion-discontinuity network). The coupling strengths are set such that visual motion integration is inhibited across image locations that are likely to represent motion boundaries. Results of a prototype sensor illustrate the potential of the approach and its functionality under real-world conditions.

Index Terms—Cellular neural networks, dynamic connectivity, gradient descent, line process, motion discontinuities, motion segmentation, neuromorphic, optimization, recurrent feedback.

I. MOTIVATION

THE ESTIMATION of visual motion is considered to be a crucial processing step for systems behaving in a dynamic visual environment. Knowing the instantaneous visual motion facilitates the cognitive understanding of the system's environment. It allows predictions of the relative dynamics of objects within this environment and the planning of appropriate actions accordingly. Typically, visual motion is estimated in a retinotopic coordinate frame and expressed as a dense vector field called the *optical flow* [1]. Such a representation naturally proposes an equivalent, distributed computational architecture where an array of identical computational units processes visual motion information at each image location in parallel. Focal-plane implementations of such arrays are constrained by limitations of the silicon area available. Analog circuit design is favorable because analog circuits allow a far more compact and efficient integration of certain computational primitives than digital circuits [2]. Also important, nonclocked analog circuits match the time-continuous nature of visual motion and therefore are not affected by temporal aliasing.

Manuscript received July 16, 2003; revised January 12, 2004. This work was supported in part by Swiss National Science Foundation, in part by the Körber Foundation, and in part by the Howard Hughes Medical Institute. This paper was recommended by Guest Editor B. Shi.

The author is with the Howard Hughes Medical Institute and Center for Neural Science, New York University, New York, NY 10003 USA (e-mail: alan.stocker@ieee.org).

Digital Object Identifier 10.1109/TCSI.2004.827619

Independent local estimation of optical flow is limited by inherent perceptual ambiguities such as the *aperture problem*. The best possible local estimate of visual motion is so-called *normal flow*, i.e., the optical-flow vector that is normal to the local brightness gradient. Several authors have reported functional analog very large-scale integrated (aVLSI) implementations of two-dimensional (2-D) normal-flow sensors. The circuit architecture of these sensors typically consists of a focal-plane array of independent, isolated processing units [3]–[5], but a serial processing scheme has also been proposed [6]. Other 2-D aVLSI motion sensors have been reported; however, these are not designed to estimate optical flow [7]–[11].

Normal-flow estimation is a perceptually rather poor measure of visual motion (see Fig. 1). It is a sparse representation, only robustly defined at image locations where the brightness gradient is significantly above the noise level. Normal flow is typically insufficient to reconstruct object motion, i.e., the vector average of the normal-flow field is not equivalent to the object motion vector, except for particular object shapes. Accordingly, normal-flow estimation is of marginal relevance to computer vision [12]. It is also important to note that the estimation of normal flow is computationally not very expensive. Implementations using modern, general purpose, digital hardware, and off-the-shelf components can provide real-time solutions for reasonable image sizes. Unless very low power consumption or a focal-plane implementation is required, dedicated aVLSI solutions do not seem advantageous over traditional, digital approaches.

Spatial integration of visual motion information, henceforth termed *motion integration*, can significantly improve the quality of the optical-flow estimate by increasing its robustness, and ultimately provide solutions to the aperture problem (Fig. 1). Motion integration requires an optimization process to find the optical-flow estimate that is in best possible agreement with the visual information observed at all locations within the integration region. Finding the optimal solution is computationally expensive and typically requires iterative numerical processing on sequential digital hardware.

However, we have shown analog network architectures that represent very efficient solutions to these optimization problems. In particular, we reported focal-plane aVLSI implementations of these architectures that estimate smooth optical flow by applying homogeneous, isotropic kernels of spatial integration [13]–[15]. A common characteristic of these networks (see also [16]) is that the optical-flow estimate is the result of a *collective and parallel computational* effort among many, only lo-

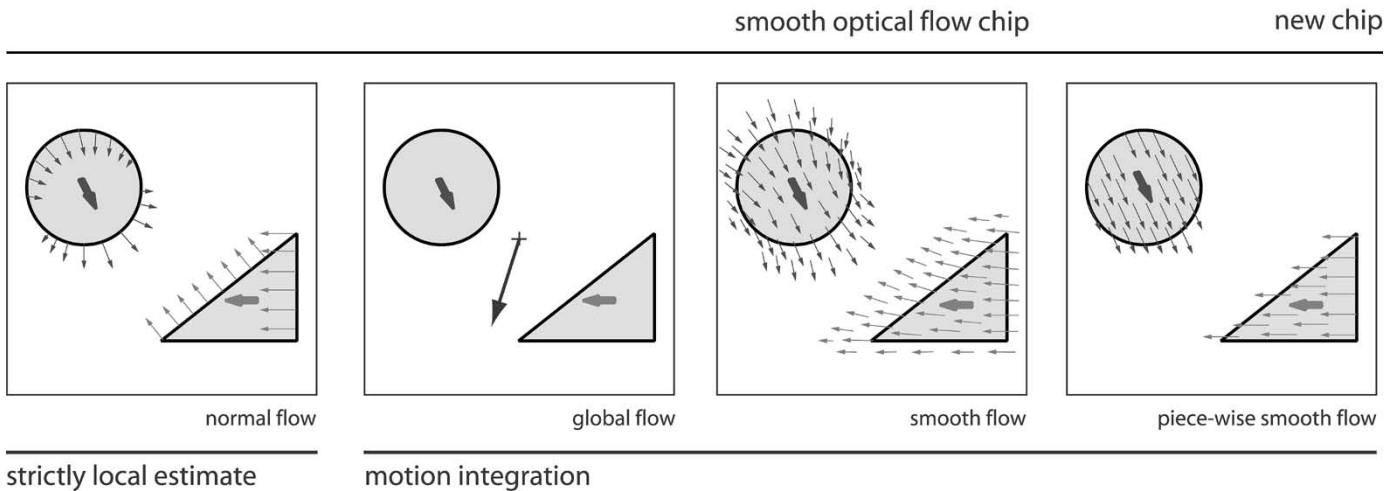


Fig. 1. *Optical-flow estimation with various forms of motion integration.* Normal flow can only be estimated at brightness edges. Global flow represents the optimal global estimate of motion, and thus fails when more than one motion source (object) is present. Smooth optical-flow estimation restricts motion integration to smooth spatial kernels, and thus does not preserve motion discontinuities. Piecewise-smooth flow results from restricting motion integration to the extent of individual motion sources. Our previously reported “smooth optical-flow chip” [13]–[15] provides flow estimates with an adjustable degree of motion integration (from normal flow to global flow); the “new chip” reported here extends this architecture to allow piecewise-smooth flow estimates.

cally coupled units, which complies with the cellular neural networks doctrine [17]. The reported focal-plane arrays robustly work under real-world conditions and have been applied in various tasks such as, e.g., a visual man–machine interface [18]. They represent a substantial improvement and extension of the global motion circuit proposed by Tanner and Mead [19], allowing a continuous adjustment of the spatial extent of motion integration. Smooth optical-flow estimates ranging from normal to global flow are possible (as illustrated in Fig. 1), which is demonstrated with particular clarity by data from a recent 30×30 array implementation [15].

Homogeneous and isotropic nearest neighbor coupling, however, imposes a tradeoff between the extent of motion integration and the preservation of local motion features. The resulting optical-flow estimate is smooth and motion discontinuities are smoothed out. Lei and Chiueh [20] proposed an implementation where motion integration is modulated according to the brightness gradients in the image. This approach is limited because brightness gradients do not usually occur only at motion discontinuities but also, e.g., within object textures. Motion discontinuities, however, usually coincide with object boundaries. Thus, finding ways to limit smoothing to the spatial extents of the individual motion sources (objects) could significantly improve motion integration, resulting in a piecewise-smooth optical-flow estimate (Fig. 1). Given our previous network architecture, such an estimate could be achieved by a mechanism that dynamically couples the appropriate motion units to form isolated ensembles that collectively estimate visual motion only within their region of support. Fig. 2 illustrates such a scheme, where motion integration is restricted to isolated ensembles of motion units A, B, and C, covering individual motion sources.

Several authors have proposed the concept of *line processes* ([21], [22]) to separate different motion sources (for review, see [23]). A line process is defined as a binary variable that determines whether to connect or disconnect local processing

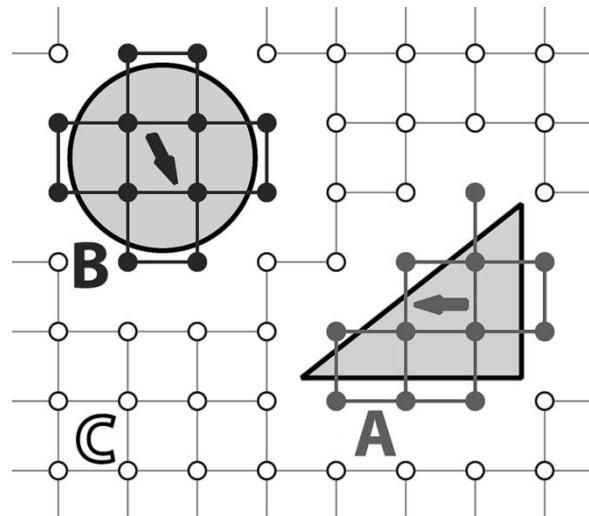


Fig. 2. *Dynamically assigned ensembles of optical-flow units.* Ensembles of processing units are dynamically assigned to individual motion sources. Here, three different motion sources are present and collective computation ideally takes place within the isolated ensembles of processing units A, B (objects), and C (background).

units. Ideally, active line processes represent the outline of the individual motion sources, thus signaling the spatial location of motion discontinuities. Finding the optimal solution is a combinatorial problem and computationally hard. Since the overall cost function of the optimization problem generally exhibits local minima, hysteresis can occur such that once active line processes will remain active even when the visual scene has changed. In traditional computer-vision algorithms that are dominantly frame based (see, e.g., [24] and [25]), this is circumvented by reinitializing the computation between each successive frame. This resetting does not take advantage of the space–time continuous nature of visual motion and is a waste of computational resources.

Of special interest, therefore, are analog, time-continuous network solutions to such nonconvex variational problems as proposed by Koch and colleagues [26]. Based on their framework, Hutchinson *et al.* [27] proposed a hybrid (analog/digital) network that provides near-optimal estimates of optical flow and motion discontinuities. They suggest that the binary line process variable be controlled by distributed digital processors according to the local optical-flow gradient and other assumptions about the typical occurrence of motion discontinuities. Harris and colleagues [28] suggested a similar approach that avoids the need for digital processors by coupling each unit to its neighbors with a two-terminal resistive device called *resistive fuse* [29]. Although the resistive fuse circuit is a very useful, nonlinearity for segmentation tasks according to gradients along one feature dimension and has been successfully applied for one-dimensional motion segmentation [30], it is not straightforward to extend it to gradients of higher dimensional features such as optical flow (speed and direction). None of the network architectures proposed above have been actually implemented in hardware.

In this paper, the author presents a focal-plane aVLSI sensor that simultaneously estimates 2-D optical flow and the appropriate motion discontinuities. It is an extension of our previous network architecture and implementation [14], [15] such that it ideally recognizes different motion sources and then dynamically separates ensembles of motion units accordingly. The design of the sensor takes into account the time-continuous nature of visual motion and the possible effects of hysteresis. The sensor is completely independent from any external reset signal. Part of this work has been described in the author's doctoral thesis [14] and in a conference paper [31].

II. COMPUTATIONAL ARCHITECTURE

The architecture of the proposed system is schematically illustrated in Fig. 3. It consists of two networks, the *optical-flow network* and the *motion-discontinuity network* which are recurrently coupled with each other at each image location. Each network tries to solve an individual task with respect to the state of the other network: “estimate the optical flow” and “find the appropriate motion discontinuities,” respectively. These tasks are defined by a number of constraints and expressed as optimization problems. Computationally, optimization consists of minimizing an appropriate cost function [32]. The dynamics of each network are derived such that they minimize these individual cost functions according to gradient descent. This is similar to network solutions proposed for other computationally hard problems (see the *traveling salesman problem* [33]), although the connectivity pattern here is limited to nearest neighbor connections.

A. Optical-Flow Network

The optical-flow network provides the optical-flow estimate $\mathbf{v}_{ij} = (u_{ij}, v_{ij})$. Each of its units receives visual input given as the spatial $(E_{x_{ij}}, E_{y_{ij}})$ and temporal gradients $(E_{t_{ij}})$ of the

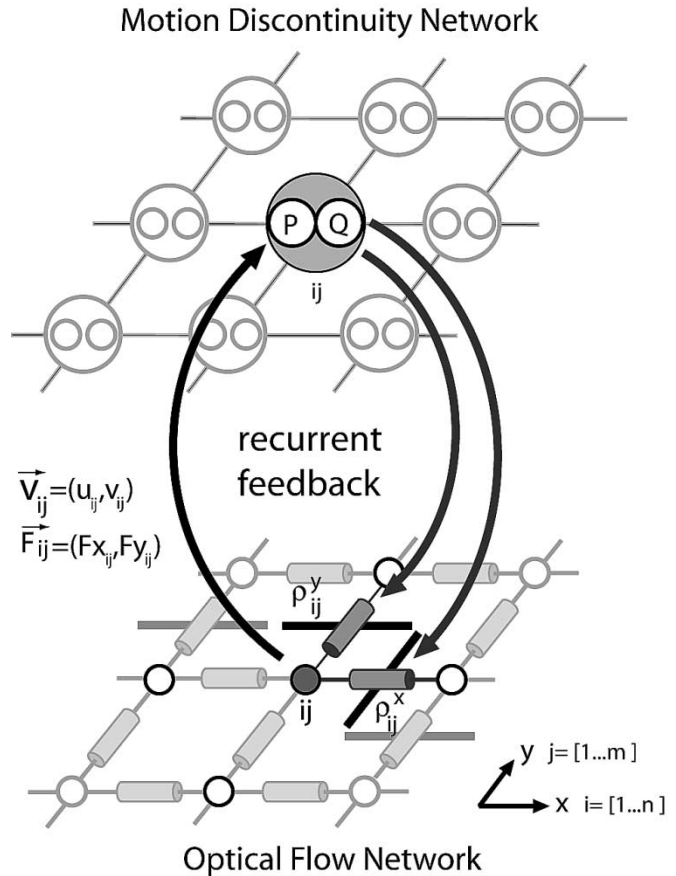


Fig. 3. System architecture. The system is composed of two networks. Each unit of the optical-flow network is coupled to a pair of units (P_{ij}, Q_{ij}) in the motion-discontinuity network that recurrently controls the lateral conductances $(\rho_{ij}^x, \rho_{ij}^y)$ of that particular optical-flow unit. Active line processes are indicated by drawn line segments in the resistive grid. For reasons of readability, only one layer of the optical-flow network is shown.

brightness distribution $E(x, y, t)$ in the image.¹ The desired optical-flow estimate minimizes the cost function

$$\begin{aligned}
 H_{\text{OF}}(\mathbf{v}) = \sum_{ij} \left[\underbrace{(E_{x_{ij}} u_{ij} + E_{y_{ij}} v_{ij} + E_{t_{ij}})^2}_{\text{a}} \right. \\
 + \underbrace{\rho_{ij}^x ((\Delta^x u_{ij})^2 + (\Delta^x v_{ij})^2) + \rho_{ij}^y ((\Delta^y u_{ij})^2 + (\Delta^y v_{ij})^2)}_{\text{b}} \\
 \left. + \underbrace{\sigma ((u_{ij} - u_{\text{ref}})^2 + (v_{ij} - v_{\text{ref}})^2)}_{\text{c}} \right] \quad (1)
 \end{aligned}$$

where Δ^x and Δ^y represent the discrete derivative operator in the x and y direction. This cost function represents the applied model of visual motion perception and is described by three constraints. It requires the optical flow to: (a) to obey the *brightness constancy constraint* [35]; (b) be *spatially smooth*; and (c) be slightly *biased* to some reference motion $\mathbf{v}_{\text{ref}} = (u_{\text{ref}}, v_{\text{ref}})$. The motion model extends the one proposed by

¹In the actual aVLSI implementation, $E(x, y, t)$ is the output of an adaptive photoreceptor array. The adaptation properties and the log compression of the applied photoreceptor circuit [34] improve the robustness of the brightness constancy constraint [14].

Horn and Schunck [36], adding a bias term (c) that improves the model in the following way.

- It increases the robustness of the optical-flow estimate.
- It resolves perceptual ambiguities.²
- It makes the diffusion length finite.

Furthermore, such a bias reflects a typical perceptual prior as observed, e.g., in the human visual motion system [37]. There is a subtle difference to our previous formulations of the cost function [14]: the impact of the smoothness constraint (b), is no longer determined by a global parameter but rather by the local variables ρ_{ij}^x and ρ_{ij}^y . Smoothness is no longer necessarily homogeneous and isotropic, but locally facilitated or inhibited. The cost function (1) is convex for any given input ($E_{x_{ij}}$, $E_{y_{ij}}$ and $E_{t_{ij}}$), any positive distribution ρ_{ij}^x , ρ_{ij}^y , and $\sigma > 0$. Thus, gradient descent on the cost function

$$\dot{u}_{ij} \propto \frac{\partial H_{\text{OF}}}{\partial u_{ij}} \quad \text{and} \quad \dot{v}_{ij} \propto \frac{\partial H_{\text{OF}}}{\partial v_{ij}} \quad (2)$$

is guaranteed to find the optimal solution [14]. We can efficiently implement gradient descent by the dynamics of an appropriate electrical network. Choosing a symmetric bias for slow speeds ($\mathbf{v}_{\text{ref}} = \mathbf{0}$) and introducing a node capacitance C , we can write its dynamics according to (2) as

$$\begin{aligned} \dot{u}_{ij} = & -\frac{1}{C} [E_{x_{ij}} (E_{x_{ij}} u_{ij} + E_{y_{ij}} v_{ij} + E_{t_{ij}}) \\ & - \rho_{ij}^x (u_{i+1,j} - u_{ij}) + \rho_{ij}^y (u_{i,j+1} - u_{ij}) \\ & + \rho_{i-1,j}^x (u_{i-1,j} - u_{ij}) + \rho_{i,j-1}^y (u_{i,j-1} - u_{ij}) \\ & + \sigma u_{ij}] \\ \dot{v}_{ij} = & -\frac{1}{C} [E_{y_{ij}} (E_{x_{ij}} u_{ij} + E_{y_{ij}} v_{ij} + E_{t_{ij}}) \\ & - \rho_{ij}^x (v_{i+1,j} - v_{ij}) + \rho_{ij}^y (v_{i,j+1} - v_{ij}) \\ & + \rho_{i-1,j}^x (v_{i-1,j} - v_{ij}) + \rho_{i,j-1}^y (v_{i,j-1} - v_{ij}) \\ & + \sigma v_{ij}] \end{aligned} \quad (3)$$

which asymptotically converge to the optimal optical-flow solution. The dynamics (3) define a network consisting of two cross-coupled resistive layers with local lateral conductances (ρ_{ij}^x , ρ_{ij}^y) and describe the current equilibrium at each node of the network. At steady state, the voltage distributions in the two resistive layers (u , v) represent the vector components of the optimal optical-flow estimate.

B. Motion-Discontinuity Network

The second network's task is to implement line processes that dynamically set the coupling in the optical-flow network such that motion integration is ideally restricted to individual motion sources. As illustrated in Fig. 3, the states of its units P_{ij} and Q_{ij} represent either the presence or absence of a local motion discontinuity in the x - and y -direction, respectively. Active units represent active line processes that disable the lateral conductances of the optical-flow network accordingly.

²Motion integration alone cannot solve, e.g., the aperture problem in the large. Under this condition, motion perception remains ambiguous, and thus the estimation problem has no unique solution unless an additional constraint is imposed. In the proposed model, this constraint is a perceptual bias that induces a slight preference for a particular motion. See [14], [15] for a detailed discussion.

In order to approximate the binary characteristics of a line process, we require the motion-discontinuity units³ to have a steep sigmoidal activation function

$$g : p_{ij} \rightarrow P_{ij} \in [0, 1] \quad (4)$$

and thus rewrite the coupling conductances in the optical-flow network in the x -direction as

$$\rho_{ij}^x = \rho_0 (1 - P_{ij}) \quad (5)$$

with ρ_0 being the globally set default smoothness conductance.

We now define the optimization problem for the motion-discontinuity network according to the following heuristically derived constraints:

- the total number of active line processes should be limited;
- line processes should preferably be active at locations where local motion information is inconsistent between neighboring locations.

The first constraint represents an activation threshold that is necessary in order not to split the visual scene into too many independent motion sources. The second constraint simply promotes line processes becoming active where they should, namely at motion discontinuities. It requires, however, defining a measure of how inconsistent local motion information is between neighboring locations. Typically, the amplitude of the optical-flow gradient $|\Delta \mathbf{v}|$ is considered to be the basic measure (see, e.g., [27], [28]). This is certainly sensible because motion discontinuities are locations of high optical-flow gradient. However, since reliable detection of motion discontinuities is a pre-requisite for a faithful optical-flow estimate, using the flow gradient to decide where motion discontinuities occur is turning into a circular argument. In practice this typically induces hysteresis, resulting, e.g., in a lag of the trailing edge of a detected motion-discontinuity contour.

To avoid this, we propose a novel measure. It not only considers the optical-flow gradient but also the error signal gradient $|\Delta \mathbf{F}|$ of the brightness constancy constraint. The local error signal is the vector \mathbf{F}_{ij} with components

$$\begin{aligned} Fx_{ij} &= -k E_{x_{ij}} (E_{x_{ij}} u_{ij} + E_{y_{ij}} v_{ij} + E_{t_{ij}}) \\ Fy_{ij} &= -k E_{y_{ij}} (E_{x_{ij}} u_{ij} + E_{y_{ij}} v_{ij} + E_{t_{ij}}) \end{aligned} \quad (6)$$

where k is a constant. \mathbf{F}_{ij} represents the gradient of the brightness constancy constraint (see (1)), and indicates how strongly the optical-flow estimate deviates from the local visual observation. Fig. 4 illustrates the situation at a typical motion discontinuity. The error signal gradient $|\Delta \mathbf{F}|$ is large when the estimated optical-flow field severely deviates from the true velocity. As indicated by the arrows, larger it becomes, the smoother the flow field is and vanishes if the flow field is equivalent to the true velocity profile. However, the optical-flow gradient $|\Delta \mathbf{v}|$ just shows the opposite behavior. A combined measure of the two gradients permits the isolation of motion discontinuities rather independent of the smoothness strength. This is advantageous because we want ρ_0 to be large to lead to strong integration of motion information within individual motion sources,

³For the sake of simplicity we will refer in the following only to one type of units (P); the analysis for the Q -units is equivalent.

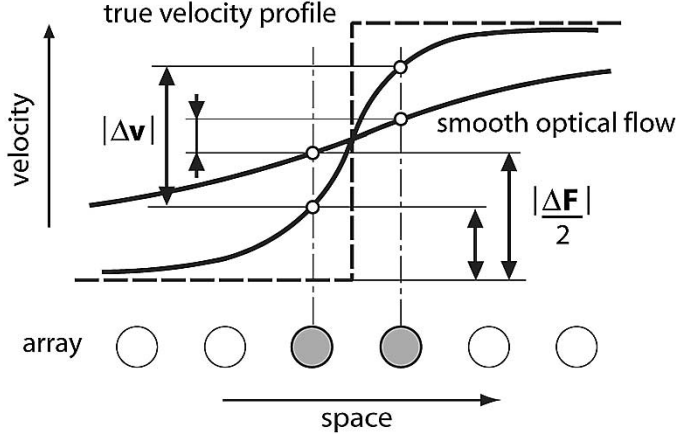


Fig. 4. *Detection of motion discontinuities.* Velocity profiles at a typical motion discontinuity are illustrated for two degrees of smoothness. The true velocity distribution (dashed line) is superimposed by two smooth velocity estimates (bold lines). While the local optical-flow gradient $|\Delta \mathbf{v}|$ decreases with increasing degree of smoothness, the error signal gradient $|\Delta \mathbf{F}|$ increases.

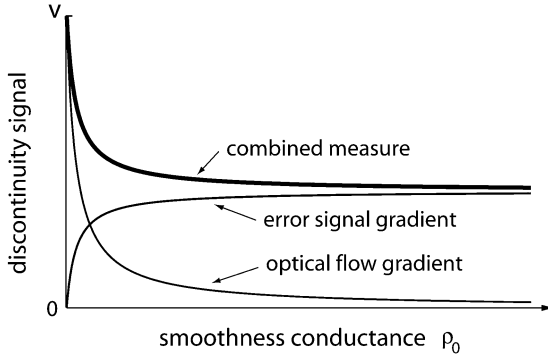


Fig. 5. *Combined gradient measure.* Gradient measures for the motion discontinuity depicted in Fig. 4 are plotted as a function of smoothness. The square amplitude of the optical-flow gradient $(\Delta \mathbf{v})^2$ and the error signal gradient $(\Delta \mathbf{F})^2$ are according to (3) and (6), as well as their sum (bold line). Whereas the amplitudes of the individual gradients can be zero depending on the smoothness strength, the combined measure always reflects the presence of a motion discontinuity.

resulting in a piecewise-smooth optical-flow field. Also, hysteresis is significantly reduced, because the combined gradient measure makes it possible to track motion discontinuities.

Fig. 5 illustrates, in a more quantitative way, the situation depicted in Fig. 4 for the two shaded units lying on each side of the motion discontinuity. The spatial brightness gradients at the two units were assumed to be equally large but the temporal gradient for the left unit was considered to be zero. The squares of the optical-flow gradient $(\Delta \mathbf{v})^2$, the error signal gradient $(\Delta \mathbf{F})^2$, and their sum were then computed according to (3) and (6) and plotted as a function of the lateral smoothness conductance ρ_0 . We clearly recognize a lower bound for the combined measure which is identical to the squares of error signal gradient for $\rho_0 \rightarrow \infty$. Thus, by applying an appropriate threshold, the motion discontinuity can be detected no matter how smooth the flow field is, which is not true if using either one of the gradient measures alone. Note that the combined measure is maximal when the smoothness conductance between the units is zero. This is exactly as it would be if a line process was active.

We formulate appropriate cost functions for each group of units of the motion-discontinuity network such as

$$H_P = \sum_{ij} \left[\alpha P_{ij} + (1 - P_{ij})(\beta \|\Delta^x \mathbf{v}_{ij}\| + \gamma \|\Delta^x \mathbf{F}_{ij}\|) + \frac{1}{R} \int_{0.5}^{P_{ij}} g^{-1}(\xi) d\xi \right]$$

and

$$H_Q = \sum_{ij} \left[\alpha Q_{ij} + (1 - Q_{ij})(\beta \|\Delta^y \mathbf{v}_{ij}\| + \gamma \|\Delta^y \mathbf{F}_{ij}\|) + \frac{1}{R} \int_{0.5}^{Q_{ij}} g^{-1}(\xi) d\xi \right] \quad (7)$$

respectively, where α , β , and γ are weighting parameters and $\|\cdot\|$ is some even-symmetric norm. The integral term represents the total activation energy needed to keep the unit's activation state high or low and is due to the finite slope of the sigmoidal activation function (4).

The cost functions (7) are convex in P_{ij} and Q_{ij} respectively, given a monotonic activation function g . Again, gradient descent is guaranteed to find the optimal solution. Accordingly, we define the network dynamics as

$$\dot{p}_{ij} = -\frac{1}{C} \left(\frac{p_{ij}}{R} + \alpha - (\beta((u_{ij} - u_{i+1,j})^2 + (v_{ij} - v_{i+1,j})^2) + \gamma((Fx_{ij} - Fx_{i+1,j})^2 + (Fy_{ij} - Fy_{i+1,j})^2)) \right)$$

and

$$\dot{q}_{ij} = -\frac{1}{C} \left(\frac{q_{ij}}{R} + \alpha - (\beta((u_{ij} - u_{i,j+1})^2 + (v_{ij} - v_{i,j+1})^2) + \gamma((Fx_{ij} - Fx_{i,j+1})^2 + (Fy_{ij} - Fy_{i,j+1})^2)) \right) \quad (8)$$

where we applied a square-norm for measuring the gradients.

We realize that the discontinuity units do not really form a network in its narrower sense because they are not directly coupled to other units of the motion-discontinuity network. Rather, the units perform a dynamic threshold operation. They are fully analog but forced (R large) to approximate, in steady-state, a binary output behavior: $P_{ij} \approx 1$ if the weighted measure of the flow gradient and the brightness constraint deviation is larger than the threshold α , and $P_{ij} \approx 0$ otherwise. Of course, this represents a rather limited model. One can think of more elaborate models that, for example, also promote the formation of continuous line segments or induce cross-inhibition between P and Q units to prevent the excessive formation of cross boundaries [14], [25]–[27]. However, the derivation of the model was strongly driven by considerations of its aVLSI focal-plane implementation and the adherent space constraints. Since the error signal \mathbf{F}_{ij} is already computed in the optical-flow network, the extraction of $\Delta \mathbf{F}_{ij}$ does not require any extensive extra circuitry.

C. Closing the Feedback Loop

Closing the recurrent feedback loop (Fig. 3), the two networks of the system together solve a typical combinatorial problem,

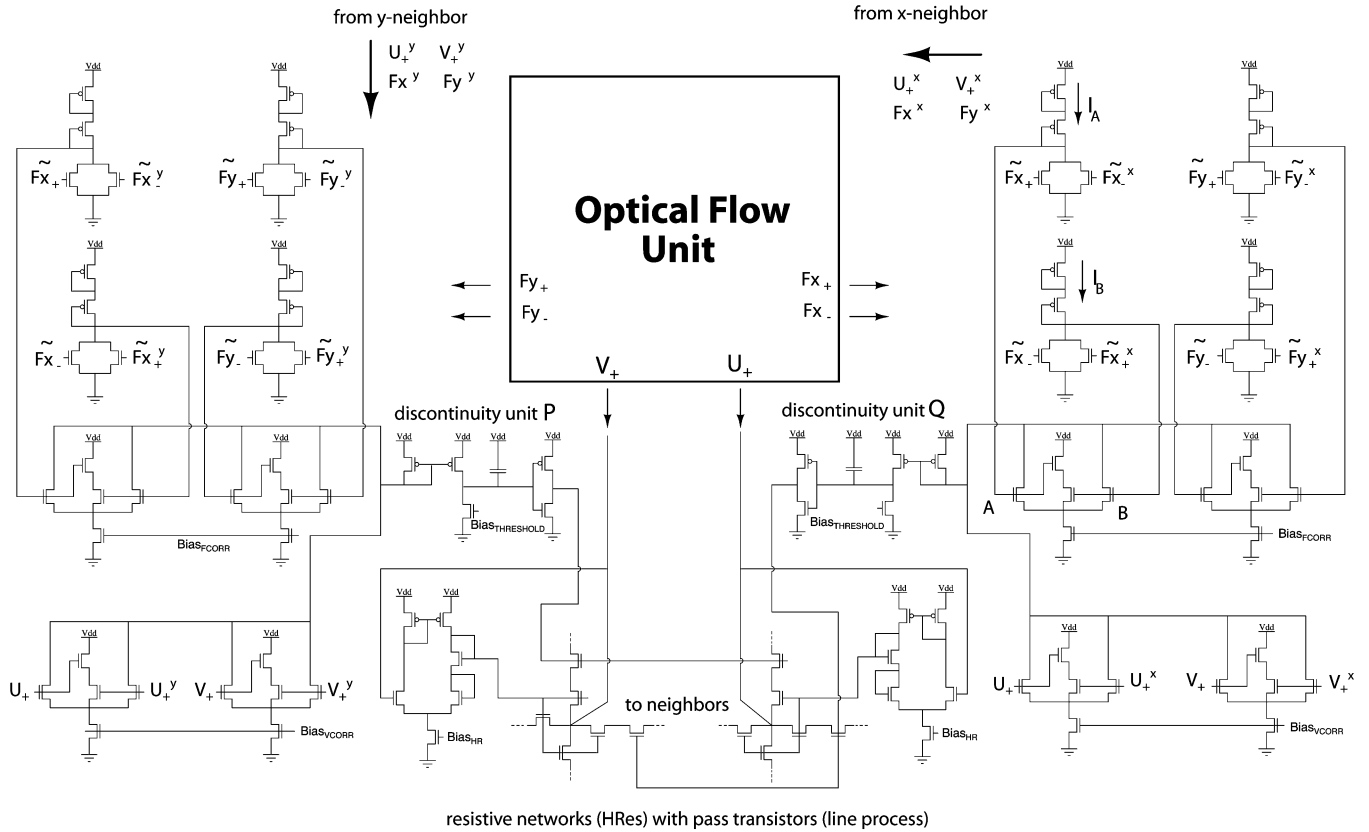


Fig. 6. Schematics of a single pixel.

which is computationally hard. Both the optical-flow network and the motion-discontinuity network are globally asymptotically stable for any given input. For weak coupling between the networks, i.e., if ρ_0 is small, we can presume that the complete system is also asymptotically stable, although not globally. For increased ρ_0 , the complete system can still be presumed to be asymptotically stable if the intrinsic time constants of both networks are sufficiently different.

III. HARDWARE IMPLEMENTATION

A prototype sensor was implemented, consisting of an 11×11 pixel array. Each pixel includes an optical-flow unit plus two discontinuity units. Fig. 6 shows the complete schematics of a single pixel. The circuit schematics of the optical-flow unit are identical to our previous implementations [14], [15] and are therefore not shown here. The optical-flow unit provides the components of the local optical-flow vector encoded as the voltages U_+ and V_+ with respect to a common virtual ground V_0 . It also provides the differential error signals $Fx = Fx_+ - Fx_-$ and $Fy = Fy_+ - Fy_-$ represented as the log-compressed voltages $\tilde{F}x_+$, $\tilde{F}x_-$, $\tilde{F}y_+$, and $\tilde{F}y_-$, respectively. In addition, each pixel receives the optical flow and error signals from the two neighboring pixels to which the coupling strengths are controlled by the pixel's motion-discontinuity units.

Table I summarizes the specifications of the sensor. On-chip scanning circuitry permits read-out of the optical-flow vector, the photoreceptor output, and the state of the motion-discontinuity units from each pixel.

TABLE I
SENSOR SPECIFICATIONS

Technology	AMS 0.8 μm BiCMOS, 2M, 2P
Supply voltage	5 V
Die size	2.7x2.95 mm ²
Array size (motion units)	11 \times 11
Pixel size	170x170 μm^2
Active elements (per pixel)	203
Fill factor*	2.1%
Power dissipation (per pixel)	80 μW
Outputs signals (scanned)	5: photoreceptor, optical flow, P and Q
typical read-out speed	1000 frames/sec

*The fill factor was previously specified as 4% which is not exactly correct [31]. This larger value represents the total area over which visual information is averaged rather than the actual photodiode size. The values are not identical due to the particular shape of the photodiode.

A. Computing the Optical-Flow Gradient

As shown in Fig. 6, the quadratic measure $(\Delta v)^2$ was implemented by summing the *antibump* output currents of two *bump-antibump circuits* [38]. Although each component of the optical-flow vector is encoded as a differential signal, the choice of a common virtual ground V_0 makes it necessary only to compare the positive voltage ground components, serving directly as inputs to the bump-antibump circuits. The bump-antibump circuit is a compact circuit to implement an even-symmetric norm. Its an-

tibump output current can be described as a function of the input voltage difference ΔV

$$I_{\text{antibump}} = I_b \frac{\frac{4}{S} \cosh^2\left(\frac{\kappa \Delta V}{2}\right)}{1 + \frac{4}{S} \cosh^2\left(\frac{\kappa \Delta V}{2}\right)} \quad (9)$$

where S is the ratio of transistor sizes (W/L) between the inner and outer transistors of the circuit and determines the width of the bump, i.e., the range where (9) approximates a squaring function. The circuit design is such that the parameter S , and thus the width of the bump approximately matches the typical linear output range of an optical-flow unit. For larger input ΔV , the output current (9) begins to saturate. Running in saturation, the total output current becomes constant in the difference of each component of the optical-flow gradient, and thus the gradient measure becomes biased toward gradients along the diagonal of its intrinsic motion coordinate frame. The total strength of the optical-flow gradient measure is determined by the bias current I_b of the two bump-antibump circuits, set by the bias voltage $\text{BiasV}_{\text{CORR}}$.

B. Computing the Error-Signal Gradient

As before, bump-antibump circuits are applied to implement the quadratic measure of the error signal gradients. However, unlike before, the appropriate input voltages to the bump-antibump circuits have to be generated first;⁴ we need to derive voltages A and B (see Fig. 6) such that their absolute difference $|\Delta_{AB}|$ is approximately proportional to the absolute difference of two differential currents, thus

$$|\Delta_{AB}| \propto |(Fx_+ - Fx_-) - (Fx_+^x - Fx_-^x)|. \quad (10)$$

The optical-flow unit provides the error signals in log-compressed form as voltages (in units kT/q), e.g.,

$$\tilde{F}x_+ = \frac{1}{\kappa_n} \log\left(\frac{Fx_+}{I_0}\right). \quad (11)$$

Since the difference of two differential signals is equivalent to the difference of their cross-component sums, we can regenerate the currents and sum them appropriately. Using twice diode-connected current adders, the joint differential current I_A can be written as

$$\begin{aligned} I_A &= I_0(\exp(\kappa_n \tilde{F}x_+) + \exp(\kappa_n \tilde{F}x_-^x)) \\ &= I_0 \exp(-\kappa_p A + V_{\text{dd}}/\kappa_p)^{\frac{\kappa_p}{\kappa_p+1}}. \end{aligned} \quad (12)$$

Solving (12) for the voltage A leads to

$$\begin{aligned} A &= -\frac{1}{\kappa_p} \left(\frac{\kappa_p + 1}{\kappa_p} \log(\exp(\kappa_n \tilde{F}x_+) \right. \\ &\quad \left. + \exp(\kappa_n \tilde{F}x_-^x)) - \frac{V_{\text{dd}}}{\kappa_p} \right). \end{aligned} \quad (13)$$

With the equivalent analysis for the current I_B and the voltage B and with the substitution of (11) into (13) we can now express the voltage difference to be

$$\Delta_{AB} = \frac{\kappa_p + 1}{\kappa_p^2} (\log(Fx_- + Fx_+^x) - \log(Fx_+ + Fx_-^x)) \quad (14)$$

⁴We will apply the following analysis for the error signal Fx in x -direction only.

where we have assumed that all κ_n are identical. We can rewrite (14) by expressing the individual components with respect to their common-mode current levels, Fx_0 and Fx_0^x . With $\Delta F = Fx - Fx^x$, we find

$$\begin{aligned} \Delta_{AB} &= -\frac{\kappa_p + 1}{\kappa_p^2} \left(\log\left(Fx_0 + Fx_0^x + \frac{\Delta F}{2}\right) \right. \\ &\quad \left. - \log\left(Fx_0 + Fx_0^x - \frac{\Delta F}{2}\right) \right). \end{aligned} \quad (15)$$

Now, the absolute difference of the logarithms (15) approximates the linear difference operator (10) if: 1) the average common-mode current level $F_{\text{avg}} = (1/2)(Fx_0 + Fx_0^x)$ is constant, and 2) the difference ΔF is small compared to F_{avg} .⁵

The error signals (6) represent the summed output current of two Gilbert multiplier circuits and the hysteretic differentiator. The common-mode current level of each multiplier circuit is half its bias current I_b , and thus constant. The differentiator circuit, however, adds some variable current to the common-mode current level which is half of its rectified output signal. Since the optical-flow network has to be tuned so that the error signals are able to level out, we can assume that the output of the differentiator is always smaller than $2I_b$, the maximal joint output of the two multiplier circuits. The variable amount of the common-mode current level is therefore smaller than I_b . Thus, F_{avg} lies in the interval $[I_b, 2I_b]$. Since the differentiator output is typically of the same order of magnitude as the multiplier outputs, usually $F_{\text{avg}} < 1.5I_b$. Nevertheless, the first condition holds only approximately. The same is true of the second: even when considering the circuit of the optical-flow network as biased so that the multiplier circuits operate in their linear range, the error signal gradient is still in a regime ($\Delta F < 0.75F_{\text{avg}}$) where (15) noticeably deviates from the linear behavior (10).

The suggested circuits only approximate the transformation of the difference of two differential current signals into the difference of two absolute voltages. In practice, however, the large deviations from ideal behavior are not dominant since they mainly occur when the optical-flow network is operating in its saturated regime, which is not recommended. In addition, the saturation characteristics of the bump-antibump circuit partially counter-balance this nonideal effect.

C. Motion-Discontinuity Units

The inputs to the motion-discontinuity units are currents that represent the combined measure of the optical-flow gradients and the error signal gradients. We recognize that the circuits of the motion-discontinuity unit (Fig. 6) implement the dynamics (8). The weighting parameters β and γ for the two gradient measures are controlled by the bias voltages of the bump-antibump circuits $\text{BiasV}_{\text{CORR}}$ and $\text{BiasF}_{\text{CORR}}$, respectively. The threshold current α is set by the voltage $\text{Bias}_{\text{THRESHOLD}}$. The sigmoidal activation function $g : p_{ij} \rightarrow P_{ij}$ is implemented by the inverting output stage, and the leak conductance R is implicitly given as the combined drain conductances of the input and threshold transistor, and is therefore high.

The outputs of the two discontinuity units each control a pair of nFET pass transistors between two neighboring units of the optical-flow network. Active line processes break the lateral

⁵In other words, these are the conditions for which a first order Taylor approximation around F_{avg} holds.

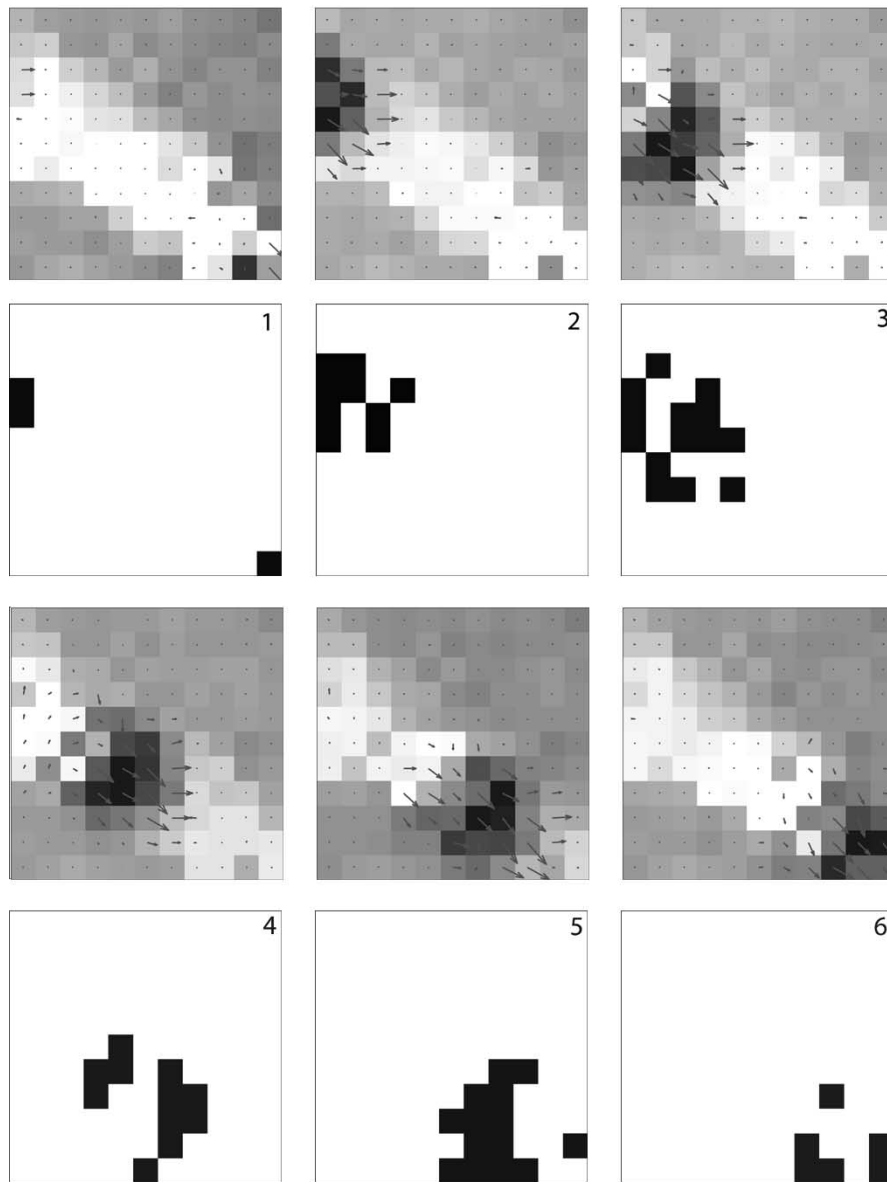


Fig. 7. *Experiment 1: Moving dot.* A dark dot on a light background is repeatedly moving from the left upper to the right lower corner of its visual field. Six frames of the estimated optical-flow field are shown, superimposed on the photoreceptor signal (note the effects of adaptation). The associated activity of the motion-discontinuity units is displayed below each frame.

coupling conductance between the units, while inactive line processes keep them at some default conductance ρ_0 of the horizontal resistor circuits, set by the voltage Bias_{HR} . Note that the horizontal resistor circuit is nonlinear and saturates for high voltage gradients. Thus, even without the motion-discontinuity network, nonisotropic smoothing takes place, reducing motion integration across large gradients in the optical-flow components [14], [20].

IV. EXPERIMENTAL DATA

An aVLSI implementation of the optical-flow network has previously been shown to work robustly for real-world stimuli with a wide range of stimulus contrasts and speeds. For a detailed account of these results, see [14] and [15].

Beyond the characterization of its individual processing units, a quantitative evaluation of the performance of the

proposed sensor is difficult, because an object related, absolute measure for the quality of the estimated optical flow is hard to define and missing so far. In the computer-vision literature, proposed computational algorithms and architectures are typically tested on representative, but nevertheless somewhat arbitrary image frames and sequences. Comparisons among different approaches are usually done on a particular, very limited set of such sequences that serve as quasi-standards. Such standards are not known in the literature on focal-plane visual motion sensors, mainly because the field is rather small.

As a consequence, the sensor's behavior is illustrated by applying two different visual experiments. Due to the small array size, the two experiments include rather simple stimuli. Despite their simplicity, they demonstrate well the advantages of any system that preserves motion discontinuities. The experiments are relatively simple to reproduce and could therefore serve to test and compare other new approaches.

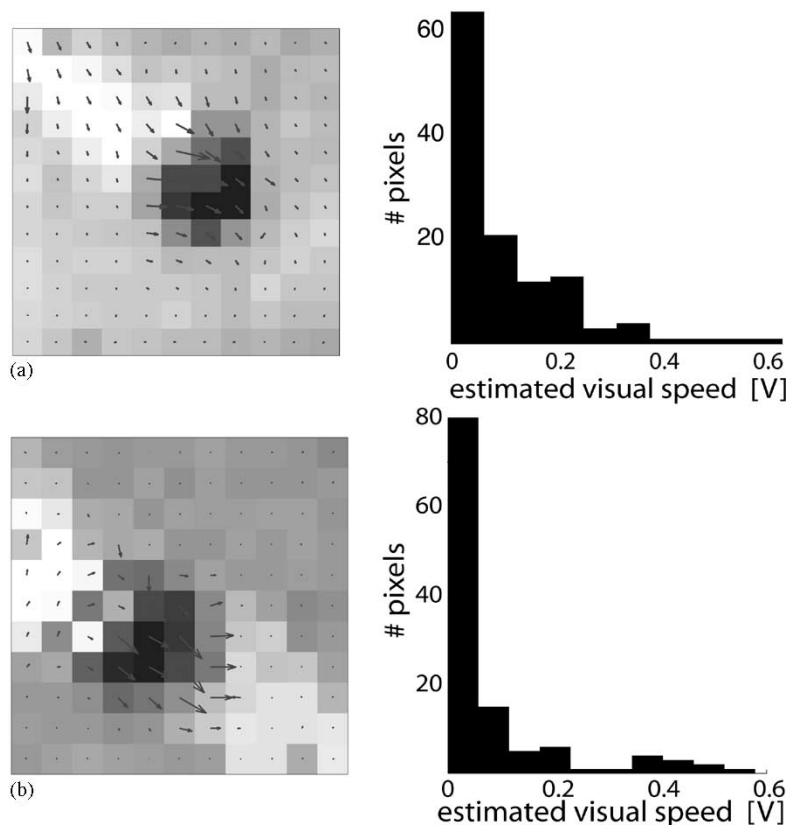


Fig. 8. *Improved optical-flow estimate.* Velocity histogram for the dot stimulus with (a) the motion-discontinuity network being disabled, and (b) being enabled (\rightarrow frame 4 in Fig. 7).

In the first experiment, the chip was presented with a stimulus consisting of a dark dot moving on a light background. The apparent dot diameter was approximately 25% of the sensor's visual field and the dot/background contrast was 72%. Fig. 7 shows six frames of the sampled responses of the sensor while the dot was moving from the upper left to the lower right corner of its visual field. The estimated optical-flow field is shown superimposed on the gray-level images representing the photoreceptor output, while the associated activity of the motion-discontinuity units (P and Q) is displayed as a binary image below each frame (logic OR). We recognize that the activity pattern of the discontinuity units approximately reflects the contour of the dark dot, although the sensor fails to establish a closed contour that completely separates figure and background. This was expected, given that the discontinuity model (7) is limited and does not promote the formation of continuous line-segments that other, more elaborate models do [14], [27].

Nevertheless, the quality of the resulting optical-flow estimate seems to be improved as it predominantly preserves a sharp flow gradient at the dot's outline. This becomes particularly evident when we compare these results with the results obtained when the motion-discontinuity network is disabled. To do so, we kept the default smoothness strength ρ_0 constant but prevented the motion-discontinuity network from becoming active by setting the threshold α to a sufficiently high value ($\text{Bias}_{\text{THRESHOLD}} = 0.8 \text{ V}$).

Fig. 8 shows a comparison of corresponding frames of the resulting optical-flow estimate under the two conditions. Plot-

ting a histogram of the estimated local visual speeds, one would ideally expect a bi-modal distribution since the visual scene contains two motion sources, the moving dot and the stationary background. As can be clearly seen in Fig. 8(a), however, isotropic smoothing does not preserve motion discontinuities, and the velocity distribution is monotonically decaying. This changes when the motion-discontinuity network is activated. Now, the velocity distribution exhibits the expected bi-modal shape, as is shown in Fig. 8(b). Comparison of the two frames also reveals the smooth trail of the optical-flow field when the motion-discontinuity network is disabled. These smooth trails originate from the long time-constant of the optical-flow network when the spatial-temporal energy of the visual input is low (as in the unstructured, stationary background) [14]. Active line processes separate the trail from the active area and thus speed up its decay.

In the second experiment, a stimulus with a less complex motion boundary was applied. As depicted in Fig. 9(a), the stimulus consisted of two tightly joined, identical sine-wave plaid patterns with 72% contrast and a spatial wavelength of approximately half the visual field of the sensor. One plaid pattern was stationary, while the other one moved horizontally to the right, thus forming a linear motion discontinuity. Compared with the first experiment, ρ_0 was increased to achieve an even smoother, almost global motion estimate. Fig. 9(b) shows the response when the motion-discontinuity units were disabled. Here, the increased smoothness conductance ρ_0 completely smooths out the motion discontinuity. The optical-flow gradient becomes

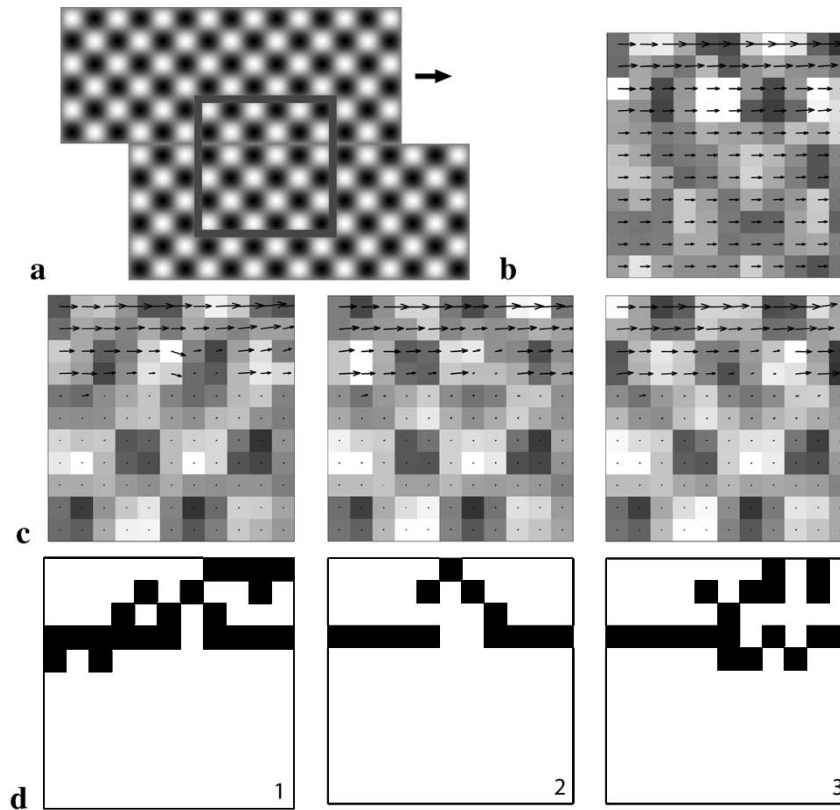


Fig. 9. *Experiment 2: Piecewise-smooth optical-flow estimation.* (a) Applied plaid pattern stimulus provides a linear motion boundary. (b) Optical-flow estimate with disabled discontinuity units. Scanned output sequences of (c) the optical-flow estimate with enabled discontinuity units and (d) their activity.

very shallow and approximately uniform across the whole visual field. Using an optical-flow gradient measure alone would hardly be sufficient to detect the underlying motion discontinuity. The combined measure, however, is sufficient for this task: enabling the discontinuity units leads to a relatively clean separation of the two motion sources, and to the segmentation of the visual scene into two areas of distinct almost piecewise-uniform optical flow [Fig. 9(c) and (d)].

V. DISCUSSION

The aVLSI focal-plane array presented here is the first functional example of a 2-D optical-flow sensor that detects and preserves motion discontinuities. It is an extension of our previous network architectures [13]–[15]. Its computational architecture consists of the optical flow and the motion-discontinuity network and can be considered a cellular neural network with dynamic, self-controlled nearest-neighbor connections. The dynamic reassessment of the coupling pattern significantly improves the optical-flow estimate compared to estimates with constant coupling strengths. The fully analog, time-continuous processing of visual information prevents the impairment of the sensor by temporal aliasing.

A novel measure for the detection of motion discontinuities, and thus the control of the local coupling, has been presented. It combines the optical-flow gradient with the gradient of the brightness constancy constraint error signal, and so permits detection and tracking of motion discontinuities, rather independent of the strength of motion integration. Motion segmentation can be performed, resulting in a piecewise-smooth optical-flow estimate. Nevertheless, the current motion-dis-

continuity network represents a limited model, basically performing a dynamic thresholding operation. There are many ways to improve this model, including line-continuation and cross-orientation inhibition. A soft winner-take-all competition could also locally adapt the effective threshold to reduce the present susceptibility to a fixed threshold value α (see (8)) [14]. However, although such enhanced models could be implemented with only local connections, they would further increase pixel size and circuit complexity. Considering that a single pixel of the current architecture already consists of over 200 active elements and has a fill-factor of only a few percent, any more complicated design might go beyond the feasible limits of focal-plane processing. Of course, advanced process technologies might shift these limits. For example, the application of extra photo-conversion layers [39] could help to overcome the problem of small fill-factors. Nevertheless, there is a limit, and careful considerations should be given to distribute the computational architecture onto multiple chips.

A sensor with the current pixel design but a larger array will improve the apparent quality of its motion estimates as well as its ability to detect motion boundaries. A reasonable resolution would also allow the application of the sensor in areas such as robotics and surveillance. Assuming a more modern $0.35\text{-}\mu\text{m}$ process, the effective pixel size could be reduced by a factor of four, so that a 64×64 array would require a silicon area of approximately $5.5 \times 5.5 \text{ mm}^2$.

ACKNOWLEDGMENT

The research described in this paper was conducted almost exclusively at the Institute of Neuroinformatics (INI), Univer-

sity and ETH Zürich, Switzerland. The author wishes to thank the many members of the hardware group of INI for fruitful discussions and support, particularly R. Douglas and the late J. Kramer. Thanks also to P. Hoyer, T. Saint, and the Associate Editor, B. Shi, for helpful comments on the manuscript.

REFERENCES

- [1] J. J. Gibson, *The Perception of the Visual World*. Boston, MA: Houghton-Mifflin, 1950.
- [2] C. Mead, "Neuromorphic electronic systems," *Proc. IEEE*, vol. 78, pp. 1629–1636, Oct. 1990.
- [3] J. Kramer, R. Sarpeshkar, and C. Koch, "Pulse-based analog VLSI velocity sensors," *IEEE Trans. Circuits Syst. II*, vol. 44, pp. 86–101, Feb. 1997.
- [4] R. Etienne-Cummings, J. Van der Spiegel, and P. Mueller, "A focal plane visual motion measurement sensor," *IEEE Trans. Circuits Syst. I*, vol. 44, pp. 55–66, Jan. 1997.
- [5] H.-C. Jiang and C.-Y. Wu, "A 2-D velocity- and direction-selective sensor with BJT-based silicon retina and temporal zero-crossing detector," *IEEE J. Solid-State Circuits*, vol. 34, pp. 241–247, Feb. 1999.
- [6] S. Mehta and R. Etienne-Cummings, "Normal optical-flow chip," in *Proc. Int. Symp. Circuits and Systems*, vol. 4, Bangkok, Thailand, May 2003, pp. 784–787.
- [7] R. Deutschmann and C. Koch, "Compact real-time 2D gradient-based analog VLSI motion sensor," in *Proc. Int. Conf. Adv. Focal Plane Arrays and Electronic Cameras*, vol. 3410, 1998, pp. 98–108.
- [8] C. Higgins, R. Deutschmann, and C. Koch, "Pulse-based 2D motion sensors," *IEEE Trans. Circuits Syst. II*, vol. 46, pp. 677–687, June 1999.
- [9] R. Harrison and C. Koch, "A robust analog VLSI motion sensor based on the visual system of the fly," *Autonom. Robot.*, vol. 7, pp. 211–224, 1999.
- [10] T. Delbruck, "Silicon retina with correlation-based velocity-tuned pixels," *IEEE Trans. Neural Networks*, vol. 4, pp. 529–541, May 1993.
- [11] C. Higgins and K. Coch, "A modular multi-chip neuromorphic architecture for real-time visual motion processing," *Anal. Integr. Circuits Signal Processing*, vol. 24, pp. 195–211, 2000.
- [12] L. Huang and Y. Aloimonos, "Relative depth from motion using normal flow: An active and purposive solution," in *Proc. IEEE Workshop Visual Motion*, Oct. 1991, pp. 196–203.
- [13] A. Stocker and R. Douglas, "Computation of smooth optical flow in a feedback connected analog network," in *Advances in Neural Information Processing Systems*, M. Kearns, S.olla, and D. Cohn, Eds. Cambridge, MA: MIT Press, 1999, vol. 11, pp. 706–712.
- [14] A. Stocker, "Constraint Optimization Networks for Visual Motion Perception—Analysis and Synthesis," Ph.D. thesis, Dep. Phys., Swiss Federal Inst. Technol., Zürich, Switzerland, Sept. 2001.
- [15] —, (2004, Jan.) Optical-flow estimation as distributed optimization problem—An aVLSI Implementation. Computer Science Dep., New York University, New York, Tech. Rep. TR2004-850. [Online]. Available: <http://www.cs.nyu.edu>
- [16] B. Shi, T. Roska, and L. Chua, "Estimating optical flow with cellular neural networks," *Int. J. Circuit Theory Applicat.*, vol. 26, pp. 343–364, 1998.
- [17] L. Chua and L. Yang, "Cellular neural networks: Theory," *IEEE Trans. Circuits Syst.*, vol. 35, pp. 531–548, Oct. 1988.
- [18] J. Heinzle and A. Stocker, "Classifying patterns of visual motion—A neuromorphic approach," in *Advances in Neural Information Processing Systems*, S. T. S. Becker and K. Obermayer, Eds. Cambridge, MA: MIT Press, 2003, vol. 15, pp. 1123–1130.
- [19] J. Tanner and C. Mead, "An integrated analog optical motion sensor," in *VLSI Signal Processing*, S.-Y. Kung, R. Owen, and G. Nash, Eds. New York: IEEE Press, 1986, vol. 2, p. 59 ff.
- [20] M.-H. Lei and T.-D. Chiueh, "An analog motion field detection chip for image segmentation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, pp. 299–308, May 2002.
- [21] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 6, pp. 721–741, Nov. 1984.

- [22] A. Blake and A. Zisserman, *Visual Reconstruction*, P. H. Winston and M. Brady, Eds. Cambridge, MA: MIT Press, 1987.
- [23] C. Stiller and J. Konrad, "Estimating motion in image sequences," *IEEE Trans. Signal Processing*, vol. 16, pp. 71–91, July 1999.
- [24] D. Murray and B. Buxton, "Scene segmentation from visual motion using global optimization," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 9, pp. 220–228, Mar. 1987.
- [25] M. Chang, A. Tekalp, and M. Sezan, "Simultaneous motion estimation and segmentation," *IEEE Trans. Image Processing*, vol. 9, pp. 1326–1333, Sept. 1997.
- [26] C. Koch, J. Marroquin, and A. Yuille, "Analog neuronal networks in early vision," in *Proc. Nat. Acad. Sci. USA*, vol. 83, 1986, pp. 4263–4267.
- [27] J. Hutchinson, C. Koch, J. Luo, and C. Mead, "Computing motion using analog and binary resistive networks," *Computer*, vol. 21, pp. 52–64, 1988.
- [28] J. Harris, C. Koch, E. Staats, and J. Luo, "Analog hardware for detecting discontinuities in early vision," *Intl. J. Comput. Vis.*, vol. 4, pp. 211–223, 1990.
- [29] J. Harris, C. Koch, and J. Luo, "A two-dimensional analog VLSI circuit for detecting discontinuities in early vision," *Science*, vol. 248, pp. 1209–1211, 1990.
- [30] J. Kramer, R. Sarpeshkar, and C. Koch, "Analog VLSI motion-discontinuity detectors for image segmentation," in *Proc. IEEE Int. Symp. Circuits Systems*, vol. 2, 1996, pp. 620–623.
- [31] A. Stocker, "An improved 2-D optical-flow sensor for motion segmentation," in *Proc. IEEE Int. Symp. Circuits Systems*, vol. 2, 2002, pp. 332–335.
- [32] J. Hopfield, "Neurons with graded response have collective computational properties like those of two-state neurons," *Proc. Nat. Acad. Sci. USA*, vol. 81, pp. 3088–3092, 1984.
- [33] J. Hopfield and D. Tank, "Neural computation of decisions in optimization problems," *Biol. Cybern.*, no. 52, pp. 141–152, 1985.
- [34] T. Delbruck and C. Mead, "Analog VLSI phototransduction by continuous-time, adaptive, logarithmic photoreceptor circuits," Caltech Computation and Neural Systems Program, Pasadena, CA, Tech. Rep. 30, 1994.
- [35] J. Limb and J. Murphy, "Estimating the velocity of moving images in television signals," *Comput. Graph. Image Processing*, vol. 4, pp. 311–327, 1975.
- [36] B. Horn and B. Schunck, "Determining optical flow," *Artif. Intell.*, vol. 17, pp. 185–203, 1981.
- [37] Y. Weiss, E. Simoncelli, and E. Adelson, "Motion illusions as optimal percept," *Nature Neurosci.*, vol. 5, no. 6, pp. 598–604, 2002.
- [38] T. Delbruck, "Bump Circuits," Caltech, Pasadena, CA, Tech. Rep. CNS Memo 26, May 1993.
- [39] S. Manabe, Y. Mastunaga, A. Furukawa, K. Yano, Y. Endo, R. Miyagawa, Y. Iida, Y. Egawa, H. Shibata, H. Nozaki, N. Sakuma, and N. Harada, "A 2-million-pixel CCD image sensor overlaid with an amorphous silicon photoconversion layer," *IEEE Trans. Electron Devices*, vol. 38, pp. 1765–1771, Aug. 1991.



Alan A. Stocker (M'97) received the M.S. degree in material sciences and biomedical engineering and the Ph.D. degree in physics from the Swiss Federal Institute of Technology, Zürich, Switzerland, in 1995 and 2001, respectively.

From 1996–2002, he was with the Institute of Neuroinformatics, Zürich, Switzerland, working on the analysis and the design of analog network architectures for visual motion perception. Since 2003, he has been a Postdoctoral Fellow at the Center for Neural Sciences, New York University, New York. His research interests are in computational neuroscience with a focus on models for primate's visual motion system. He is also interested in the analysis and development of neurally inspired, computationally efficient hardware models of visual perception and their applications in robotics.

Dr. Stocker is a member of the Swiss Society of Neuroscience.