AN IMPROVED 2D OPTICAL FLOW SENSOR FOR MOTION SEGMENTATION

Alan A. Stocker

Institute of Neuroinformatics
University and ETH Zürich
Winterthurerstrasse 190
8057 Zürich, Switzerland

ABSTRACT

A functional focal-plane implementation of a 2D optical flow system is presented that detects an preserves motion discontinuities. The system is composed of two different network layers of analog computational units arranged in a retinotopical order. The units in the first layer (the optical flow network) estimate the local optical flow field in two visual dimensions, where the strength of their nearest-neighbor connections determines the amount of motion integration. Whereas in an earlier implementation [1] the connection strength was set constant in the complete image space, it is now *dynamically and locally* controlled by the second network layer (the motion discontinuities network) that is recurrently connected to the optical flow network. The connection strengths in the optical flow network are modulated such that visual motion integration is ideally only facilitated within image areas that are likely to represent common motion sources. Results of an experimental aVLSI chip illustrate the potential of the approach and its functionality under real-world conditions.

1. MOTIVATION

The knowledge of visual motion is valuable for a cognitive description of the environment which is a requisite for any intelligent behavior. Optical flow is a dense representation of visual motion. Such a representation naturally favors an equivalent computational architecture where an array of identical, retinotopically arranged computational units processes in parallel the optical flow at each image location. Successful aVLSI implementations of such architectures have been reported (see *e.g.*[2]) that demonstrated real-time processing performance in extracting optical flow. Although local visual motion information is sufficient for many applications, its inherent ambiguity (which is *e.g.* expressed as the *aperture problem*) makes the purely local

(normal) optical flow estimate of these processors unreliable and often incorrect.
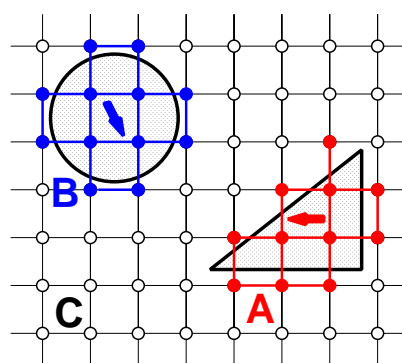
Fig. 1. *Different motion sources and their appropriate regions-of-support.* Three different motion sources are induced by two moving objects and the background. The collective computation is ideally restricted to the isolated sets of processing units A, B (objects) and C (background).

The estimation quality can be increased significantly if visual motion information is spatially integrated. In [3], a motion chip that globally integrates and thus performs a collective estimation of visual motion amongst all the units in the complete image space, is presented. If multiple motion sources[1] are present, however, such a global estimate becomes meaningless. Earlier, we presented an improved focal-plane processor that restricts collective computation to smooth isotropic kernels of variable size, resulting in a smooth optical flow estimate [1]. Ideally, integration should be limited to the extents of the individual motion sources. Such a scheme, as illustrated in Figure 1, provides an optimal optical flow estimate but requires the processing array to be able to connect and separate groups of units *dynamically*. Resistive network architectures applying such dynamical linking have been proposed before [4]. However, to our knowledge there exists only one attempt to implement such an approach [5]. In this one-dimensional processing

---

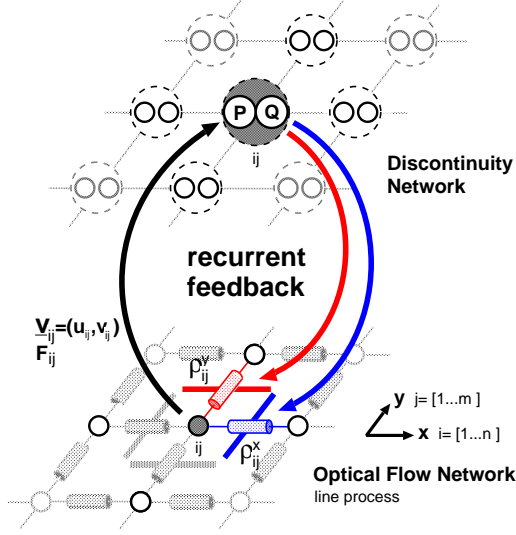[1]*e.g.* a single moving object on a stationary but structured background

Fig. 2. *System architecture.*

array, nearest-neighbor connections in between units were dynamically set according to the local optical flow gradients.

In the following, we present a novel 2D focal-plane processor that provides enhanced optical flow estimates using dynamically controlled connections between its computational units.

2. COMPUTATIONAL ARCHITECTURE

The computational architecture of the processor is schematically illustrated in Figure 2. It consists of the *optical flow network* and the *discontinuity network* that are recurrently connected.

The optical flow network is the physical embodiment of a dynamical system that, in steady state, solves the constraint optimization problem of minimizing the cost function

$$H_{OF} = \sum_{ij} (F_{ij} + \rho_{ij} S_{ij} + \text{bias term}). \qquad (1)$$

This cost function represents the *model* of visual motion the systems applies and is described by three constraints imposed on the expected and allowed optical flow fields: It requires the optical flow at each location (i, j) to obey the *brightness constraint*[2] (F_{ij}), to be *smooth* (S_{ij}) and *biased*. Applying gradient descent on the cost function directly proposes the necessary network dynamics in order to solve the problem. Thereby, the components of the optical flow vectors u_{ij}, v_{ij} are represented by the analog states of the units in the network. The network closely follows previous approaches of solving optimization problems in networks [6],

[2] thus applying a gradient-based optical flow approach

although its units have a linear activation function and connectivity is restricted to nearest neighbors. Further details on the network architecture and an extended analysis can be found in [1, 7].

In the context of this paper, there are two important points to consider: i) the weight of the smoothness constraint in the overall optical flow model is determined by the conductance pattern ρ_{ij} that effectively determines the connectivity strength of the local nearest-neighbor connections; ii) the network is proven to be globally asymptotically stable for any given, positive distribution ρ_{ij}. As a consequence, we can preserve discontinuities in the optical flow estimate if we find a possibility to control the smoothness conductances ρ_{ij} locally at motion boundaries; and still, the network remains asymptotically stable.

The additional discontinuity network provides this possibility where the states of its units P_{ij} and Q_{ij} (see Figure 2) recurrently control the lateral conductance ρ_{ij}^x and ρ_{ij}^y of the optical flow network in the x- and y-direction respectively. For simplicity we refer in the following only to one type of units (P). We assume the discontinuity units to have a very narrow sigmoidal activation function $g : p_{ij} \to P_{ij} \in [0, 1]$, meaning that a motion discontinuity at node ij is, basically, either detected $(P_{ij} = 1)$ or not $(P_{ij} = 0)$. Thus we rewrite the local conductances in the x-direction as

$$\rho_{ij}^x = \rho_0(1 - P_{ij}). \qquad (2)$$

The task, and thus the behavior of the discontinuity network is defined again as optimization problem with the following constraints: i) the total network activity should be low in order not to split the visual scene into too many independent motion sources and ii) activity should be high preferably at locations where the optical flow gradient $(\Delta \boldsymbol{v})$ is large *and* the brightness constraint is severely violated (ΔF). As we will see later, the violation of the brightness constraint serves as an inevitable error measure to achieve a piece-wise smooth optical flow estimate and thus perform motion segmentation. The constraints are combined to form the cost function

$$H_P = \sum_{ij} \Big((1 - P_{ij})[\beta\|(\Delta\boldsymbol{v})_{ij}\| + \gamma\|(\Delta F)_{ij}\|] $$
$$+ \alpha P_{ij} + \tfrac{1}{R}\int_{0.5}^{P_{ij}} g^{-1}(\xi)d\xi \Big) \qquad (3)$$

where α, β and γ are weighting parameters and $\|\cdot\|$ is some symmetric measure. The integral term represents the total activation energy needed to keep the unit's activation state high or low. Gradient descent on (3) then leads to the following dynamics

$$\dot{p}_{ij} = -\frac{1}{C}\Big(\frac{p_{ij}}{R} + \alpha - [\beta\|(\Delta\boldsymbol{v})_{ij}\| + \gamma\|(\Delta F)_{ij}\|]\Big). \qquad (4)$$

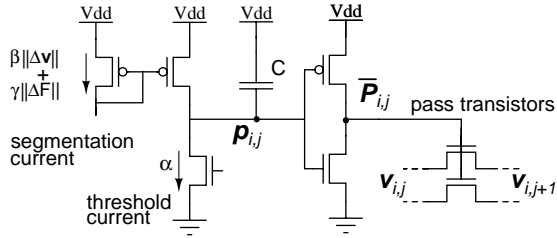We recognize that the units in the network perform a threshold operation, which approximates in steady state a binary

Fig. 3. *Schematics of a discontinuity unit.*

output behavior (due to the boundedness of *g*) if the output conductance R is large: $P_{ij} = 1$, if the weighted measure of the flow gradient and the brightness constraint deviation is larger than a threshold α, and $P_{ij} = 0$ otherwise.

Closing the feedback loop, the two relatively simple network stages of the system solve a typical *combinatorial problem*, which is computationally hard. Unlike in other network solutions of such problems [6], the network architecture is non-homogeneous. The discontinuity network thereby performs a typical *line process* [8], although it remains fully deterministic. Hence, the found solutions might be only sub-optimal which can be reflected by hysteretic behavior in the activity of the discontinuity network.

3. HARDWARE AVLSI IMPLEMENTATION

An 11x11 array of the described optical flow system has been implemented in a double-poly double-metal $0.8\,\mu$m BiCMOS process. Each pixel consists of an optical flow unit plus two discontinuity units. The schematics of the optical flow unit is basically as reported in [1] although improved [7]. The estimated optical flow field is encoded as the continuous voltage distributions U_{ij} and V_{ij} in two resistive layers with respect to some reference potential, where the output signal range is roughly ± 0.5 V. The optical flow units can reliably report the speed of visual motion over almost 3 orders of magnitude.

The schematics of a single discontinuity unit are shown in Figure 3. The circuit approximates the dynamics (4) with the output P_{ij} being inverted. The error measures ($\|\Delta v\|$ and $\|\Delta F\|$) are implemented by bump circuits [9] that provide the local segmentation current accordingly. The output of the discontinuity units controls a pair of pass transistors sitting in between two neighboring units of the optical flow network in order to break the lateral conductances or to leave them at some preset value ρ_0.

In total, a single pixel consists of roughly 200 active elements, occupying a chip area of $(170\,\mu\text{m})^2$. A substantial fraction of this area, however, is used for all the nearest-neighbor connections of the different signals. The fill-factor is at low 4% and power consumption is 80μW/pixel in steady-state.
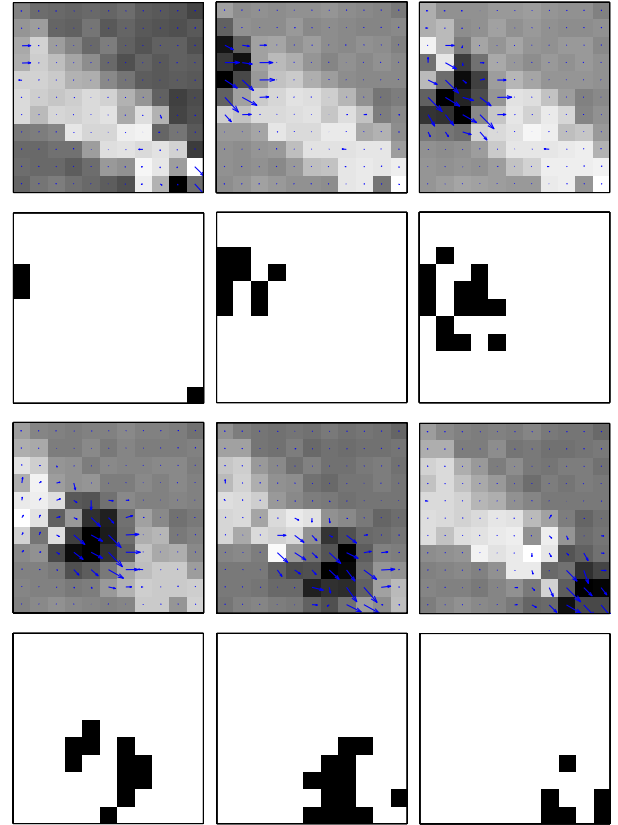

Fig. 4. *Detecting motion discontinuities.* The scanned sequence of the chip's output while seeing a dark dot on a light background, moving from the left upper to the right lower corner of its visual field.

4. RESULTS

We waive a detailed characterization of the optical flow units (which can be found in [7]) and report instead the response of the complete system in two visual experiments, performed under real-world conditions.

In the first example, the chip was presented with a stimulus consisting of a dark moving dot on a light background. Figure 4 shows the sampled responses of the chip 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 onto the images of the photoreceptor output, while the associated activity of the discontinuity units (P and Q) is displayed as binary images below each frame. Note that the activity pattern of the discontinuity units approximately reflects the contour of the dark dot. However, the chip has difficulties to achieve a closed contour that completely separates figure and background. Nevertheless, the optical flow estimate is improved insofar as it predominantly preserves a sharp flow gradient at the dot's outline.
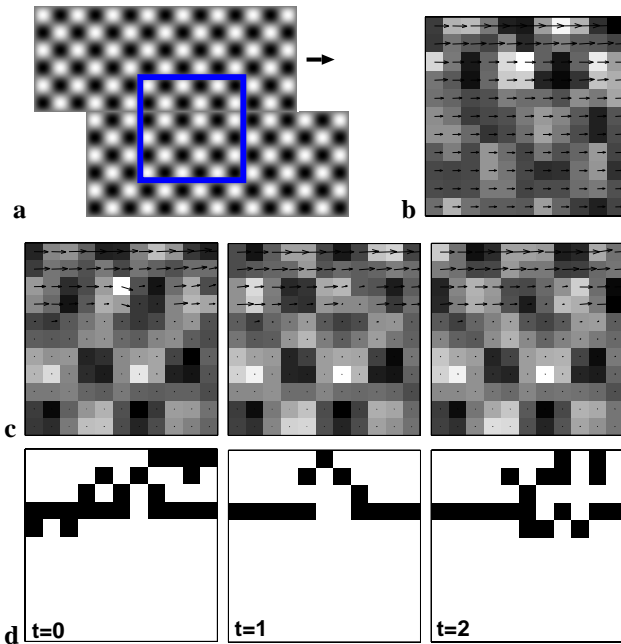
Fig. 5. *Piece-wise smooth optical flow estimation – motion segmentation.* (a) The plaid pattern stimulus providing a linear motion boundary. (b) Optical flow estimate with discontinuity units being disabled. Scanned output sequence of (c) the optical flow and (d) the discontinuity network.

In the second example, a stimulus with a less complex motion boundary was applied (Figure 5a). The stimulus consisted of two tightly joined, identical sinewave plaid patterns. One pattern was stationary while the other one moved horizontally to the right, thus forming a linear motion discontinuity. In a first instance, the motion discontinuity units were disabled. As shown in Figure 5b, the smoothness constraint forces the flow field to be smooth, thus assigning uniform motion to the moving plaid (as desired) but also to the stationary pattern. Enabling the discontinuity units (Figure 5c,d), however, leads to a clear separation of the two motion sources and thus to the segmentation of the visual scene in two areas of distinct, piece-wise smooth optical flow. Note, that this is hardly possible without using the error signal generated by the brightness constraint violations.

5. DISCUSSION

The presented focal-plane implementation is an example of the successful physical translation of collective computational principles in recurrent network architectures. The implementation provides near-optimal solutions to the computationally hard problem of motion segmentation. The chip's dynamical reassessment of its connectivity pattern did prove to enhance optical flow estimation as compared to previous implementations with no [2], or constant interaction strengths [3, 1].

An increased array size of the processor will certainly improve the apparent quality of its motion estimates as well as its ability to detect motion boundaries. Furthermore, local interactions of the motion discontinuity units in a soft-WTA (winner-takes-all) manner can adapt the effective threshold locally to reduce the present susceptibility to a fixed threshold value. Nevertheless, the current implementation is already close to the feasible limits of focal-plane implementation. Any increase in connectivity might require to split up the system into multiple chips.

6. REFERENCES

[1] A. A. Stocker and R. J. Douglas, "Computation of smooth optical flow in a feedback connected analog network," in *Advances in Neural Information Processing Systems 11*, M. S. Kearns, S. A. Solla, and D. A. Cohn, Eds., Cambridge, MA, 1999, pp. 706–712, MIT Press.

[2] R. Etienne-Cummings and P. Mueller, "A focal plane visual motion measurement sensor," *Trans. on Circuits and Systems I*, vol. 44, no. 1, pp. 55–66, January 1997.

[3] J. Tanner and C.A. Mead, "An integrated analog optical motion sensor," in *VLSI Signal Processing, 2*, S.-Y. Kung, R. Owen, and G. Nash, Eds., p. 59 ff. IEEE Press, 1986.

[4] J. Hutchinson, C. Koch, J. Luo, and C. Mead, "Computing motion using analog and binary resistive networks," *Computer*, vol. 21, pp. 52–64, March 1988.

[5] J. Kramer, R. Sarpeshkar, and C. Koch, "Analog VLSI motion discontinuity detectors for image segmentation," in *Intl. Symposium on Circuits and Systems*. IEEE, 1996, pp. 620–623.

[6] J.J. Hopfield and D.W. Tank, "Neural computation of decisions in optimization problems," *Biological Cybernetics*, , no. 52, pp. 141–152, 1985.

[7] A. A. Stocker, *Constraint Optimization Networks for Visual Motion Perception - Analysis and Synthesis*, Ph.d. thesis no. 14360, Swiss Federal Institute of Technology ETHZ, Zürich, Switzerland, September 2001.

[8] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 6, no. 6, pp. 721–741, November 1984.

[9] T. Delbruck, "Bump circuits," Tech. Rep. CNS Memo 26, Caltech, Pasadena, California 91125, May 1993.