ISCAS 2004, Vancouver
Canada, 23-26 May
2004

ANALOG INTEGRATED 2-D OPTICAL FLOW SENSOR WITH PROGRAMMABLE PIXELS

Alan A. Stocker

Center for Neural Science
New York University
4 Washington Place
New York, NY-10003, U.S.A.

Rodney J. Douglas

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

ABSTRACT

We present a framework for real-time visual motion perception consisting of a novel analog VLSI optical flow sensor with reconfigurable pixels, connected in feedback with a controlling processor. The 2-D sensor array is composed of motion processing pixels that can be individually recruited to form dynamic ensembles that collectively compute visual motion. This flexible framework lends itself to the emulation of multi-layer recurrent network architectures for high-level processing of visual motion. In particular, attentional modulation can easily be incorporated in the visual motion processing. We show a simple example of visual tracking that demonstrates the potential of the framework.

1. MOTIVATION

Estimation of visual motion is an important aspect of visual processing in freely behaving and cognitive autonomous systems. However, it is difficult to obtain a reliable estimate because visual motion is inherently ambiguous. The ambiguity can be resolved by estimating the visual motion in terms of a model that relates the observed spatio-temporal brightness changes in a visual scene with the underlying physical world producing it. The better the model, the better is the understanding of the visual scene, and so the better is the estimate of visual motion. The selection of the model is a trade-off between computational cost and functional reliability. Simple models, such as Normal Flow, offer computationally trackablity and real-time performance at the cost of poor estimates under some conditions. Elaborate models, on the other hand, provide more generally acceptable results but are computationally expensive. For example, the estimation of visual motion of an object usually requires the integration of visual information across its outline, an operation that requires *dynamical* recombination of the individual units of a retinotopic array of local motion processors. This approach has been used to model cortical motion processing [1], and also in computer vision (see *e.g.* [2]). Both these cases require highly recurrent, distributed computation whose simulation on standard serial processors is computationally expensive and so prohibits their deployment in real-time applications.

However, experimental evidence indicates that biological brains can perform this task, and so integrated hardware systems are being developed, which are strongly inspired by biological neural processing of visual motion. For example, we have previously described an analog 2-D optical flow sensor that smoothly integrates visual motion information and applies an *a priori* model to resolve ambiguities [3]. And, we have extended this circuit to include a segmentation process that dynamically reconnects the individual units of its processing array to enforce motion integration only within objects [4]. Others have proposed a related approach [5]. These circuits offer dense distributed computation together with real-time performance; but they lack flexibility because they are fully hard-wired. As a step towards overcoming this disadvantage, we present here a hybrid framework that combines the computational efficiency of aVLSI processing arrays with the flexibility of a general purpose, digital micro-processor.

In comparison with a general cellular neural network universal machine CNN-UM [6], our framework has the advantage of being tailored to the requirements of visual motion estimation. In particular, the time-continuous nature of visual motion is preserved by time-continuous processing of visual information, which is not the case in the CNN-UM.

2. FRAMEWORK ARCHITECTURE

The architecture of the framework is based on a novel programmable analog integrated 2-D optical flow sensor (Figure 1). This focal-plane array consists of identical, locally connected processing units that collectively estimate 2-D optical flow. The resulting optical flow estimate $\boldsymbol{v} = (u, v)$

correspondance to: *alan.stocker@nyu.edu*

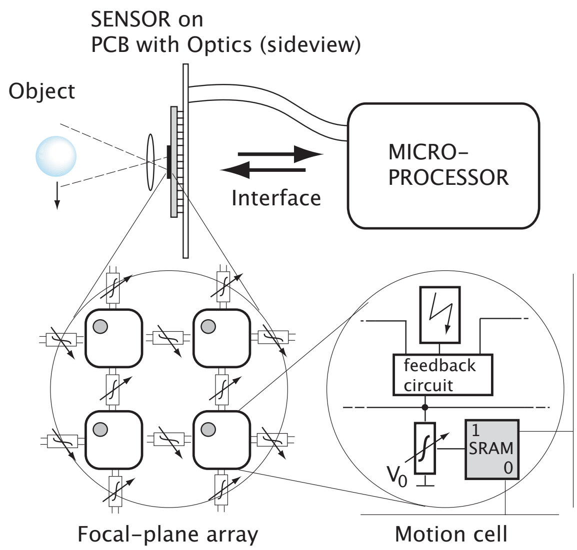

Figure 1: *Schematics of the framework architecture.* The framework consists of the analog 2-D optical flow chip and an external microprocessor. The zoomed-in areas depict the cellular neural network structure of the sensor and one of its units. Each unit has an individually addressable SRAM cell that controls the shunting conductance.

is given as the voltages $U_{i,j}$ and $V_{i,j}$ at each discrete node i, j of the array and represents the *optimal estimate* of the optical flow, given the instantaneous visual input (each unit has an adaptive photoreceptor [7]) and the built-in model of visual motion. The model assumes that visual motion is locally *translational, smooth* and *biased toward some a-priori motion*. The derivation of the model, and its mapping onto two parallel analog resistive networks are explained in detail elsewhere [8]. Note that the conductances of the resistive networks determine the strength of the smoothness, and the bias imposed, respectively. We will consider these conductances as free, local *parameters* that strongly determine the non-isotropic spatial integration of motion information. We have previously presented a completely integrated sensor that dynamically controls its lateral conductances in order to segment motion [4]. The present framework uses isotropic [1] lateral conductances ρ, but permits the individual bias conductances σ to be set externally (see Figure 1). Each unit includes a static 1-bit memory cell (SRAM) whose state $S_{ij} \in [0, 1]$ sets the local bias conductances σ_{ij} to one of two globally defined and controllable values, thus

$$\sigma_{ij} = \begin{cases} \sigma_0 & \text{if } S_{ij} = 0 \\ \sigma_1 & \text{if } S_{ij} = 1; \end{cases} \tag{1}$$

Assuming linear conductances, we can write the dynamics of each node i, j in the two resistive networks as

$$\begin{aligned} \dot{u}_{ij} &= -\frac{1}{C}[E_{x_{ij}}(E_{x_{ij}}u_{ij} + E_{y_{ij}}v_{ij} + E_{t_{ij}}) - \\ &\quad \rho(u_{i+1,j} + u_{i-1,j} + u_{i,j+1} + u_{i,j-1} - 4u_{ij}) + \\ &\quad \sigma_{ij}(u_{ij} - u_{\text{ref}})] \quad \text{and} \\ \dot{v}_{ij} &= -\frac{1}{C}[E_{y_{ij}}(E_{x_{ij}}u_{ij} + E_{y_{ij}}v_{ij} + E_{t_{ij}}) - \\ &\quad \rho(v_{i+1,j} + v_{i-1,j} + v_{i,j+1} + v_{i,j-1} - 4v_{ij}) + \\ &\quad \sigma_{ij}(v_{ij} - v_{\text{ref}})], \end{aligned} \tag{2}$$

where E_x, E_y and E_t are the spatial and temporal gradients of the photoreceptor signals and $(u_{\text{ref}}, v_{\text{ref}})$ is the *a-priori* motion vector [8].

Typically, $\sigma_1 \gg \sigma_0$ and σ_0 is small. Thus, if the memory bit is not set, then the optical flow estimate is only slightly biased toward the reference motion vector $(u_{\text{ref}}, v_{\text{ref}})$, which allows the sensor to resolve ambiguous visual input and increase its robustness at low stimulus contrasts [3, 8]. On the other hand, if the memory bit is set, then the large σ_1 strongly biases the local optical flow estimate and if large enough, basically **shunts** the resistive node to the reference motion vector. Thus, depending on the value σ_1 the response of a local motion processing unit can be reduced or the unit can even be functionally *disabled*.

A key effect of a local increase in the bias conductance, is that the *diffusion length* of the resistive network at this local node is reduced. For a linear resistive network [2] the diffusion length is given as the root ratio $\sqrt{\rho/\sigma}$ [9], which indicates how quickly a voltage signal attenuates along the resistive network. Thus, in the optical flow chip, the diffusion length is a direct measure of the width of the region of motion integration. If σ is very large, information exchange is greatly reduced; in this way a shunted unit provides a barrier for motion integration.

In an array with linear lateral conductances, shunted units would significantly influence the estimates of their neighbors. However, our conductances are highly non-linear and saturating: The *HRes-circuits* [9] have a sigmoidal current-voltage characteristic and saturate for voltage difference larger than $\Delta V_{sat} = 150\text{mV}$. This has the positive effect that the diffusion length decreases sharply for local voltage gradients larger than ΔV_{sat}. Thus, shunted units have little effect on their neighbors when the estimated optical flow is significantly larger than the reference motion. Figure 2 illustrates this property by showing the simulated response of an optical flow sensor (1-D only) for linear and non-linear lateral conductances. Figure 2a shows the estimate for a noisy velocity profile (dashed line) in the case where no units are shunted. For both types of conductance the estimated flow field is smooth. However, the saturating conductances only induce smoothing across small motion gradients, while pre-

[1] in the linear approximation

[2] in the continuous limit.

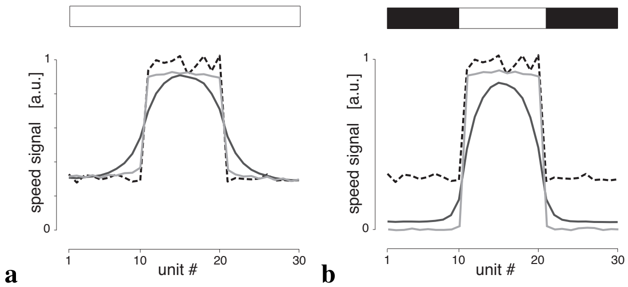

Figure 2: *Non-linear lateral coupling and the shunting influence.* The simulated response for the optical flow network with either non-linear (gray curve) or linear (black curve) lateral conductances. The dashed line represents the noisy true velocity profile. The bar on top of each figure represents the state of the memory cells (black=active). (a) No shunting. (b) The units seeing the slower motion are shunted to the reference motion $v_{\text{ref}} = 0$.

Technology	0.8 μm BiCMOS, 2M, 2P
Supply voltage	5 V
Die size	2.7x2.9 mm^2
Array size (motion units)	15×15
Pixel size	132x132 μm^2
Fill factor	4.1%
read/write cycle	typically < 5 ms

Table 1: Sensor specifications.

serving the motion discontinuities. When the units seeing the lower velocity are shunted (Figure 2b), the response of the active units is not affected in the non-linear case, but significantly reduced in the linear one.

3. IMPLEMENTATION

The present chip layout design follows closely our previous designs [3, 8], except for the SRAM cell added to each unit. The bias conductance is incorporated as the transconductance g_m of a five-transistor transconductance amplifier, with g_m being proportional to the bias current I_b of the amplifier. Thereby, I_b is controlled by two, parallel connected bias transistors. While one bias transistor is always on, the state of the SRAM cell directly enables or disables the second bias transistor. In this way, the local bias conductance σ_{ij} can be programmed to take on one of two freely controllable values.

The chip includes a voltage-scanner for sequential read-out of the optical flow vector and the photoreceptor voltages of each unit. It also contains a 4-bit row- and column address register to individually address and set the state of each memory cell. For flexibility of testing, the optical flow chip was interfaced with a PC via DAQ card. With these arrangements, we could easily achieve several hundred read/write cycles per second.

4. EXPERIMENTS AND RESULTS

To demonstrate the functionality of the framework, we present the results of two experiments. In both experiments, the ratio of the bias conductances for activated versus non-activated shunt was approximately 1000.

4.1. Static Selection

The first experiment shows the effect of statically shunting an arbitrary ensemble of units. A static ensemble of units forming the letter "A" were shunted (disabled) in the center of the array (see Figure 3b, black=high bias conductance). Then, a moving orthogonal sine-wave plaid pattern (photoreceptor signal shown in Figure 3a) was presented to the chip, inducing a uniform optical flow field. Figure 3c shows that the shunting almost completely suppressed the response in the affected units, while the effect of shunting on their nearest neighbors is small due to the non-linear lateral conductances (see Figure 2).

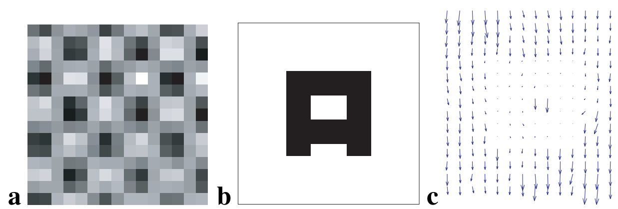

Figure 3: *Shunting the units of the optical flow sensor.* (a) A moving sine-wave plaid pattern induces a uniform optical flow field, but responses of units with an activated (black) shunting conductance (b) are suppressed (c).

4.2. Tracking by selective bottom-up attention

The second experiment now illustrates how our proposed framework can emulate multi-layer recurrent feedback processing. Two moving objects (black dots on white background) are observed by the system. Now, a recurrently connected second layer network is emulated on the PC. A bottom-up attention selective process is implemented by a simplified winner-take-all WTA network, that – once visual speed crosses a threshold – selects the location with largest optical flow vector and shunts all units in the optical flow network accordingly except a given small neighborhood around the winning location. Figure 4 shows four frames of data of the experiment. The first frame shows the starting condition before the recurrent loop was activated.

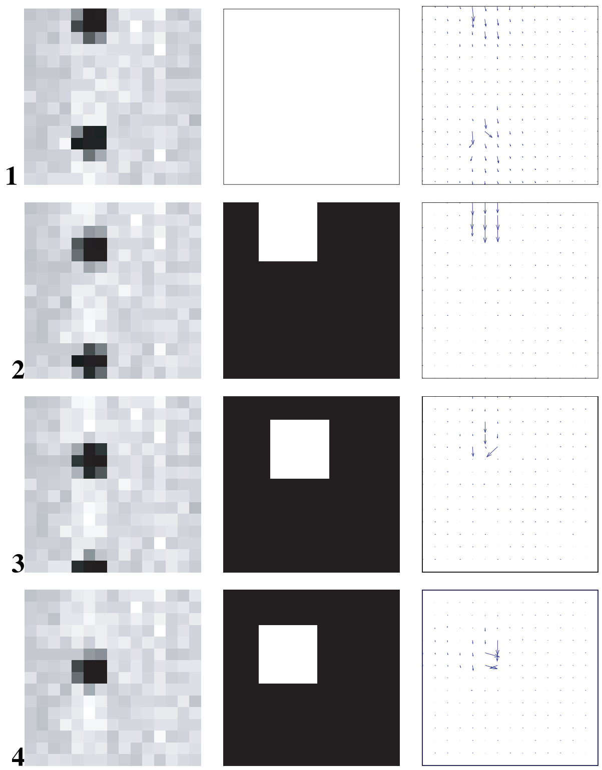

Figure 4: *Tracking through selective attention (bottom-up).* Four frames of a feature-based attention process. Columns: photoreceptor output, shunt states (low=white, high=black), resulting optical flow.

All bias conductances are low and the sensor estimates a smooth optical flow field for both objects. As soon as the recurrent loop is activated, the WTA network selects one object and shunts the units in the optical flow sensor such that the estimation of optical flow is suppressed elsewhere. In this way, the selected object is tracked by dynamically adapting the shunting pattern. Once the object leaves the visual field or stops moving, threshold is not met and shunting is released in the whole array, allowing any new moving object to win the competition.

Note that, because we can freely determine the strength of the shunting conductance, the responses do not have to be suppressed as completely as in this particular experiment. Partial suppression permits an object outside the selected area to win the competition and gain the attentional spotlight if it is moving significantly faster than the currently selected object.

5. CONCLUSIONS

We have presented a framework for emulating recurrent multi-layer network architectures for high-level processing. The present approach extends the potential of our previously reported smooth optical flow sensor [3] by permitting the spatial control of visual motion integration. Either top-down or bottom-up attentional modulation of visual motion processing can be implemented. The hybrid approach combines the computational power of the time-continuous aVLSI optical flow sensor with processing flexibility of a high-level programming language on a standard micro-processor.

Acknowledgments

This work was entirely conducted at the Institute of Neuroinformatics, University and ETH Zürich, Switzerland. The authors thank the Swiss National Foundation, the Körber Foundation and the Howard Hughes Medical Institute for funding.

6. REFERENCES

[1] T.J. Sejnowski and S.J. Nowlan, “A model of visual motion processing in area MT of primates,” in *The Cognitive Neuroscience*, MIT, Ed., pp. 437–449. MIT, 1995.

[2] D.W. Murray and B.F. Buxton, “Scene segmentation from visual motion using global optimization,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 9, no. 2, pp. 220–228, March 1987.

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

[4] A.A. Stocker, “An improved 2-D optical flow sensor for motion segmentation,” in *IEEE Intl. Symposium on Circutis and Systems*, 2002, vol. 2, pp. 332–335.

[5] M.-H. Lei and T.-D. Chiueh, “An analog motion field detection chip for image segmentation,” *IEEE Trans. on circuits and systems for video technology*, vol. 12, no. 5, pp. 299–308, May 2002.

[6] T. Roska, A. Zarandy, Zöld S., P. Földesy, and P. Szolgay, “The computational infrastructure of analogic CNN computing - part 1: The CNN-UM chip prototyping system,” *IEEE Transactions on Circuits and Systems-1: Fundamental Theory and Applications*, vol. 46, no. 1, pp. 261–268, January 1999.

[7] T. Delbruck and C.A. Mead, “Analog VLSI phototransduction by continuous-time, adaptive, logarithmic photoreceptor cirucuits,” Tech. Rep. 30, Caltech Computation and Neural Systems Program, 1994.

[8] 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, March 2002, www.cns.nyu.edu/∼alan/publications/publications.htm.

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