UTILITY AND FEASIBILITY OF A CENTER SURROUND EVENT CAMERA

Tobi Delbruck, Chenghan Li, Rui Graca, Brian Mcreynolds
Inst. of Neuroinformatics, UZH-ETH ZUrich
Zurich, Switzerland. Primary Contact: tobi@ini.uzh.ch.

ABSTRACT

Standard dynamic vision sensor (DVS) event cameras output a stream of spatially-independent log-intensity brightness change events so they cannot suppress spatial redundancy. Nearly all biological retinas use an antagonistic center-surround organization. This paper proposes a practical method of implementing a compact, energy-efficient Center Surround DVS (CSDVS) with a surround smoothing network that uses compact polysilicon resistors for lateral resistance. The paper includes behavioral simulation results for the CSDVS (see sites.google.com/view/csdvs/home). The CSDVS would significantly reduce events caused by low spatial frequencies, but amplify the informative high frequency spatiotemporal events.

Index Terms— retina, pixel, neuromorphic

1. INTRODUCTION

Dynamic Vision Sensor (DVS) event cameras [1] based on the pixel architecture illustrated in Fig. 1 output a stream of pixel-level brightness change events that are proving useful for quick, low power, high dynamic range vision applications [2].

A shortcoming of the DVS is that it responds to all local brightness changes, generating output from any moving image gradient—even low frequency and relatively uninformative gradients—and from brightness changes caused by fluctuating illumination. Many artificial lighting systems (sodium, LED, and fluorescent) flicker at some frequency. These lighting fluctuations can cause a storm of DVS events that are largely uninformative, requiring either increasing the DVS event threshold, decreasing the photoreceptor bandwidth, or both, limiting the ability to transmit informative events about scene reflectance changes.

This problem can be solved by imitating a key feature of mammalian retinas. They have a lateral “surround” network consisting of a 2D mesh of horizontal cells that are connected by conductive gap junction synapses [4]. The surround averages local photoreceptor activity over space and time. This photoreceptor and surround are combined antagonistically with a Center-Surround (CS) arrangement. This way, the surround can cancel the output activity from uniform regions. So far, DVS cameras have not included such CS architecture because the functional advantages were not evident and it was not known how to implement it in a compact and precise form.

This paper proposes a compact and energy-efficient Center Surround Dynamic Vision Sensor (CSDVS) design. The key contribution of this paper is a new compact surround design. Instead of using bulky and imprecise transistors for lateral surround resistors (like past designs), the surround consists of fixed lateral polysilicon resistors combined with a controllable transverse transconductance. The response of the resulting surround will be effectively instantaneous, and its size can be controlled over a wide range. The CSDVS pixel would increase the circuit area by about 20%, but would significantly decrease low spatial frequency output, particularly in response to fluctuating illumination.

Fig. 2 illustrates the CSDVS pixel circuit. It uses a resistive network driven by a transverse conductance \( G \) from the inverted photoreceptor signal \( V_p \) to represent the antagonistic horizontal cell surround signal \( V_h \). The difference \( V_{p+} - V_h \) suppresses events from groups of pixels with similar photoreceptor output. The operating principle and circuit are described in Secs. 3 and 5.

C. Li was with UZH-ETH at time of work; is now with inivation.com. Funded by EU project Visualise (600954), Swiss National Science Foundation project SCIDVS (185069), and the USAF Academy Department of Physics, Instructor Faculty Pipeline.
2. RELATED WORK

Neuromorphic engineers have long dreamed of implementing an electronic model of the biological retina. Fukushima’s discrete electronics model from the 1970s used a resistor mesh to model the horizontal cell network [5]. The earliest integrated silicon retinas of Mahowald and Mead [6–10] featured a horizontal spreading network constructed from transistors using Mead’s Horizontal Resistor (HRES) circuit [7, Ch. 7]. Numerous other imaginative silicon vision sensors featured transistor-based spatial and spatiotemporal filtering at the focal plane [11–17]. These devices had complex pixels (which reduced resolution) and lots of transistor mismatch, which produced excessive salt and pepper Fixed Pattern Noise (FPN) in the output. In-pixel digital calibration circuits were bulky and limited to compensating a single current mirror [16, 17]. At the same time, the computer vision community was being treated to clean megapixel CMOS Image Sensor (CIS) cameras, Moore’s law, and the Internet, so silicon retina development largely stalled out, despite some persistent and beautiful work from the Yagi lab in Osaka [18], [19] and Ruedi at CSEM [20, 21].

Still, CIS cameras had—and still have—nagging problems of limited sample rate, limited dynamic range, redundant output, motion blur, and power-latency trade off [22]. These problems kept interest alive. Since the mid-2000s, DVSs and subsequent event cameras using the active logarithmic photoreceptor and switched capacitor change detector from [1] reduced FPN and improved overall performance, opening up the field for event camera applications [2]. Some event cameras even include activity-driven [23] or sampled [24] intensity values. Resolution and readout bandwidth have increased dramatically, which is great for applications like self-driving cars that must see small objects far away [25].

However, biological Retinal Ganglion Cell (RGC) spikes are the result of vastly sophisticated computations relevant for survival, and it seems with the drive for more and smaller pixels, event camera design is diverging from biology. With this rich historical background—and with industry occupied with the megapixel DVS race [25–27]—it is a good opportunity to revisit the idea of implementing at least a simple antagonistic center surround. The main questions we answer are:

1. Would a center-surround event camera be useful?
2. Is it possible to design a compact and precise pixel?

Preliminary work was reported in Li’s PhD thesis [3, Chapter 5.1]. This paper extends that report by a better understanding of pixel dynamics and behavioral simulation.

3. FUNCTIONAL PRINCIPLE

The CS computation (illustrated by the inset in Fig. 2) takes the difference between the photoreceptor output signal $V_{p+}$ and the average $V_h$ of the surrounding neighbors over a space constant $L$, where $L$ is the distance from the origin of an input to the node where the surround signal has decayed to $1/e$ of the signal at the origin node. This CS computation is a spatial highpass filter; spatial frequencies lower than $\sim 1/L$ pixels are filtered out.

Past CS silicon retinas have mainly output the static difference between the $V_{p+}$ center and the $V_h$ surround, modeling sustained RGCs [4]. The CSDVS pixel models a transient retinal pathway such as is found in most peripheral RGCs: its output consists of asynchronous changes in $V_{p+} - V_h$ that exceed the ON and OFF threshold $\theta$:

$$|\Delta(V_{p+} - V_h)| > \theta.$$  

After each event, the value of $V_{p+} - V_h$ is memorized by the change detector.

4. FUNCTIONAL UTILITY

To model CSDVS, we branched our video to events camera simulator v2e [28] to include an optional antagonistic surround network. A user can specify the space constant $L$ and
Fig. 3. Comparison of simulated normal DVS and CSDVS response to a flashing spot.

Fig. 4. Comparison of normal DVS and CSDVS simulated responses with highlighted differences. A: Flashing spot. B: Moving gradient. C: Panned outdoor cloudy scene. D: Flashing illumination.

5. PIXEL DESIGN

Mead [7, Chapter 7] showed that a one-dimensional discrete resistive network has a space constant length $L$ that has the relation

$$L = \frac{1}{\sqrt{RG}}. \quad (2)$$

For a 2D resistive mesh, (2) still describes the response to an edge [7, Ch. 7, App. C]. Feinstein’s analysis showed for a 2D mesh that (2) still approximately holds for $L \gg 1$ [29].

5.1. Time domain

Fig. 2 outlines a potential CSDVS circuit. Each horizontal cell surround node $V_h$ is driven by the inverted photoreceptor output $V_p$ through transconductor $G$. It has dynamics determined by (3):

$$C \frac{dV_h}{dt} = G(V_{p+} - V_h) - \frac{1}{R} \sum_{j=NSEW} (V_h - V_j), \quad (3)$$

Sec 5 shows that $\tau$ should be much smaller, but it makes the Euler stepping of (3) extremely slow.

See sites.google.com/view/csdvs/home

Sec 3 shows that $\tau$ should be much smaller, but it makes the Euler stepping of (3) extremely slow.

Sec 3 shows that $\tau$ should be much smaller, but it makes the Euler stepping of (3) extremely slow.
where $C$ is the capacitance, and NSEW means the 4 nearest neighboring $V_b$ nodes.

### 5.2. Design of surround

Previous CS retina designs used some type of CMOS transistor surround, which allowed control of surround space and time constant independently. This choice required many transistors and the transistor mismatch caused a lot of FPN. But shows that the larger the space constant $L$, the smaller the transconductance $G$ needs to be, which means it might be possible to use normal (fixed) resistors for $R$ and to implement $G$ with transistors. That would still allow control of $L$ but the resistors would be more compact and better matched. Therefore, we reconsider using polysilicon resistors for the lateral resistor $R$, rather than the bulky HRES circuit. The computations in this section are order-of-magnitude since we do not yet have detailed circuit implementations.

Unsalicided polysilicon is a standard resistor offered in many processes. For example, in a 180nm process we have used for DVS chips, the unsalicided polysilicon has a sheet resistance of $2\,\Omega/\square$ and a $3\sigma$ mismatch of 5% for a 2um wide 100um long resistor. If we use 1um wide and 5um long unsalicided polysilicon as one resistor (to fit along one edge of a $5 \times 5um^2$ pixel), its resistance is 10k$\Omega$, with a likely $3\sigma$ matching to 10%.

Now means that $G \sim 1/(RL^2)$. The transconductor $G$ could be implemented with a 5-transistor transconductance amplifier or even a 2-transistor source follower, since the DC level is ignored by the DVS change detection. Since subthreshold transconductance $G$ is related to bias current $I_G$ by $G \sim I_G/U_T$, then

$$I_G \sim \frac{U_T}{RL^2} \quad \text{and} \quad L \sim \sqrt{\frac{U_T}{R I_G}},$$

(4)

where $U_T$ is the thermal voltage (25 mV at room temperature). Thus, for $L = 10$ px, with $R = 10$ k$\Omega$, $I_G \sim 10$ nA, which is a small current. For a megapixel array, the total $G$ bias current would only be 10mA, a fraction of the total supply current.

The surround temporally low-pass filters the photoreceptor through $\tau = C/G$. However, because the lateral resistance $R$ is only a few k$\Omega$, $G \sim 1/(RL^2)$ is still large. $C$ mainly consists of the summing amplifier input capacitance $C_{h}$ (Fig. 5B). Overestimating $C = 1pF$ and $R = 100k\Omega$ and taking $L = 10$, then

$$\tau \sim C/G = RCL^2 = 10^5\Omega \times 1^{-12}F \times 10^2 = 10us,$$

(5)

corresponding to a cutoff frequency $f_{\text{cutoff}} = 1/(2\pi\tau) \approx 10kHz$, which is about 10X faster than the maximum photoreceptor bandwidth. The surround would effectively respond instantaneously to the photoreceptor input, allowing the cancellation of redundant events caused by flickering illumination.

To detect changes in the difference $V_{+} - V_b$, according to A, the surround must be subtracted from the photoreceptor. Fig. 5A shows a photoreceptor circuit that produces opposing output voltages $V_{+}$ and $V_{-}$; the $V_{-}$ drives the input to the surround. Fig. 5B shows a switched capacitor circuit that sums the positive photoreceptor and negative surround voltages to detect significant events. The relative sizes of $C_{+}$ and $C_{-}$ are adjusted by circuit simulation to compensate for $V_{+}$ and $V_{-}$ gain differences. The large $L$ reduces the effect of expected large $G$ mismatch.

### 6. CONCLUSIONS

The proposed CSDVS design would provide a surround with a controllable size. The surround would effectively be instantaneous. CSDVS would amplify high spatial frequencies and significantly reduce DVS activity in uniform and smoothly varying areas of the scene. It would have a dramatic effect on reducing activity in uniform areas of the scene caused by time-varying lighting.

The proposed surround would add 4-8 transistors (depending on the transconductor design) and 2 narrow stripes of polysilicon resistor. The CSDVS pixel would use about 10 fewer large analog transistors than a design using the HRES. Hence, the CSDVS design would be feasible with a modest increase in pixel complexity. Combined with the switched capacitor DVS change detection, we expect much less FPN than in past CS silicon retinas.

Although CSDVS (like DVS) remains a gross simplification of the vastly more sophisticated biological retina, it may form the next useful abstraction in the delicate dance between introducing features and increasing pixel complexity. It may benefit to dynamically aggregate photoreceptors for the center and to dynamically modulate both center and surround radii as is seen in biology [4], and to exploit $V_b$ feedback in some manner like [10, 15].

Acknowledgments – We thank SC Liu for useful comments.

---

4Salicide is a highly conductive alloy of metal and silicon that normally decreases the polysilicon gate conductor resistance. If the metalization is blocked, the conductance is controlled by the polysilicon doping.

---

![Fig. 5: A: CSDVS photoreceptor circuit that produces opposing output voltages $V_{+}$ and $V_{-}$. B: Summing switched capacitor change detector amplifier. Adapted from [3].](image-url)
