A 240x180 10mW 12us Latency Sparse-Output Vision Sensor for Mobile Applications

Raphael Berner, Christian Brandli, Minhao Yang, Shih-Chii Liu, and Tobi Delbruck
University of Zurich and ETH Zurich, Zurich, Switzerland
Inst. of Neuroinformatics, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland
raphael@ini.phys.ethz.ch, +41 44 635 30 38

Abstract

This paper proposes a 0.18um CMOS vision sensor that combines event-driven asynchronous readout of temporal contrast with synchronous frame-based active pixel sensor readout of intensity. The sensor is suitable for mobile applications because it allows low latency at low data rate and therefore, low system-level power consumption. The image frames can be used for scene analysis and the temporal contrast events with 12us latency can be used to track fast moving objects.

(Keywords: CMOS, vision sensor, AER, DVS, APS, neuromorphic, machine vision, wireless sensor networks)

Introduction

The use of conventional frame-based imagers for machine vision or in sensor networks has the drawback that they produce redundant data which consumes processing power and output bandwidth. The dynamic vision sensor (DVS) overcomes this redundancy by detecting and outputting temporal contrast events i.e. relative illumination changes [1]. Each pixel asynchronously sends out an ON event if the log-compressed light intensity at the pixel increases by a fixed amount and an OFF event when it decreases. This redundancy reduction leads to low latency and low power consumption, but the lack of access to the static scene content restrict its application in scene analysis tasks. The asynchronous time-based image sensor (ATIS, [3]) overcomes this limitation by adding an event-triggered PWM intensity readout to each DVS pixel. This PWM readout provides high dynamic range, but it uses a large on-pixel area and an additional photodiode; and it triples the amount of data to be transferred, since every DVS event is followed by two PWM events encoding the intensity.

The proposed "apsDVS" sensor combines the advantages of DVSs and active pixel sensors (APSs) at the pixel level. It outputs image frames through the synchronous APS pathway and simultaneously outputs low latency temporal contrast events through the asynchronous DVS. The intensity information is recorded through an APS circuit added to the DVS pixel. The shared photo-diode (PD) and small size of the APS circuit lead to an apsDVS pixel area that is 60% smaller than the ATIS pixel area in the same process.

Sensor design

The pixel circuit (Fig. 1) allows simultaneous operation of the temporal contrast detection and the frame-based intensity readout without interfering with each other. Because the DVS photoreceptor (MN5/MN6/MP1) holds PD at a virtual ground, the photocurrent is available at the drain of MN5, allowing the APS intensity readout with transistors MN1 to MN3. The cascode transistor MN4 protects the drain of MN5 from voltage transients due to the reset of Vaps. To prevent saturated pixels from affecting the DVS pathway, during integration column reset CR is set to a voltage that keeps the cascode transistor (MN4) in saturation.

The DVS part of the pixel is similar to [1]: the photo-current is logarithmically encoded by the photoreceptor output voltage Vpr. The self-timed switched-cap amplifier amplifies the deviation from the last reset-level. This deviation is compared against two thresholds. As soon as one of the thresholds is crossed, an event is communicated to the periphery and the switched-cap amplifier is reset to store the new illumination level. To communicate the DVS events, the sensor uses word-serial burst mode address event representation (AER) communication circuits evolved from [4]. The frame readout uses a single-tap rolling-shutter column parallel analog output. Correlated double sampling is performed by off-chip digital subtraction of signal and reset value. Increased dynamic range is enabled by multiple readouts during a single integration cycle.

Experimental Results

Fig. 2(a) shows the combined output of both readouts when the sensor observes a person catching a football. The output illustrates the sparseness of the DVS output (colored) – the events are mainly produced by the moving ball while most pixels stay silent. Fig. 2(b) demonstrates the temporal accuracy and resolution of the DVS events: A 5ms slice of DVS events (colored) acquired 75ms after the APS frame highlights the contours of the ball. Fig. 3 shows a space-time view of the DVS data generated by a 100Hz spinning rectangle. Because of the asynchronous nature of the DVS output, the sensor can capture high-speed motion exceeding 40k pixels/s across the sensor.

Table 1 compares the apsDVS to other sensors. Although APS dark current is high owing to the intervening transistors, the DVS circuit accesses the nwell photodiode directly, resulting in a 120dB dynamic range for the DVS pathway extending down to 0.1lux operation for high contrast inputs. The equivalent frame rate of the DVS is twice the pixel bandwidth (24kfps at 1.5klux scene illumination). The chip includes a fully configurable bias generator circuit [5] that

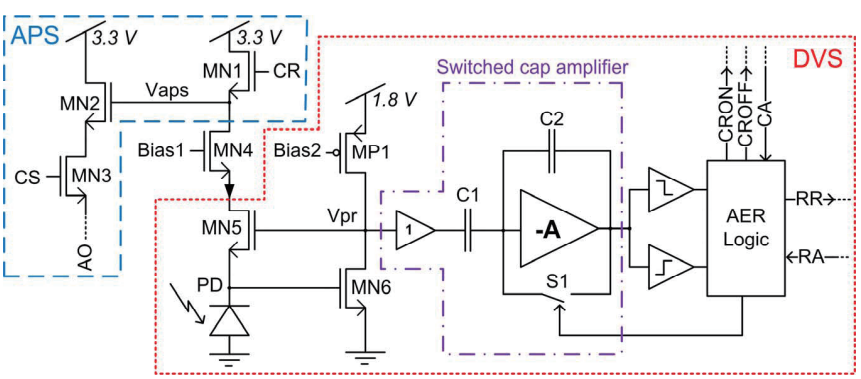

Fig. 1 apsDVS pixel schematic. The detailed transistor level schematic can be found in [1].

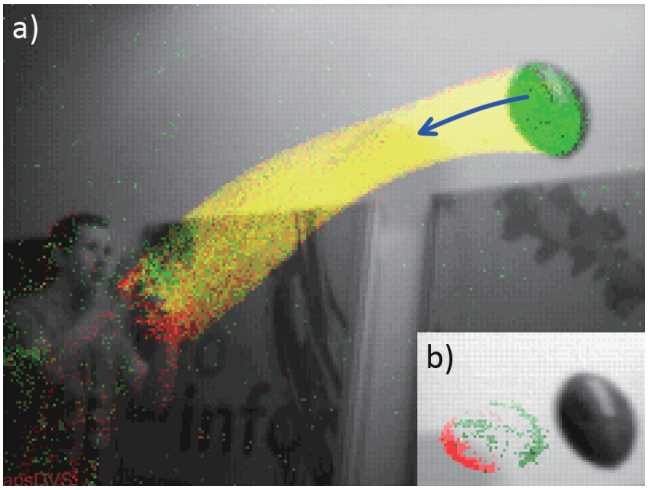

Fig. 2 (a) Combined 378ms snapshot showing an APS frame (gray) and the sparse DVS events (green/red) of a flying ball caught by a person. Most of the events originate from the moving ball. (b) 5ms slice of DVS events of the ball in the air, taken 75ms after the APS snapshot.

generates PVT tolerant biases. The quad mirrored pixel layout shares bias lines and source contacts. A minor checkerboard FPN in the APS readout may be an artifact of the layout.

Conclusion

With the combination of frame-based image readout and asynchronous data-driven temporal contrast readout, the apsDVS outputs information on static as well as dynamic visual scene content. This dual readout is achieved through a shared photodiode and by adding only 4 transistors to the original DVS pixel. The 240x180 sensor achieves a minimum DVS latency of 12us and 10mW power consumption which is 10 times less than the ATIS. The DVS pathway has a dynamic range of 120dB with 12% contrast detection threshold and 3.5% contrast matching; the APS readout has 57dB dynamic range with 1% FPN.

The temporal contrast events enhance the frame-based approach in several ways. In mobile devices, autonomous robots, and machine vision applications where power consumption and latency are critical factors, these events can be used to track moving features in a frame. With this approach, the processing latency and cost are decreased; and a low frame rate can be used which reduces the power consumption. In wireless sensor networks, the events can serve as activity cues to control the frame rate or the readout of a region of interest to reduce the output data rate.

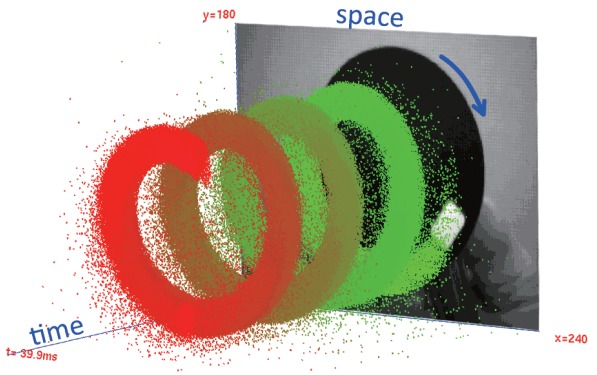

Fig. 3 Space-time 3D view of DVS events during 40ms of a white rectangle spinning on a black disk at 100Hz. Green events are older and red events are newer.

Acknowledgements

This research was supported by the European Union funded project SeeBetter (FP7-ICT-2009-6), the Swiss National Science Foundation through the NCCR Robotics, ETH Zurich, and the University of Zurich.

REFERENCES

[1] P. Lichtsteiner, C. Posch, and T. Delbruck, "A 128×128 120 dB 15 µs Latency Asynchronous Temporal Contrast Vision Sensor", ***J. Solid State Circuits***, vol.43, pp. 566-576, 2008.

[2] J. Lenero-Bardallo, T. Serrano-Goterredona, and B. Linares-Barranco, " A 3.6us Latency Asynchronous Frame-Free Event-Driven Dynamic-Vision-Sensor ", ***J. Solid State Circuits***, vol. 46, pp.1443-1455, Jun. 2011.

[3] C. Posch, D. Matolin, and R. Wohlgenannt, "A QVGA 143 dB Dynamic Range Frame-Free PWM Image Sensor With Lossless Pixel-Level Video Compression and Time-Domain CDS ", ***J. Solid State Circuits***, vol. 46, pp. 259-275, 2011.

[4] K. Boahen, "A Burst-Mode Word-Serial Address-Event Link-I: Transmitter Design ", ***IEEE Trans. on Circuits and Systems I: Regular Papers*** , vol. 51, pp. 1269-1280, Jul. 2004.

[5] M. Yang, et al., "Addressable Current Reference Array with 170dB Dynamic Range", ***Proc.of ISCAS***, pp. 3110-3113, May 2012

TABLE I Specifications compared to other data-driven vision sensors

	This work	Posch et al. [5]	Lenero et al. [3]
Functionality	**async. temp. contrast+APS**	async. temp. contrast +level crossing intensity	only async. temp. contrast
CMOS Technology	**0.18um 1P6M MIM**	0.18um 1P6M MIM	0.35um 2P4M
Array size	**240 x 180**	304 x 240	128 x 128
Pixel size um^2	**18.5 x 18.5**	30 x 30	35 x 35
Fill factor	**22%**	20%,10%	8.7%
Pixel complexity	**44 transistors, 2 MIM caps, 1 pd**	77 transistors, 3 caps, 2 pds	N.A.
Supply voltage	**1.8V / 3.3V (pixels)**	3.3V analog, 1.8V digital	3.3V
Power consumption	**7.4-13.5mW**	50-175mW	132-231mW
Dynamic range	**120dB DVS, 57dB APS**	Intensity 125dB, DVS N.A.	>100dB (56dB intrascene)
Contrast sensitivity	**12%**	13% @ 100lx, 30% @1klux	10%
FPN	**1 % APS, DVS 3.5%**	<0.25% intensity, DVS N.A.	DVS 4%
Max. bandwidth	**30MEvents/s, 40fps**	N.A.	N.A.
Min. latency	**12us @ 1klux (mean of 20 pixels)**	<4us @ 1klux (fastest pixel)	3.6us @ 25klux (fastest pixel)