Color Temporal Contrast Sensitivity in Dynamic Vision Sensors

Diederik Paul Moeys[1], Chenghan Li[1], Julien N.P. Martel[1], Simeon Bamford[2], Luca Longinotti[2],
Vasyl Motsnyi[3], David San Segundo Bello[3], Tobi Delbruck[1]
[1]Institute of Neuroinformatics (INI), University of Zürich and ETH Zürich, Switzerland
[2]iniLabs GmbH, Zürich, Switzerland, [3]IMEC research institute, Leuven, Belgium

***Abstract*— This paper introduces the first simulations and measurements of event data obtained from the first Dynamic and Active Vision Sensors (DAVIS) with RGBW color filters. The absolute quantum efficiency spectral responses of the RGBW photodiodes were measured, the behavior of the color-sensitive DVS pixels were simulated and measured, and reconstruction through color events interpolation was developed.**

I. Introduction

Dynamic Vision Sensors (**DVS**) [1] introduced the concept of frame-free image sensors. These bio-inspired sensors solely output the x-y position of temporal changes in logarithmic intensity, asynchronously for each pixel using an Address Events Representation (**AER**). The increased dynamic range (>100 dB) and reduced redundancy of information can allow 1) more robust vision in badly controlled lighting, 2) reduced power consumption and 3) sub-millisecond latency. Various generations of DVS sensors have been developed over the last decade. The introduction of the **DAVIS** provide intensity read-out (Active Pixel Sensor, **APS**) by just adding 5 transistors and sharing the same photodiode, thereby not increasing pixel area greatly [2].

Color is a useful feature for vision systems. Color image sensors using a color filter array (**CFA**) are widely available. Previous DVS sensors were almost all gray cameras (no CFA), or tried to achieve color sensitivity by using buried double or triple junction color separation. Color is not essential for many problems as proven by the many nocturnal animals that are monochromats and the large market for gray cameras. However, color is a discriminative cue that can be used, for instance to track an object (e.g. in RoboCup or traffic signs) using a simple method segmenting the image based on, for example, the hue component of each pixel.

The **CDAVIS** sensor [3] introduced a heterogeneous optical array composed of one white (no filter) DAVIS pixel every three APS pixels having overlying RGB color filters, after the basic concept introduced by J. Maxwell in 1855. The CDAVIS provides high-quality global or rolling-shutter APS frame using pinned photodiode technology, but not a color DVS read-out. Thus, the color readout has limited sample rate (30 Hz) and dynamic range (60 dB), like conventional image sensors.

A color DVS sensor [4] was attempted with blue/red change detection with the use of two photodiodes at different depths. The results suffered from noise, low sensitivity, and poor color separation and had to use poorly optimized n-well and p-source junctions as photodiodes. Ref. [5] showed a tri-color sensor using buried triple junctions available in a deep-submicron process that included a deep n-well option. It had poor color separation and the intensity-coded event output is redundant and unfairly allocates more AER bandwidth to brighter areas.

The results reported in this paper are the first study of color-sensitive DVS pixels with overlying CFA. The underlying sensor from which these color events are generated, is a DAVIS circuit inspired by [6] and [4], called SDAVIS192. SDAVIS192 was fabricated with a CMOS Image Sensor (**CIS**) 180 nm process of Towerjazz, with RGBW sensitivity thanks to RGB color filters and no filter for white. The sensor also has 1 µm-thick 15x15 µm microlenses on each 18.5x18.5 µm pixel.

This work also shows results of a scene reconstruction obtained with SDAVIS192 events. A simple method for reconstruction is used to show that DVS has indeed color sensitivity, but other methods such as [7] and [8] could be used to attain better results. In addition, we suggest a computational approach to reconstruct color intensities from color events.

II. Pixel Characterization and Simulation

An approximate model of a DVS pixel with a color filter is shown in Fig. 1. When a surface reflects light to the DVS sensor, it is focused by the lens on the pixel array. The light spectrum is modified depending on the spectral absorption characteristics of the color filter. Once the light has reached the photodiode, it undergoes the final spectral absorption to produce electron-hole pairs in the photodiode. This process depends on the depth of the photodiode and on the doping profile. It can be approximated by the tendency of silicon to favor red and near infrared (**NIR**) light; shallow blue-generated holes in the n-type photodiode tend to recombine at the surface. Thus the photon-to-electron efficiency always has a quantum efficiency less than one. The DVS processing described in [1]

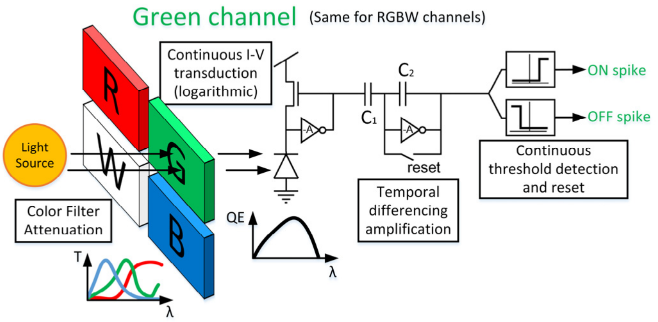
Green channel
(Same for RGBW channels)
Continuous I-V transduction (logarithmic)
C₂
C₁
ON spike
OFF spike
R
G
B
Light Source
reset
Continuous threshold detection and reset
Color Filter Attenuation
Temporal differencing amplification
QE
T
λ
λ

Fig. 1 Simplified block diagram of DVS pixels with color filters. Graphs show CFA transmission coefficients T vs wavelength λ and photodiode QE vs λ.

This research is supported by the European Commission project VISUALISE (FP7-ICT-600954), SeeBetter (FP7-ICT-270324) and Samsung. We would like to thank the Sensors group at INI Zürich and iniLabs GmbH.

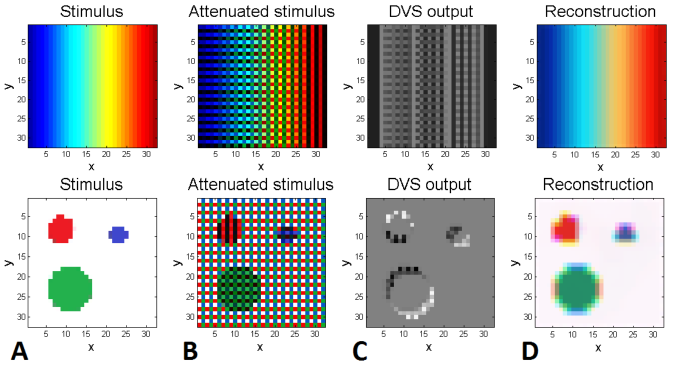

Fig. 2 Synthetic stimuli and responses. **A**: two synthetic stimuli: a rainbow varying sinusoidally in intensity in time and three moving color blobs. **B**: stimuli as seen by the silicon after color filtering. **C**: DVS histogram output (ON events are represented in white, OFF events in black and lack of events in grey), showing responses of the color DVS pixels. **D**: simple reconstruction [9].

is then applied to the sensed current: the logarithmic voltage conversion, the differencing amplification and the threshold comparison. The last step gives rise to ON and OFF events depending on if the positive or negative temporal contrast is large enough to trigger one of the two comparators. The idea of a color event, is that if for example a blue bar moves over a red background, at the leading edge, R DVS pixels will respond with a OFF events while B DVS pixels will respond with ON events. G pixels also respond depending on the change of G pixel photocurrent across the edge. The opposite responses would occur at the trailing edge.

The model described in Fig. 1 was implemented in MATLAB and used to generate synthetic DVS responses for various stimuli of various contrast levels and colors (Fig. 2). A 32x32 input frame-based video was the stimulus to the software sensor model. Every pixel value (ranging 0-255) is interpreted as a combination of three pure wavelength components: red at 650 nm, green at 510 nm and blue at 470 nm (based on the manufacturer's datasheet). Videos of example stimuli, are available at [9].

To compute the attenuation of each color filter on the RGB value of the color, together with the wavelength-dependent silicon absorption, our model takes into account the measured absolute external quantum efficiency (**QE**, generated electrons over incoming photons) of the sensor, as a function of wavelength λ. "External" here means the effective QE of the entire pixel, including the effects of reflectance, absorption, and particularly the limited Fill Factor (FF) of the pixel (21.2%). The QE measurements were performed at IMEC, using a computer controlled monochromator to generate narrow band monochromatic light from a wide-band white light source, and comparing the measured photocurrents with a calibrated reference photodiode. The resulting QE curve for each color channel is shown in Fig. 3. The peak QE of the W channel is 18% at 600 nm. This relatively high QE, so close to the ideal maximum, set by the FF of the pixel, can be linked to the pixel's special features: the buried photodiode (to double the junction's area), the extra deep p-implant under the n-well (to prevent it from stealing photo-generated carriers) and the microlens (to focus the light on the photodiode). The cut-off at 300 nm is caused by the UV absorption of the cover glass. The three RGB color channels show that while B and G are attenuated by 30% compared with R, they have sharp selectivity. R almost exactly follows the white response from 600 nm to longer wavelengths. All channels respond to NIR light at 850 nm, where B and G display a secondary peak of about 70% of the main peak's height. R only has a secondary 20% peak at 350 nm, in the UV region. These responses show that a filter blocking NIR (hot mirror), commonly used on color cameras, could increase color selectivity but it was not used in this study.

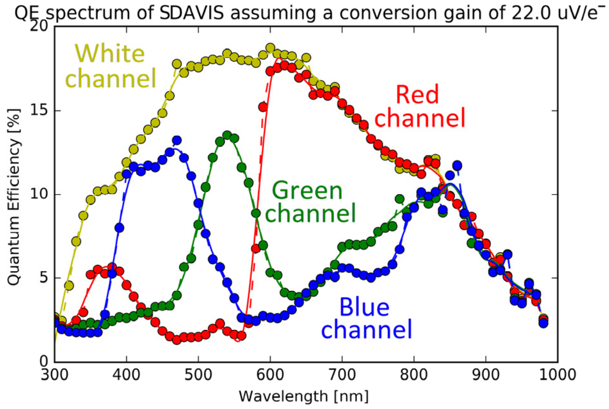

Fig. 3 Measured external QE spectrum of the RGBW channels of SDAVIS192.

When modeling the DVS operation, all absorption factors are therefore already taken into account (Fig. 2B), before the logarithmic transformation of the sum of each RGB input component, since the DVS circuit is colorblind. The temporal contrast is then computed by summing the logarithm of each pixel intensity value ("**brightness**") of a particular frame of the stimulus with the previous memorized brightness. When the difference exceeds the ON or OFF threshold of the DVS comparators, θ_{ON} and θ_{OFF} respectively, an event is generated and the new brightness is memorized. The entire mechanism can be approximated by Eq. (1), if all attenuations to the intensity I are incorporated in a single factor α, and the refractory period of the pixel is ignored. This means that the pixel will produce as many events as the contrast would dictate. The condition for generating events is thus:

$$\begin{cases}\Delta\ln(\alpha I) < -\theta_{OFF} & \text{OFF event}\\ \Delta\ln(\alpha I) > \theta_{ON} & \text{ON event}\end{cases} \tag{1}$$

where α and I encode for the three RGB components. To model multiple events due to high contrast moving edges, the temporal contrast is divided by the threshold it exceeds, and the resulting number is floored to give the total number of events the edge would produce. Fig. 2C shows this by presenting a 2D histogram of DVS activity that represents the events between two frames of the input stimulus. Videos showing the response of the single top corner pixels are available at [10].

III. Reconstruction

Although intensity reconstruction from events might not be needed in an event-driven algorithm, in this study, it was used to better understand the color DVS responses. Results from a naïve reconstruction are first presented, and then results of a more sophisticated reconstruction are presented.

A. Algorithm for naïve reconstruction

The simplest way to reconstruct intensity from events consists in using the inverse model of the DVS pixel presented in Section II. In order to reconstruct a color video from the DVS output, a frame is needed as an initial "condition". The new generation of DVS sensors, the DAVIS [2], includes a conventional APS circuitry that could supply this initial frame. The algorithm is summarized in the following steps, where simple reconstruction of each color channel is computed separately.

1. The DVS events produced by each color pixel are binned in 2D histograms of regular time intervals of duration t_b. Therefore $1/t_b$ is the frame-rate of the reconstructed video. The events are timestamped with microsecond resolution and typically jitter is less than 1 ms, therefore t_b could be chosen ranging over a few hundred microseconds up to a few seconds. In the following experiments t_b was set to 2 ms to achieve a frame rate of 500 Hz.
2. θ_{OFF} and θ_{ON} are measured to estimate the sizes of the log intensity steps. These are known in simulation, but for real data, a starting point can be estimated from the contrast sensitivity measurements of SDAVIS192. One of the two thresholds can then be adjusted to balance the number of ON and OFF events; if the object passes sideways a leading edge is always followed by a trailing edge. Since noise and physical movements do not always match this assumption, a manual correction factor is added. Every binned frame, containing $\Delta\ln(\alpha I)$ must then be multiplied by θ_{OFF} or θ_{ON}, depending on if the value of the pixel is negative or positive.
3. The logarithm of the first frame (either real or gray), is used to recreate all of the following logarithm frames. Each channel is then separately exponentiated element by element and multiplied by the inverse of the overall attenuation factors (the external QE).
4. Spatial color interpolation of each pixel's R, G, or B channel value is computed using to estimate the full RGB color values of each pixel: this step is known as demosaicking.
5. Each of the color reconstructed channel outputs (R, G and B) then represents one of the three RGB channels of the final reconstruction, which is computed by assuming that what each channel sees is monochromatic, with a light intensity matching the peak wavelength λ of the filter. This assumption breaks if, for example, a non-negligible amount of IR light falls on the R channel. For this reason, the sensor should use an IR-cut filter in the future.
6. The "gray world" assumption, stated in Eq. (2), is used to balance colors by multiplying each channel pixel (R_i, G_i and B_i) of a reconstructed frame by the largest color mean, i.e $\max(R_m, G_m, B_m)$ and dividing it by the mean of the channel (R_m, G_m or B_m).

$$\begin{cases} R_{gw} = R_i(\max(R_m, G_m, B_m))/R_m \\ G_{gw} = G_i(\max(R_m, G_m, B_m))/G_m \\ B_{gw} = B_i(\max(R_m, G_m, B_m))/B_m \end{cases} \tag{2}$$

7. Finally, the contribution of W is used to scale all channels. Since the average of a W frame W_m should roughly correspond to the average of all three RGB channel means $\mathrm{mean}(R_m, G_m, B_m)$, their ratio is used to scale the RGB channels to achieve white-balanced channels R_{wb}, G_{wb} and B_{wb}, as Eq. (3) shows.

$$\begin{cases} R_{wb} = R_{gw}W_m/\mathrm{mean}(R_m, G_m, B_m) \\ G_{wb} = G_{gw}W_m/\mathrm{mean}(R_m, G_m, B_m) \\ B_{wb} = B_{gw}W_m/\mathrm{mean}(R_m, G_m, B_m) \end{cases} \tag{3}$$

B. A computational approach for intensity reconstruction

We also present results from an algorithm that attempts a reconstruction without an explicit modelling of the DVS pixel. The goal is to reduce the effects of spurious events, and mismatch between pixels, smoothing the resulting intensity reconstruction by assuming that the spatial gradient $\vec{G}$ of the intensity I derives from a real scalar potential: $\vec{G} = \vec{\nabla} I$. This potential obliges the spatial gradients to form a conservative vector field. Because of noise, the spatial gradients computed from an event time surface do not generally form such a conservative field. However, it can be enforced by looking at its divergence $\mathrm{div}(\vec{G}) = \mathrm{div}(\vec{\nabla} I) = \nabla^2 I$ and solving for I. This Poisson equation gives the intensity as if it was coming from a conservative $\vec{G}$. Gradients computed from an event time surface can be assumed to contain a clean part (a part deriving from I) and a noisy part (caused by spurious events) such as: $\vec{G} = \vec{G}_{clean} + \vec{G}_{noise}$. The divergence cancels this noisy part since $\mathrm{div}(\vec{G}) = \mathrm{div}(\vec{G}_{clean})$ by invoking Helmholtz decomposition. This summarizes the algorithm's principle:

1. We compute a gradient $\vec{G}$ from the event time surface: it contains noise since the time surface contains noisy events.
2. We solve for the intensity I that would have created the conservative part of the gradient (the clean part: $\vec{G}_{clean}$) solving the Poisson equation $\mathrm{div}(\vec{G}) = \nabla^2 I$.
3. From the reconstructed I a new estimation of $\vec{G}$ using $\vec{G} = \vec{\nabla} I$ can be derived (which then only contains the clean part).
4. This estimate is blended to the noisy estimate.

This procedure can be iterated over many times and cleans $\vec{G}$, while reconstructing a "regularized" I. Real time reconstruction is possible on a laptop.

IV. Results

A. Naïve reconstruction

The results of the simulation's reconstruction are shown in Fig. 2D and are available at [9]. Since the initial frame is known and events are generated with known parameters, the reconstruction is limited only by the quantization from the threshold. Fig. 4B shows recorded DVS raw data and Fig. 4C the linear reconstruction results obtained with the unpublished 188 x 192 SDAVIS192 sensor for the stimuli of Fig. 4A. The distorted shapes in the lower part of the image are caused by the fact that the vision chip was not placed correctly in the center of the cavity of the package. The sensor therefore sees the distortion caused by the edges of the 1/3" 2.6 mm lens. Fig. 4B shows there is no immediately visible difference in the separated DVS channel outputs. This means that no matter what we think the color of an object is, its spectrum is much wider

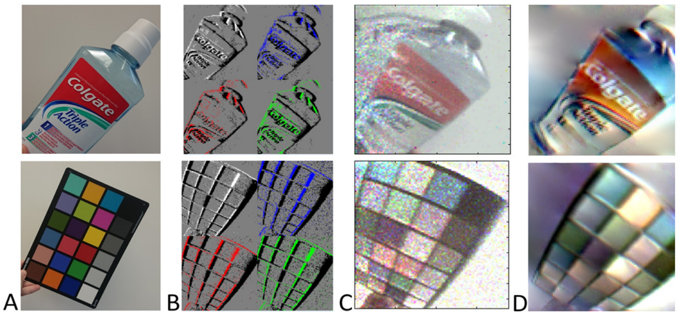

Fig. 4 SDAVIS192 raw data and reconstructions. **A**: stimuli: a moving mouthwash bottle and Macbeth color checker chart. **B**: color RGBW DVS events. Events shown in RGBW represent color ON events of RGBW type respectively. OFF events of all types are shown in black. **C**: naïve reconstruction from the 2 ms DVS bins [11]. **D**: computational approach.

than a single monochromatic color, and produces photocurrent in all RGB channels (as already clear from Fig. 3). To assess whether the imperceptible difference in response really exists and encodes for color difference, a simple image reconstruction can be used to check if the reconstruction results in an image with recognizable shapes and colors.

Fig. 4C shows that the color DVS pixels do encode for color and allow a basic color reconstruction. The naïve reconstruction is shown starting from a white background after approximately 300 2 ms-frames have been integrated. The best results were obtained for objects moving in front of a white background, where the initial frame could be pure white. As can be seen from Fig. 4C, where a mouthwash bottle and a Macbeth color checker chart were waved in front of the sensor, all colors can be detected and gray is preserved as gray, but there is noise; many pixels have the wrong color. Videos of these reconstructions starting from a white frame and played at 10 Hz frame rate, are available at [11]. The trail behind the mouthwash bottle is a result of the imbalance in θ_{OFF} and θ_{ON} and non-idealities including finite refractory period. Tuning imbalance is limited by inter-pixel mismatch in θ_{OFF} and θ_{ON}, resulting in different DVS pixel firing rates. If the threshold that should detect the trailing edge is not as low as the one detecting the leading edge, then the pixel might not fire for colors that only contain a low amount of the three principal RGB filters, also producing a trail.

Strong coupling between APS frame capture and immediately subsequent DVS activity in this SDAVIS192 prototype prevented us from using APS frames as starting frames. The imbalance or total lack of the type of event with less sensitivity (ON in the case of SDAVIS192) also means that the trails of objects add up to the other colors, blurring the reconstruction.

B. Computational approach

The reconstruction following the computational approach we proposed is compared to the naïve reconstruction method and can be seen in Fig. 4D. As can be seen in Fig. 4C, while the first method works with its known limitation, the computational approach of Fig. 4D instead allows a much sharper reconstruction with the removal of the noisy background. Though colors in the Macbeth color checker chart of Fig. 4D are not rendered exactly (signifying that further color correction would be needed), these are nonetheless more uniform than in Fig. 4C, with less salt-and-pepper noise.

V. Conclusion

The paper shows the first results from a DVS sensor with RGBW color filters. The event data generated encodes for fast color change information. The video interpolation from the DVS data show moderately good initial results with a white background as the starting point of the reconstruction, although there is large room for improvement and more powerful nonlinear probabilistic methods such as the one previewed in section IV.B will be adapted for color with better results. These computational methods, currently under development and shortly available, should improve the image deterioration introduced by DVS noise and coupling events, as well as compensate the mismatch in threshold sensitivities of the sensor. Standard color calibration techniques should also be used to improve the reconstruction, but are also beyond the scope of this work.

References

[1] P. Lichtsteiner, C. Posch, and T. Delbruck, “A 128 x 128 120 dB 15 µs Latency Asynchronous Temporal Contrast Vision Sensor,” *IEEE J. Solid-State Circuits*, vol. 43, no. 2, pp. 566–576, 2008.

[2] C. Brandli, R. Berner, M. Yang, S.-C. Liu, and T. Delbruck, “A 240x180 130 dB 3 us Latency Global Shutter Spatiotemporal Vision Sensor,” *IEEE J. Solid-State Circuits*, vol. 49, no. 10, pp. 2333–2341, Oct. 2014.

[3] C. Li *et al.*, “Design of an RGBW color VGA rolling and global shutter dynamic and active-pixel vision sensor,” in *2015 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2015, pp. 718–721.

[4] R. Berner and T. Delbruck, “Event-based color change pixel in standard CMOS,” in *Proceedings of 2010 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2010, pp. 349–352.

[5] L. Farian, J. A. Lenero-Bardallo, and P. Hafliger, “A Bio-Inspired AER Temporal Tri-Color Differentiator Pixel Array,” *IEEE Trans. Biomed. Circuits Syst.*, vol. PP, no. 99, pp. 1–1, 2015.

[6] T. Serrano-Gotarredona and B. Linares-Barranco, “A 128 x 128 1.5% Contrast Sensitivity 0.9% FPN 3 µs Latency 4 mW Asynchronous Frame-Free Dynamic Vision Sensor Using Transimpedance Preamplifiers,” *IEEE J. Solid-State Circuits*, vol. 48, no. 3, pp. 827–838, Mar. 2013.

[7] M. Cook, L. Gugelmann, F. Jug, C. Krautz, and A. Steger, “Interacting Maps for Fast Visual Interpretation,” in *The 2011 International Joint Conference on Neural Networks (IJCNN)*, 2011, pp. 770–776.

[8] P. Bardow, A. J. Davison, and S. Leutenegger, “Simultaneous Optical Flow and Intensity Estimation From an Event Camera,” presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 884–892.

[9] Diederik Paul Moeys, “Color-sensitive DVS simulation of stimuli and reconstruction.” [Online]. Available: https://www.youtube.com/watch?v=KlOhuiTOEQs&feature=youtu.be. [Accessed: 31-Oct-2016].

[10] Diederik Paul Moeys, “Color-sensitive DVS simulation and analysis.” [Online]. Available: https://www.youtube.com/watch?v=iipQGISTzpo&feature=youtu.be. [Accessed: 31-Oct-2016].

[11] Diederik Paul Moeys, “Real color RGBW DVS event reconstruction from white frame.” [Online]. Available: https://www.youtube.com/watch?v=xTsR7i-ho1U&feature=youtu.be. [Accessed: 31-Oct-2016].