Real-Time, High-Speed Video Decompression Using a Frame- and Event-Based DAVIS Sensor

Christian Brandli, Lorenz Muller and Tobi Delbruck
Institute of Neuroinformatics, University of Zurich and ETH Zurich, Zurich, Switzerland

***Abstract*—Dynamic and active pixel vision sensors (DAVISs) are a new type of sensor that combine a frame-based intensity readout with an event-based temporal contrast readout. This paper demonstrates that these sensors inherently perform high-speed, video compression in each pixel by describing the first decompression algorithm for this data. The algorithm performs an online optimization of the event decoding in real time. Example scenes were recorded by the 240x180 pixel sensor at sub-Hz frame rates and successfully decompressed yielding an equivalent frame rate of 2kHz. A quantitative analysis of the compression quality resulted in an average pixel error of 0.5DN intensity resolution for non-saturating stimuli. The system exhibits an adaptive compression ratio which depends on the activity in a scene; for stationary scenes it can go up to 1862. The low data rate and power consumption of the proposed video compression system make it suitable for distributed sensor networks.**

I. INTRODUCTION

Most video compression is done using frame-based sensors which read out every pixel once per frame, no matter whether the pixel changed or not. To compress this stream of full frames, redundancy can be reduced using spatial image compression, temporal motion compression or a combination of both [1]. Acquiring redundant information to later remove it, is inefficient and consumes computational resources and power. Moreover the temporal resolution is limited to the frame rate.

Dynamic vision sensors (DVSs [2], [3]) are a novel type of bio-inspired vision sensors which detect changes in the scene illuminance and report them as a stream of asynchronous events. DVSs have high dynamic range, high temporal resolution, low latency and they produce a sparse output. The pixels perform a simple form of video compression: If nothing is happening in the visual scene, the pixels do not report anything (except for some noise events). While DVS have successfully been applied in machine vision tasks [4], attempts to recover the compressed video signal were less successful or computationally heavy, mostly because DVSs cannot measure absolute intensities. A new generation of DVSs allows overcoming this problem by offering access to the absolute light intensities: The asynchronous time-based image sensor (ATIS) [5] does a change triggered intensity readout in the time domain; the dynamic and active pixel vision sensor (DAVIS) [6] is a more recent sensor that allows the acquisition of active pixel sensor (APS) intensity frames (i.e. CMOS image sensor technology). The ATIS has a high dynamic range of over 143dB for static scenes and it can be applied in compressive sensing setups to reconstruct static backgrounds [7]. However, the ATIS suffers from motion artifacts: The integration time for dark pixels can take up to 2s and the readout can be interrupted when changes in the scene occur, which makes the intensity readout of the sensor blind to fast movements. The intensity readout of DAVIS on the other hand offers a global, fixed integration time and can also be used to compress dynamic scenes.

This paper proposes the first real-time event decompression algorithm for the DAVIS. It allows online optimization of the event decoding.

II. THE DAVIS

The DAVIS [6] offers two concurrent types of readouts (Fig. 1a): It produces an asynchronous stream of DVS address events which encode changes in the pixel brightness i.e. temporal contrast, and at any point in time an exposure can be started (pixel reset in Fig. 1a), resulting in the output of a standard still image. The DVS events are generated in the following way: In the pixel the photocurrent is continuously logarithmically converted into a voltage which is internally sampled. Deviations from this sampled voltage are continuously amplified and when an ON (or OFF) threshold voltage is crossed, the pixel asynchronously requests to send

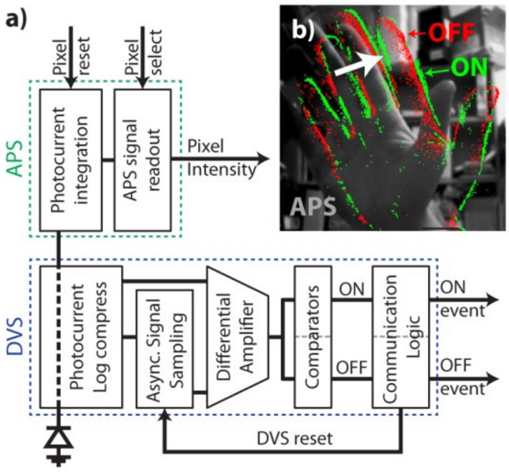


Fig. 1. The DAVIS. a) pixel block diagram and b) sensor output overlay of a waving hand (following the white arrow): grayscale APS background image and 4200 red/green rendered ON/OFF events of a 20ms time slice in the event stream.

This research is supported by the European Union funded projects SEEBETTER and VISUALIZE, the Swiss National Science Foundation through the NCCR Robotics, ETH Zurich, and the University of Zurich.

978-1-4799-3432-4/14/$31.00 ©2014 IEEE

out an ON (or OFF) event, depending on the sign of change. Peripheral circuits arbitrate between requesting pixels and transmit pixel addresses using the address-event representation (AER) [8]. As soon as the pixel transmits its event, it is reset (DVS reset in Fig. 1a), a new internal voltage is sampled and the process starts over. Fig. 1a) shows the output of a DAVIS: Since the DVS events are triggered asynchronously, they respond with less latency (typically sub-ms) to the motion of the hand than the scanned-out APS frames.

III. Decoding the DAVIS Output

A first approach to decompress a DVS signal was previously implemented in the jAER software (unpublished, code open-source [9]). jAER is an open-source, Java-based framework to acquire, process, analyze, visualize and store data from event-based sensors which use the address event representation (AER) communication protocol (such as the DVS, the DAVIS, silicon cochleas or event-based convolution chips). Apart from the required mechanical shutter, this first method has other drawbacks: The error in the brightness update per event which originates from the inter-pixel mismatch of the ON and OFF thresholds is integrated. By integrating the error, the image quality degrades quickly to the point where the scene can no longer be recognized. A second approach to decompress the DVS output uses a network of parameter maps to infer the intensity image [10]. While this approach performs qualitatively better than the first one, it is computationally heavy and the sensor motion is constrained to rotations.

The proposed algorithm uses the sampled still images obtained from the DAVIS which help to decompress the DVS event stream in three ways:

1. The images can replace the mechanical shutter by serving as starting point for the event decoding.
2. By continuously acquiring images, the integration of the DVS update error is reset on a regular basis so that it cannot grow too large.
3. The sampled absolute light intensities allow estimating the absolute temporal contrast (intensity steps) encoded by an event.

For an efficient implementation, the task of the proposed algorithm is divided into two parts:

1. Decompressing the data by adding intensity steps on the arrival of events (Fig. 2).
2. Estimating the intensity steps for the events.

The parameter space (i.e. circuit biasing) in which the DAVIS can operate is high-dimensional and the pixel response functions are difficult to model. For this reason an online modeling approach is applied that estimates the intensity step encoded by an event based upon the latest intensity samples.

The event-based decompression of the image (part 1) works as follows: Because the DAVIS pixel does a log compression of the photocurrent, the algorithm converts the intensity samples ($\hat{\mathrm{I}}_s$) of an acquired image to the log domain: $I_s = \log(\hat{\mathrm{I}}_s)$. This log-compressed image is then taken as the starting point for the event decoding, forming the initial decompressed image I_d. At the arrival of an ON/OFF event at pixel i, the current estimate of the ON/OFF event steps (positive $\delta_{ON,i}$ and negative $\delta_{OFF,i}$) is added to the to the log-compressed intensity image I_d at pixel i (= multiplication in linear domain). To obtain the final decompressed image $\hat{\mathrm{I}}_d$, I_d is exponentiated and displayed.

The estimates of $\delta_{ON,i}$ and $\delta_{OFF,i}$ (part 2) are updated when a new image (with index f) is sampled. The new image allows the calculation of the decompression error $E = I_d - I_s$ of the most recent decompression which used the estimates δ^{f-1}. To optimize the δ's, the error is used as a feedback signal and a fraction $\eta_e \cdot E$ of E is subtracted from the current $\delta_{ON,i}$and $\delta_{OFF,i}$. In case $\delta_{ON,i}$ is underestimating or $\delta_{OFF,i}$ is overestimating (as in Fig. 3), the errors become negative and the feedback makes the $\delta_{ON,i}$ larger and/or the $\delta_{OFF,i}$ less negative. However, when a pixel has seen both ON and OFF events between two image samples, it is not clear whether $\delta_{ON,i}$ or $\delta_{OFF,i}$ caused the error. Therefore the feedback term is multiplied by a factor that expresses how much $\delta_{\mathrm{ON},i}$was to blame for the error (the same error might partly have been caused by $\delta_{\mathrm{OFF},i}$). This second factor is given by the fraction of total events at this pixel i that were ON events $\frac{N_{\mathrm{ON},i}}{N_{\mathrm{tot},i}}$. The proposed method of error attribution requires only the number of events per polarity and not the full event history between two image samples, which makes it memory-efficient. The update is given by Eq.(1) ($\delta_{\mathrm{OFF},i}$ is computed equivalently):

$$\delta^{f}_{\mathrm{ON},i} = \delta^{f-1}_{\mathrm{ON},i} - \eta_e \cdot \mathrm{E}_i \cdot \frac{N_{\mathrm{ON},i}}{N_{\mathrm{tot},i}} \qquad (1)$$

A problem of this update rule is that it converges slowly, because each pixel needs to receive many events to provide a reliable estimate of $\delta_{\mathrm{ON},i}$. By making use of the fact that $\delta_{\mathrm{ON},i}$ is approximately the same (δ_{ON}) for all pixels, the convergence

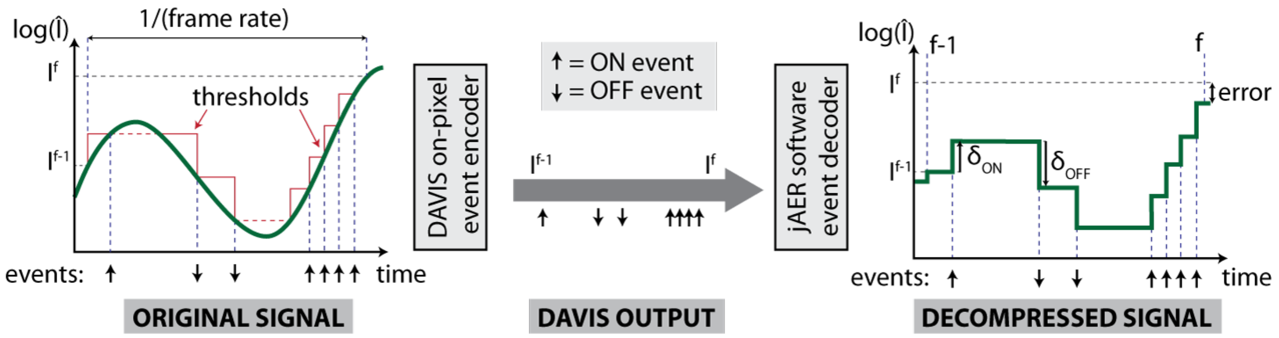
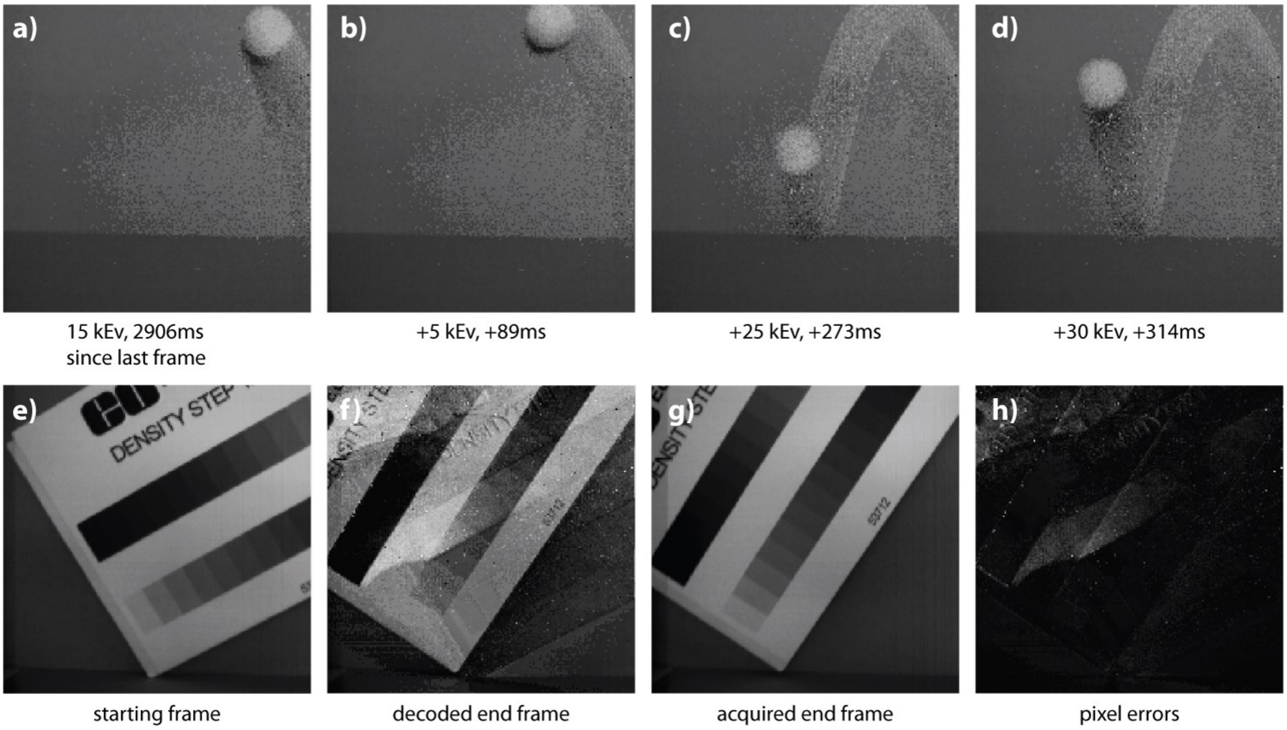


Fig. 2. Schematic representation of the proposed compressive on-pixel event encoder and software event decoder. I^f and I^{f-1} denote log-compressed intensities.



15 kEv, 2906ms since last frame

+5 kEv, +89ms

+25 kEv, +273ms

+30 kEv, +314ms

starting frame

decoded end frame

acquired end frame

pixel errors

Fig. 3. Decompressed DAVIS output. a-d) Decompressed images of a bouncing ping pong ball. The initial frame was caputured with no ball present. a) displays the redonstruction with the ball entering the scene 1.2s after the the initial frame. b-d) show the decompression at arbitrary timesteps relative to a). e-f) Comparison of a decompressed image to the original image after sweeping the DAVIS over a density step chart (each step representing 0.1 density step or 12.6% change of reflectance). e) is the initial frame, f) shows the decompressed image after 1.2Mio Events and 2.8s. g) displays the sampled frame at the same time as f). h) shows the difference between the sampled frame and the decompressed image which is the compression error per pixel (darker pixels mean smaller error). Recordings were made with η_e = 0.4 for global estimates, η_e = 0.004 for local estimate and η_a = 0.4.

can be sped up substantially. The global δ_{ON} is given by the average of all pixel estimates of $\delta_{\text{ON},i}$. For greater stability, this average is mixed with the previous estimate of δ_{ON} at a rate η_a (i.e. δ_{ON} is low-pass-filtered):

$$\delta_{\text{ON}}^{f} = \delta_{\text{ON}}^{f-1} \cdot (1-\eta_a) + \eta_a \cdot \sum\nolimits_{i\,\in\,\text{Pixels}} \frac{\delta_{\text{ON},i}^{f}}{N_{\text{pixels}}} \qquad (2)$$

$$\delta_{\text{ON},i}^{f} = \delta_{\text{ON}}^{f} \qquad (3)$$

The fast converging update rule shown in Eq.(2) and (3) is only used initially. Once this average has converged sufficiently, the pixel-wise update rule given by (1) is used, which makes it possible to model inter-pixel variance (i.e. mismatch in the pixel thresholds).

Beyond what we have described in detail above, the ON and OFF steps are also made dependent on the time dt_{last} elapsed since the last event arrived. This is necessary because in the current reset scheme of the DAVIS (and DVS), signal is discarded during the refractory period between the time the pixel crosses the threshold and the time the pixel reset is released [11]; a dependency on dt_{last} allows estimating this signal loss. The dt_{last} dependency is implemented using a binning approach where each of N_{bin}=8 logarithmically spaced dt_{last} bins contains separate δ steps which are adapted using the same error attribution principle as above.

The code of the presented algorithm can be found in the open-source jAER framework [9] in the class called *ApsFrameExtrapolationISCAS* which is in the project package *ch.unizh.ini.jaer.projects.davis.frames*.

IV. Results and Discussion

To evaluate the algorithms performance, several scenes were recorded and the intensity images were decompressed. To demonstrate the high temporal resolution of the presented compression approach, a bouncing ping pong ball was captured (Fig. 3a-d). The decoding of the DAVIS output ran in real-time in jAER [9] on a Core-Duo i7 3.07GHz desktop computer. The decompressed video can be rendered event by event as seen in Fig. 3 and even fast movements can be investigated at sub-ms resolution. The ball leaves a weak trace in the decompressed images because low contrast scenes handicap a good δ estimation.

To assess the image quality of the decompression, an Edmund density step chart (spanning a wide range of intensities) was rotated in the field of view of the DAVIS camera (Fig. 3e-h). The figure shows that the reconstruction is recognizable, but also that lag from the previous image sample can be observed. Furthermore the image exhibits some pixelated noise. These effects are thought to be caused by a non-optimal δ estimation, and the large temporal contrast quantization of the pixels.

Because this is the first publication of an algorithm for the decompression of temporal contrast events, there is no standard way of quantifying its results. Therefore the average pixel error after decompression is suggested as a reasonable metric for decompression accuracy. A uniform white bar with was swiped across the sensor's full field of view containing a homogeneous background of intensity I_{init}. After the sweep the intensity of the decompressed background is I_{dec}. Ideally I_{dec} should be equal to I_{init}; the error is defined as the average absolute difference between I_{dec} and I_{init}. For a short exposure of 1.42ms, an error of 0.5 DN (digital number: APS output ADC counts) was obtained, while an exposure of 5.23ms (the white bar intensity being just above the APS saturation level) resulted in an error of 2 DN. The result that stimuli close to the APS saturation lead to more error, illustrates the difficulty of mapping the wide logarithmic response of the DVS to the clipped linear response of the APS readout. These rather small DN errors are clearly not indicative of the actual appearance of the reconstructed images, which have clear lag artifacts.

The design of the DAVIS pixel leads to a dynamic data compression ratio *CR* that depends on the amount of change in a scene (in the form of total events per second eps from the sensor):

$$CR = \frac{\text{Uncompressed Data Size}}{\text{Compressed Data Size}} = \frac{R \cdot S \cdot f_e}{R \cdot S \cdot f_s + eps \cdot b} \quad (4)$$

where R is the intensity resolution in bits, S the size of the image in pixels, f_e the equivalent frame rate i.e. the frame rate which is required to capture a scene at the same temporal resolution, f_s is the sampling frame rate and b the size of one DVS event in bits. For the DAVIS (R=10bit, S=43200, b=32bit), the equivalent frame f_e rate corresponds to 1/temporal jitter of the events (i.e. the temporal precision of the events) which is approximately 1/500us = 2kHz at indoor illumination levels. The maximum compression ratio is therefore achieved when the sensor is run at low frame rate and no activity is present in the scene; a background noise activity of 2keps and 1Hz sampling frame rate yield a compression ratio of 1862. If an equivalent frame rate of 200Hz is assumed (because the temporal resolution might not be required to be higher) and with an event activity rate of 500keps (hand waving in front of sensor), the compression ratio is still 5. Under the assumption that most of the events are caused by contours, the compression ratio increases for higher spatial resolution sensors because the image size S grows quadratically with the scale factor while the length of the contours grows approximately linearly.

V. Conclusion

This paper presents the first publication on an algorithm for the decompression of asynchronous temporal contrast events. The algorithm is applied on the output of a dynamic and active pixel vision sensor (DAVIS) which performs in-pixel video compression in the form of temporal contrast events. The system allows real-time decompression of video information with sub-ms resolution and a dynamic, activity-dependent compressive ratio. These characteristics suggest two obvious fields of application: Firstly, by reducing the output data, the system allows reducing power consumption for data processing and communication so that it can be applied in surveillance tasks or distributed sensor networks, where most of the time nothing is happening and power consumption is of importance. Secondly, the system allows a simple inspection of fast processes without the necessity of a strong light source.

To achieve a more precise compression, a lossless encoding should be used which can be done by applying a DVS pixel reset scheme that does not discard signals (companion paper [11]). Increasing the sensitivity of the pixel [12] can further improve the quality of the compression by increasing the intensity resolution and decreasing the quantization effects, at least for linear decompression[11]. The decoding algorithm could further be improved by modeling intensities outside of the intensity range of the APS (but still in the DVS range), which would allow extending the dynamic range of the reconstruction outside the dynamic range of the APS readout.

Acknowledgment

The authors would like to thank Matthew Cook for fruitful discussions.

References

[1] A. Beach, *Video compression*. Berkeley, Calif.: London: Peachpit; Pearson Education [distributor], 2007.

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

[3] T. Delbruck, B. Linares-Barranco, E. Culurciello, and C. Posch, "Activity-Driven, Event-Based Vision Sensors," in *Proceedings of 2010 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2010, pp. 2426–2429.

[4] Z. Ni, C. Pacoret, R. Benosman, S. Ieng, and S. RéGnier, "Asynchronous Event-Based High Speed Vision for Microparticle Tracking," *J. Microsc.*, vol. 245, no. 3, pp. 236–244, Mar. 2012.

[5] 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," *IEEE J. Solid-State Circuits*, vol. 46, no. 1, pp. 259–275, Jan. 2011.

[6] R. Berner, C. Brandli, M. Yang, S.-C. Liu, and T. Delbruck, "A 240x180 10mW 12us Latency Sparse-Output Vision Sensor for Mobile Applications," presented at the Symposium on VLSI Circuits, Kyoto Japan, 2013, pp. C186 – C187.

[7] G. Orchard, J. Zhang, Y. Suo, M. Dao, D. T. Nguyen, S. Chin, C. Posch, T. D. Tran, and R. Etienne-Cummings, "Real Time Compressive Sensing Video Reconstruction in Hardware," *IEEE J. Emerg. Sel. Top. Circuits Syst.*, vol. 2, no. 3, pp. 604–615, 2012.

[8] K. A. Boahen, "Point-to-Point Connectivity Between Neuromorphic Chips Using Address Events," *IEEE Trans. Circuits Syst. II Analog Digit. Signal Process.*, vol. 47, no. 5, pp. 416–434, May 2000.

[9] "Accumulate Image Function in jAER Software," *jAER Open Source Project*. [Online]. Available: https://sourceforge.net/p/jaer/wiki/Home/. [Accessed: 17-Sep-2013].

[10] 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.

[10] M. Yang, S.-C. Liu, and T. Delbruck, "Comparison of Spike Encoding Schemes in Asynchronous Vision Sensors: Modeling and Design," in *International Symposium on Circuits and Systems (ISCAS) 2014*, Melbourne, Australia, 2014 (submitted).

[12] J. A. Lenero-Bardallo, T. Serrano-Gotarredona, and B. Linares-Barranco, "A 3.6 μs Latency Asynchronous Frame-Free Event-Driven Dynamic-Vision-Sensor," *IEEE J. Solid-State Circuits*, vol. 46, no. 6, pp. 1443–1455, Jun. 2011.