https://doi.org/10.2352/EI.2024.36.9.IQSP-259

Benchmarking Motion Blur of Video Frame Interpolation using Hybrid EVS+CIS against CIS

Kamal Rana, Sean Fausz, Zhiyao Yang, Fangwen Tu, Qinyi Wang, Boyd Fowler, Andreas Suess, Bo Mu; OMNIVISION, Santa Clara, California 95054, USA

Abstract

Event-based vision Sensors (EVS) utilize smart pixels capable of detecting whether relative illumination changes exceed a predefined temporal contrast threshold on a pixel level. As EVS asynchronously read these events, they provide low-latency and high-temporal resolution suitable for complementing conventional CMOS Image Sensors (CIS). Emerging hybrid CIS+EVS sensors fuse the high spatial resolution intensity frames with low latency event information to enhance applications such as deblur or video-frame interpolation (VFI) for slow-motion video capture. This paper employs an edge sharpness-based metric-Blurred Edge Width (BEW) to benchmark EVS-assisted slow-motion capture against CIS-only solutions. The EVS-assisted VFI interpolates a CIS video steam with a framerate of 120 fps *by 64x, yielding an interpolated framerate of* 7680 fps. *We observed that the added information from EVS dramatically outperforms a* 120 fps *CIS-only VFI solution. Furthermore, the hybrid EVS+CIS-based VFI achieves comparable performance as high-speed CIS-only solutions that capture frames directly at* 480 fps *or* 1920 fps *and incorporate additional CIS-only VFI. These, however, do so at significantly lower data rates. In our study, factors $\sim$ 2.6 and $\sim$ 10.5 were observed.*

Introduction

Video Frame Interpolation (VFI) generates synthesized intermediate frames between two consecutive keyframes to produce smooth and visually appealing videos (see Figure 1). The quality of the intermediate synthesized frame relies on the amount of motion between the two successive keyframes. The motion blur present in the keyframes is another factor impacting the quality of interpolated frames. Hence, mobile phone cameras capture videos of fast-moving objects with high frame rates to limit the amount of motion blur. Capturing high frame rates produces smoother videos but leads to high data and power consumption.

Hybrid camera sensors include image and event pixels. Event pixels report continuous intensity changes when they exceed a predefined contrast threshold [2]. Consequently, these novel hybrid sensors provide additional low-latency event information that can mitigate the two significant issues in video frame interpolation by deblurring keyframes and providing additional intensity information between the two consecutive captured frames, which is useful for VFI.

Data-driven techniques built for fusing events with intensity frames mainly depend on deep learning, contain millions of parameters, and thus are computationally expensive and, therefore, challenging for use in mobile platform applications. We utilize a low-computation, inexpensive, mathematical-based technique called an Event-Double Integral algorithm (EDI) to fuse an image with its corresponding events [5]. The EDI algorithm can deblur the keyframes and interpolate frames between two consecutive deblur keyframes; thus, we anticipate achieving high-quality video frame interpolation.

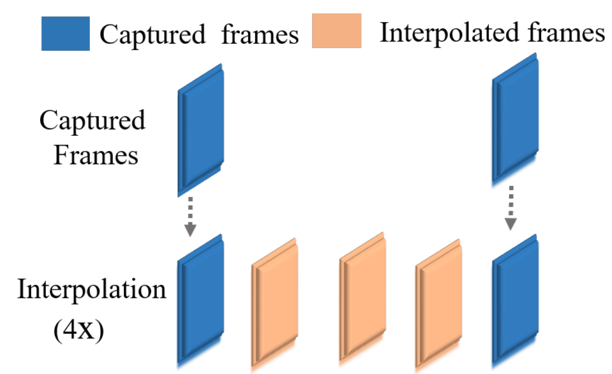

Figure 1. *Schematic of video frames interpolation mode in mobile phones.*

This work compares the VFI performance of a hybrid CIS+EVS VFI solution to a CIS-only solution. For this, we consider a use-case where we upscale a CIS key framerate of 120 fps to 7680 fps. To evaluate the quality of interpolated frames, we use *Blurred Edge Width* (BEW) metric that quantifies the sharpness of edges. Moreover, we compare the VFI performance of a hybrid CIS+EVS sensor with two phones offering high key framerates of 480 fps and 1920 fps. We observed that VFI using hybrid CIS+EVS provides comparable performance to high framerate phone cameras but with significantly lower data rate and potentially power consumption.

Related Work

We are unaware of a CIS-only VFI method with low computational cost. We consider RIFE (Real-Time Intermediate Flow Estimation) being a state-of-the- art CIS-only VFI method for further comparison against hybrid EVS+CIS VFI. Similar to other state-of-the-art methods, RIFE uses a deep learning network for bidirectional optical flow between intensity frames for VFI [4]. Deep learning (data-driven) methods based on hybrid CIS+EVS like Time Lens and Adhoc Deburring outperform VFI based on CIS-only [6]. However, these methods are deep learning-based and computationally expensive. The Event Double Integral (EDI) method combining frame and event information enables VFI with low computational complexity [5]. It can deblur the captured intensity frames and provide high-quality synthesized interpolated frames.

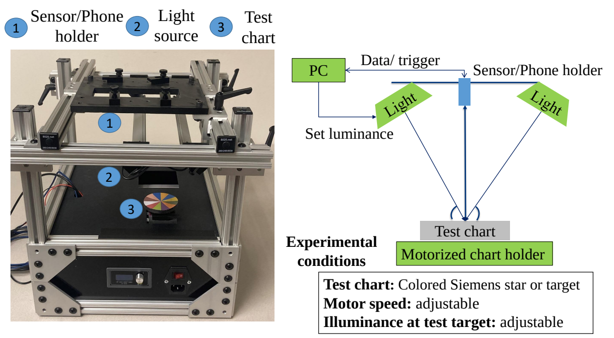

Figure 2. *(a) front-view of the apparatus. (b) schematic of the appartus setup*

Experimental Setup

A motorized disk surface is used to depict and rotate a Siemens star chart in a controlled manner (see Figure 2). The rotation of the target provides a repeatable evaluation of motion blur as a function of object speed and illuminance. A fixture is used to displace a camera at a controlled distance from the target to yield comparable FOV between different cameras. Instead of using a two-camera system with a collocated CIS and EVS camera, we utilize a dedicated hybrid EVS+CIS sensor [3]. This mitigates spatial and temporal alignment, avoids optical artifacts from occlusion/parallax, and reduces package size and cost. To compare hybrid EVS+CIS versus CIS-only at a comparable base framerate, we utilize the CIS channel of the hybrid sensor to emulate the CIS-only performance. For comparison with high-speed CIS-only solutions, we swap the hybrid sensor against commercial mobile phones implementing such sensors.

We used a hybrid CIS+EVS sensor with a CIS key framerate of 120 fps and a resolution of 1080p. To compare with high-framerate cameras, we used two flagship Phones: Phone A (480 fps, 720p, and 2x frame interpolation) and Phone B (1920 fps, 720p, and 4x frame interpolation).

The rotation speed of the color chart can be controlled in the range of 0 rpm to 300 rpm. The setup has two light-emitting diode (LED) floodlights with a 5600 K color temperature, placed at a 45 degree angle relative to the test chart. The illuminance at the test chart can be controlled from 0 to 12.000 lx. In the experiment setup, the camera remains stationary while the color disk rotates.

Blurred Edge Width Calculation by Curve Fitting

The blurred edge width method scans the intensity profile along an edge and measures the number of pixels it takes to change the intensity from 10 % to 90 % of its local range (see Figure 3) [1]. A sharper edge leads to lower BEW and vice-versa. The presence of noise in actual data creates challenges in calculating the accurate BEW Value. To solve the mentioned issues, we fit a sigmoid function that matches the curve, and then we calculate the BEW of the fitted curve. The fitted sigmoid function is shown below as $S(x)$:

$$S(x) = |Y_1 - Y_2| \frac{1}{1+e^{-(ax+b)}} + \min(Y_1, Y_2) \tag{1}$$

Here, Y_1 and Y_2 are the minimum and maximum luma values,

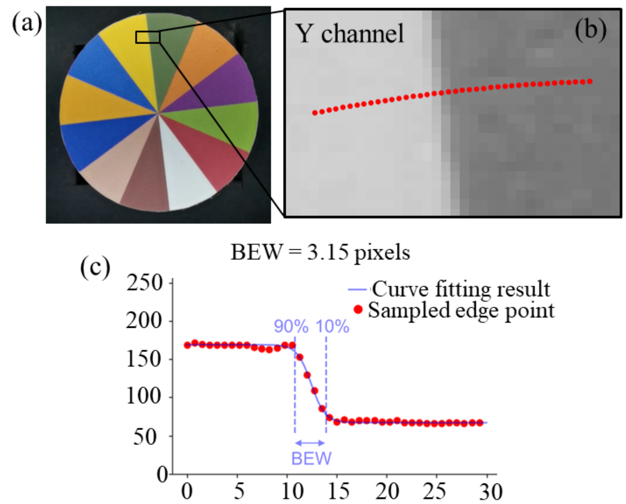

Figure 3. *(a) Color chart with edges detected. (b) pixels valued scanned at particular locations at the edges. (c) A sigmoid function fitted to pixel values and calculated BEW.*

$|Y_1 - Y_2|$ is the edge contrast, and $\min(Y_1, Y_2)$ controls the up/down shift of the fitting curve. The fitting parameters a and b are introduced to control the slope and position of the edge obtained by regression among sampled edge points.

Event Double Integral Method for hybrid CIS+EVS

The Event Double Integral algorithm fuses intensity images with event information to deblur captured frames and interpolate the frames between newly deblur frames. Here, we briefly explain the outline of the EDI algorithm: an event can be described as a direct delta function, $e(t) = p \cdot \delta(t)$, and based on the sum of events, we get proportional change in intensity between reference time f and t as $E(t) = \int_f^t e(s)\,\mathrm{d}s$. Here, p is polarity and can be $+1$ or -1 based on event generation with an increase or decrease in luminance. Using $E(t)$, we can get a latent image sequence at any time t as $L(t) = L(f)\exp(cE(t))$. Here, c is the contrast threshold, which means the pixel will trigger an event if there is a change in intensity above c. $L(f)$ is the latent frame at reference time f that can be easily calculated using a blurry image and the sum of events information. Using $L(t)$, we can calculate video frames at any time t between two captured frames.

Analysis of Experimental Results

We implemented the EDI algorithm on hybrid CIS+EVS with a CIS framerate of 120 fps to deblur the keyframes and interpolated 63 synthesized intermediate frames ($120 \times 64 =$ 7680 fps)(see Figure 4). We observed that the interpolated frames have higher BEW than the deblurred keyframes (see Figure 6). Moreover, we noticed that the BEW of actual keyframes is higher than that of deblurred keyframes, and the difference in BEW between them increases with increasing motion speed. Furthermore, we compared the hybrid CIS+EVS VFI performance versus CIS-only VFI. At a significantly higher motion speed, we observe that the CIS+EVS VFI-based low computation cost EDI algo-

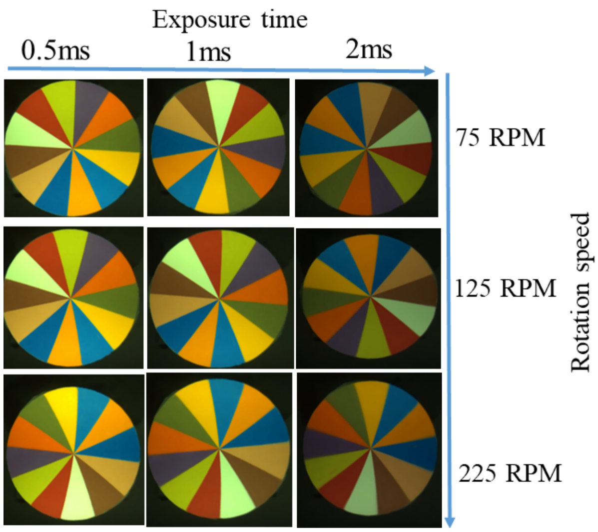

Figure 4. *Data samples collected from hybrid CIS+EVS sensor with various exposure times and the rotation speed of the color disk.*

	Phone A	Phone B	Hybrid Sensor (CIS +EVS)
Key Frame Rate	480 frames	1920 frames	120 frames
Interpolated Frame Rate	960 frames	7680 frames	7680 frames
Data Rate (Mb/s/pixel)	5.72×10^{-4}	22.8×10^{-4}	2.16×10^{-4}

Figure 5. *The figure depicts the key framerates, interpolated framerate, and the data rate normalized to pixels of Phone A, Phone B, and hybrid CIS+EVS sensor.*

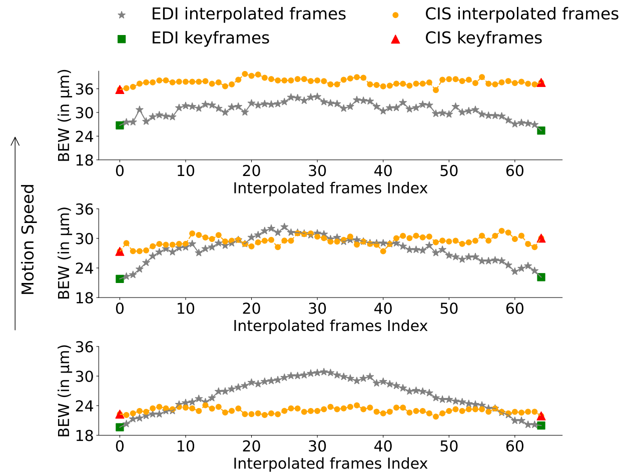

Figure 6. *The figure shows the BEW of CIS-only based interpolated frames versus interpolated frames based on CIS+EVS at different motion speeds. The figure also depicts the BEW of CIS+EVS keyframes and CIS-only keyframes.*

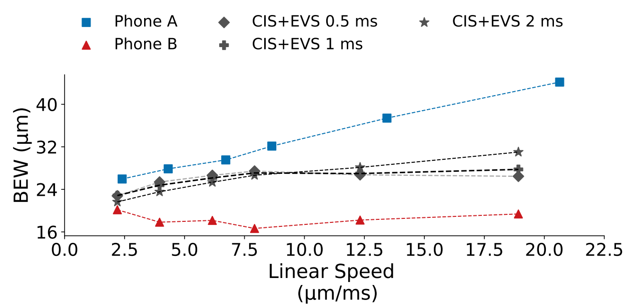

Figure 7. *The figure compares the BEW of Phone A, Phone B, and the hybrid CIS+EVS sensor at different motion speeds. The exposure time of the hybrid sensor is manually set to: 0.5, 1, and* 2 ms*, while the exact exposure times of Phone A and Phone B are unknown. We estimate the maximum possible exposure time of Phone A ($<$ 2 ms) and Phone B($<$ 0.5 ms) based on their framerates.*

rithm outperforms the CIS-only VFI method based on the computationally expensive deep learning-based algorithm RIFE (see Figure 6).

To compare the VFI performance of the hybrid CIS+EVS with CIS framerate of 120 fps against high frame rate cameras, we study the BEW of Phone A, Phone B, with hybrid CIS+EVS sensor with various exposure times and motion speeds. We observed that the BEW of the hybrid sensor increases with an increase in exposure time. Moreover, we discovered that at high motion speed, the hybrid sensor outperforms Phone A (480 fps) as the BEW of Phone A increases sharply (see Figure 7). In contrast, the BEW of the hybrid sensor slowly rose with increasing motion speed. Furthermore, we noticed that even with the increase in motion speed, there was little change in the BEW of Phone B (1920 fps).

Various parameters, such as exposure time and lens design, affect the BEW of key and interpolated frames of Phone A, Phone B, and the hybrid CIS+EVS sensor. For a fair comparison between the BEW of phone cameras and hybrid CIS+EVS sensor, we calculated the ratio of BEW of interpolated synthesized frames to keyframes, thus only evaluating the quality of interpolated frames relative to the keyframes (see Figure 8). The ratio of BEW interpolated frames w.r.t captured frames is almost 1 for Phone A and Phone B. The probable cause of the constant relative ratio is that Phone A and Phone B only interpolate 2 and 4 times, respectively. For the hybrid CIS+EVS sensor, the ratio is larger than 1 at low motion speed. As can be seen in Figure 6, at low temporal contrast hybrid EVS+CIS VFI based on the EDI method struggles to generate interpolated frames with low BEW. At high motion speed, however, the ratio reduces below 1 as hybrid deblur and VFI reduce the BEW of both the deblurred key frames as well as the interpolated frames.

The data rate of the phone cameras based on CIS-only is calculated as data rate = frame rate × raw image size. The camera raw image size is the same irrespective of the scene and depends on the pixel count. Phone A, with a framerate of 480 fps, has a

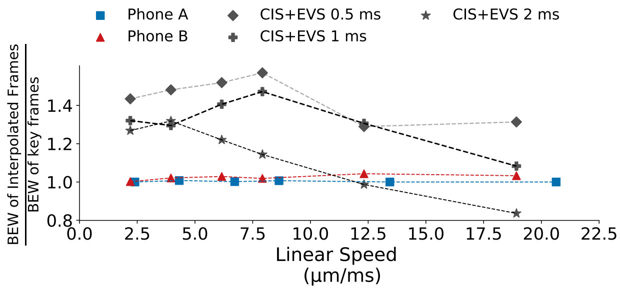

Figure 8. The figure compares the ratio of BEW of interpolated to keyframes at different motion speeds of Phone A, Phone B, and Hybrid CIS+EVS sensor. The exposure time of the hybrid sensor is manually set at different exposure times: 0.5, 1, and 2 ms, while the exact exposure times of Phone A and Phone B are unknown. We estimate the maximum possible exposure time of Phone A (< 2 ms) and Phone B(< 0.5 ms) based on their framerates.

four times lower data rate than Phone B operating at 1920 fps, as they have the same resolution of 720p. In hybrid CIS+EVS, the data rate is the sum of the data rate coming from intensity frames and events. The data rate of events is scene-dependent as each event pixel independently outputs events based on local changes in intensity. Consequently, the data rate of Hybrid CIS+EVS is scene-dependent. We compared the data rate of the hybrid sensor with different motion speeds and the event rate (see Figure 9). We observed that even with significant motion present in the scene, the data rate coming from events is significantly lower than that of the CIS-only solution.

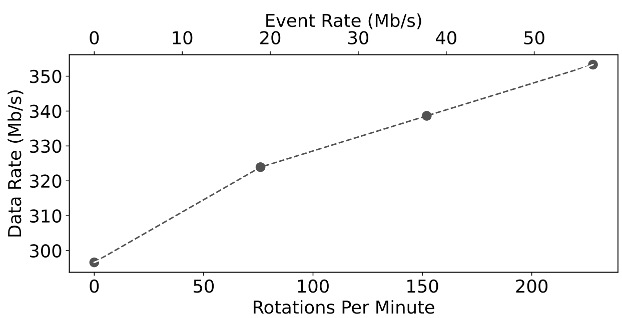

Figure 9. The hybrid CIS+EVS sensor data rate consists of the data rate of the CIS framerate and the number of events generated per second. The number of events depends on the amount of motion; with more motion, the number of events increases, and thus, the total data rate increases. In this figure, the events are originating from a rotating color disk.

We compared the data rate of Phone A, Phone B, and the hybrid sensor (see Figure 5). As Phone A, Phone B, and the hybrid sensor have different resolutions, we normalized the data rate to pixel count for a fair comparison. At significant motion speed, the data rate of the hybrid CIS+EVS sensor is ~ 2.5 and 10.5 times lower than of Phone A or B. Furthermore, we compare the data rate with the BEW of the phones and the hybrid CIS+EVS sensor. The hybrid CIS+EVS sensor achieves a lower data rate (less than 2.5 times) and lower BEW than Phone A (480 fps) (see Figure 10). Similarly, for Phone B, the hybrid CIS+EVS achieves a significantly lower data rate (less than 10.5 times) but at a comparatively higher BEW.

	BEW	**Data Rate (Mb/s/pixel)**
Phone A (960 frames, 2x Interpolation)	30	5.72×10^{-4}
Phone B (7680 frames, 4x Interpolation)	16	22.8×10^{-4}
Hybrid Sensor (CIS+EVS) (7680 frames, 64x Interpolation)	24	2.16×10^{-4}

Figure 10. The figure compares data rate with the BEW of Phone A, Phone B, and the hybrid CIS+EVS sensor. Here, BEW is calculated as the average over different motion speeds.

Conclusion and Discussion

In this work, we compared the performance of Video Frame Interpolation of hybrid CIS+EVS versus CIS-only cameras. We studied a hybrid image sensor that provides a CIS framerate of 120 fps as well as EVS information. We found that at significant motion speed, VFI using hybrid CIS+EVS generates sharper images than the CIS-only method. Also, we observed that the EDI-based method deblurs the captured frames with substantial motion, and the BEW of deblurred keyframes itself is significantly lower than that of blurred keyframes. Our results show that EDI, a low computational algorithm based on hybrid data, outperforms CIS alone VFI based on complex deep learning state-of-the-art methods in the presence of significant motion blur.

In addition, we compare the performance of hybrid CIS+EVS with two high frame rate flagship cameras: Phone A with a key framerate of 480 fps and Phone B with a key framerate of 1920 fps. The hybrid sensor achieves higher quality VFI results with a lower 2.5 times data rate compared to Phone A. Phone B achieves a lower BEW than the hybrid sensor but with a 10.5 times increased data rate compared to the hybrid CIS+EVS sensor. Our results show that VFI using a low-framerate hybrid CIS+EVS sensor achieves comparable BEW to high-frame cameras with very low data rates and potentially low power consumption. In the future, we plan to compare the VFI results of hybrid CIS+EVS sensor at higher motion speeds and different luminance levels. Moreover, we intend to compare VFI between phones and the hybrid sensor using different image quality metrics.

References

[1] Hai Dinh, Qinyi Wang, Fangwen Tu, Brett Frymire, and Bo Mu. Evaluation of motion blur image quality in video frame interpolation. *Electronic Imaging*, 35:262–1, 2023.

[2] Guillermo Gallego, Tobi Delbrück, Garrick Orchard, Chiara Bartolozzi, Brian Taba, Andrea Censi, Stefan Leutenegger, Andrew J Davison, Jörg Conradt, Kostas Daniilidis, et al. Event-based vision: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 44(1):154–180, 2020.

[3] Menghan Guo, Shoushun Chen, Zhe Gao, Wenlei Yang, Peter Bartkovjak, Qing Qin, Xiaoqin Hu, Dahei Zhou, Masayuki Uchiyama, Shimpei Fukuoka, et al. A 3-wafer-stacked hy-

brid 15mpixel cis+ 1 mpixel evs with 4.6 gevent/s readout, in-pixel tdc and on-chip isp and esp function. In *2023 IEEE International Solid-State Circuits Conference (ISSCC)*, pages 90–92. IEEE, 2023.

[4] Zhewei Huang, Tianyuan Zhang, Wen Heng, Boxin Shi, and Shuchang Zhou. Real-time intermediate flow estimation for video frame interpolation. In *European Conference on Computer Vision*, pages 624–642. Springer, 2022.

[5] Liyuan Pan, Richard Hartley, Cedric Scheerlinck, Miaomiao Liu, Xin Yu, and Yuchao Dai. High frame rate video reconstruction based on an event camera. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(5):2519–2533, 2020.

[6] Stepan Tulyakov, Daniel Gehrig, Stamatios Georgoulis, Julius Erbach, Mathias Gehrig, Yuanyou Li, and Davide Scaramuzza. Time lens: Event-based video frame interpolation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 16155–16164, 2021.