DOI 10.5162/SMSI2025/B6.4

Ev2Gray: Generating Intensity Images from Event Streams

Johannes Baßler[1] *and Michael Heizmann*[1]
[1] Karlsruhe Institute of Technology, Karlsruhe, Germany
johannes.bassler@kit.edu

Summary: This paper presents a novel method for extracting absolute intensity information from event-based vision sensors, addressing their limitation in capturing static scene regions. Our approach utilizes a controlled occlusion mechanism that triggers events across both dynamic and static areas. By analyzing event patterns during occlusion and revelation phases, we reconstruct intensity values for each pixel.

Keywords: Event-based Vision, Intensity Reconstruction, Neuromorphic Sensing, Occlusion Mechanism, High Dynamic Range Imaging

Introduction

Event-Based Vision (EBV) sensors employ a fundamentally different output type compared to frame-based camera sensors. While frame-based camera sensors generate absolute intensity information for all image pixels at fixed time intervals, EBV sensors produce asynchronous and sparse events that encode relative brightness changes in the spatio-temporal domain. Each event e is represented by a tuple of four values: (x, y, t, p), where (x, y) denotes the pixel coordinates, t the timestamp, and p the polarity of the event, which defines the sign of the brightness change. This unique approach allows EBV sensors to offer significant advantages such as high temporal resolution, low latency, and high dynamic range. However, a critical limitation of event streams is their inherent difficulty in semantic interpretation of the scene, as they lack information about non-moving regions such as static backgrounds and do not provide comprehensive structural information. This absence of complete spatial context poses significant challenges for traditional computer vision algorithms and hinders direct semantic understanding, necessitating novel approaches to bridge the gap between event-based data and meaningful scene interpretation. Many state-of-the-art event-based vision methods still depend on structural information from frame-based cameras [1] – a requirement our approach aims to address. To address these limitations, we propose a novel method that, firstly, triggers events in all parts of the image, including static regions, in a controlled manner, and secondly, extracts absolute intensity images from the resulting event stream data without the need for an additional sensor. This approach aims to combine the advantages of event-based sensing with the comprehensive spatial information typically associated with frame-based cameras, potentially enabling more robust semantic interpretation of event-based data.

Related Work

While numerous methods, such as `E2VID` [2] and `ET-Net` [3], perform intensity reconstruction for videos using deep learning algorithms, to the best of our knowledge, there is only one other method capable of translating static scenes into intensity images using an event-based camera. The `EvTemMap` method initially darkens the scene completely using the aperture and then maps the timing of the first positive event for each pixel to absolute intensity values during the continuous opening of the aperture. As will be demonstrated, our method differs significantly in both the generation of occlusion and the computation approach. Unlike `EvTemMap` [4], our approach requires only partial occlusion and evaluates multiple events per pixel rather than just the first, which should positively affect noise performance.

Method

Event Generation Process We employ a moving opaque strip that traverses the entire image plane, completely occluding light from the scene in a spatially continuous and sparse manner. This strip induces controlled intensity changes across all pixels, regardless of whether they correspond to static or dynamic regions of the scene. As the opaque strip moves across the field of view, it creates two distinct transitions for each pixel:

1. **Occlusion Transition:** When the strip enters a pixel's view.
2. **Revelation Transition:** When the strip exits a pixel's view.

The number of events generated during these transitions is roughly proportional to the magnitude of the intensity change. The key insight of

DOI 10.5162/SMSI2025/B6.4

our method is that the intensity level during full occlusion is uniform across all pixels. This provides a common reference point from which we can derive absolute intensity information.

Temporal Window Identification The first step in our procedure is to identify the relevant temporal windows for each pixel, specifically the occlusion and revelation phases. The method for determining these windows should be optimized based on the motion direction of the opaque strip. For instance, given a strictly horizontal motion of a vertical strip, we can select these time windows uniformly for all vertical pixel rows, resulting in a more robust and efficient determination of these regions. We define the temporal windows as:

$$T_{\text{occlusion}}(x,y) = [t^{\text{occ}}_{\text{start}}(x,y), t^{\text{occ}}_{\text{end}}(x,y)], \quad (1)$$

$$T_{\text{revelation}}(x,y) = [t^{\text{rev}}_{\text{start}}(x,y), t^{\text{rev}}_{\text{end}}(x,y)], \quad (2)$$

where x and y represent the pixel coordinates.

Intensity Value Assignment Once the temporal windows are established, we count the number of events generated during the occlusion and revelation phases with respect to the polarity of the events. We define the following quantities:

$$N_{\text{occ}}(x,y) = \sum_{t=t^{\text{occ}}_{\text{start}}(x,y)}^{t^{\text{occ}}_{\text{end}}(x,y)} \mathbb{I}\{e_{(x,y,t,-1)}\} - \mathbb{I}\{e_{(x,y,t,1)}\} \quad (3)$$

$$N_{\text{rev}}(x,y) = \sum_{t=t^{\text{rev}}_{\text{start}}(x,y)}^{t^{\text{rev}}_{\text{end}}(x,y)} \mathbb{I}\{e_{(x,y,t,-1)}\} - \mathbb{I}\{e_{(x,y,t,1)}\} \quad (4)$$

After calculating the sums based on the polarity and number of events in the respective phases, we compute the difference to assign a non-normalized intensity value to each pixel. This process can be formalized as:

$$I_{x,y} = N_{\text{occ}}(x,y) - N_{\text{rev}}(x,y), \quad (5)$$

where $I_{x,y} \in \mathbb{Z}$ represents the non-normalized intensity value for pixel (x,y). This integer value encodes the relative brightness of each pixel, with lower values indicating darker regions and higher values indicating brighter regions in the original scene. The final step involves normalizing the intensity values to obtain a standard 8-bit grayscale image. We employ min-max scaling to map the intensity values to the range $[0,255]$:

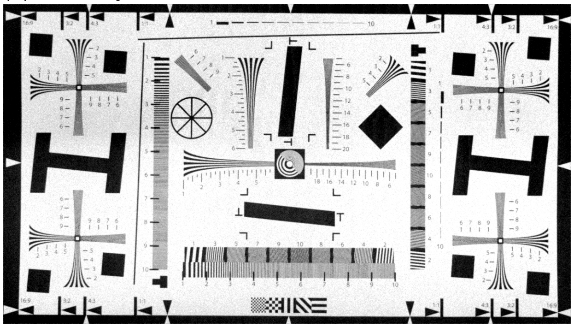
$$I_{normalized}(x,y) = \left\lfloor \frac{I_{x,y} - I_{min}}{I_{max} - I_{min}} \cdot 255 \right\rfloor \quad (6)$$

This normalization produces a grayscale image compatible with standard image processing algorithms and display systems.

Results

We present a proof of concept for our proposed method using two examples: an ISO 12233 chart and a picture of the Karlsruhe Castle. These examples demonstrate our method's capability to reconstruct absolute intensity images from event-based data in both structured test patterns and complex real-world scenes. The experiments were conducted using a Prophesee EVK4 event camera, with a vertical black stripe moving horizontally across the scene to generate the controlled occlusion. These visual results validate the successful generation of intensity information from event streams, including static regions, while detailed quantitative evaluation is reserved for future work.

(a) Intensity reconstruction of the Karlsruhe Castle.



(b) Intensity reconstruction of an ISO 12233 test chart.

Fig. 1: Examples of intensity reconstruction using the proposed method.

References

[1] D. Cazzato and F. Bono, "An application-driven survey on event-based neuromorphic computer vision," Information, vol. 15, no. 8, 2024.

[2] H. Rebecq, R. Ranftl, V. Koltun, and D. Scaramuzza, "Events-to-video: Bringing modern computer vision to event cameras," IEEE Conf. Comput. Vis. Pattern Recog. (CVPR), 2019.

[3] W. Weng, Y. Zhang, and Z. Xiong, "Event-based video reconstruction using transformer," in Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2021.

[4] Y. Bao, L. Sun, Y. Ma, and K. Wang, "Temporal-mapping photography for event cameras," 2024.