

# Simultaneous Mosaicing and Tracking with an Event Camera

Hanme Kim<sup>1</sup>  
hanme.kim@imperial.ac.uk  
Ankur Handa<sup>2</sup>  
ah781@cam.ac.uk  
Ryad Benosman<sup>3</sup>  
ryad.benosman@upmc.fr  
Sio-Hoi Ieng<sup>3</sup>  
sio-hoi.ieng@upmc.fr  
Andrew J. Davison<sup>1</sup>  
a.davison@imperial.ac.uk

<sup>1</sup> Department of Computing,  
Imperial College London,  
London, UK

<sup>2</sup> Department of Engineering,  
University of Cambridge,  
Cambridge, UK

<sup>3</sup> INSERM, U968, Paris, F-75012, France;  
Sorbonne Universités, UPMC Univ Paris 06, UMR\_S 968,  
Institut de la Vision, Paris, F-75012, France;  
CNRS, UMR\_7210, Paris, F-75012, France

An event camera is a silicon retina which outputs not a sequence of video frames like a standard camera, but a stream of asynchronous spikes, each with pixel location, sign and precise timing, indicating when individual pixels record a threshold log intensity change (positive or negative). By encoding only image change, it offers the potential to transmit the information in a standard video but at vastly reduced bitrate, and with huge added advantages of very high dynamic range and temporal resolution.

In this paper, we show for the first time that an event stream from an event camera (e.g. Figure 1(b)), with no additional sensing, can be used to track accurate camera rotation while building a persistent and high quality mosaic of a scene (e.g. Figure 1(d)) which is super-resolution accurate and has high dynamic range; we use the first commercial event camera [1] (Figure 1(a)). Our method involves parallel camera rotation tracking and template reconstruction from estimated gradients (e.g. Figure 1(c)), both operating on an event-by-event basis and based on probabilistic filtering.

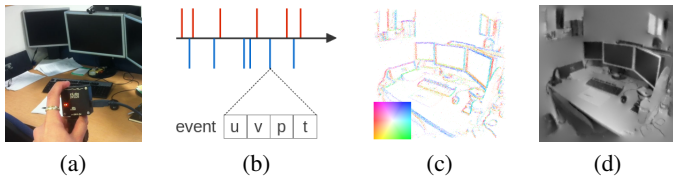


Figure 1: Proposed algorithm: (a) scene and DVS camera; (b) event stream; (c) estimated gradient map; (d) reconstructed intensity map.

In our particle filter based tracking, the posterior density function at time  $t$  is represented by  $N$  particles, each of which is a set consisting of a hypothesis of the current state  $\mathbb{R}_i^{(t)} \in \mathbf{SO}(3)$  and a normalised weight  $w_i^{(t)}$ . As a new event is received, all particles are perturbed based on a constant position motion model; we perturb the current  $\mathfrak{so}(3)$  vector on the tangent plane with Gaussian noise independently in all three axes and reproject it onto the  $\mathbf{SO}(3)$  unit sphere to obtain the corresponding predicted rotation. The noise is the predicted change the current rotation might have undergone since the previous event was generated. The weights of these perturbed particles are then updated through the measurement update step which applies Bayes rule to each particle and normalised subsequently. A measurement given an event, the current state  $\mathbb{R}_i^{(t)}$  and the previous state  $\mathbb{R}_i^{(t-\tau_c)}$ , where  $\tau_c$  is the time elapsed since the previous event at a specific pixel, is a log intensity difference between the corresponding intensity map positions which is to be used to calculate the likelihood for each particle, essentially asking ‘how likely was this event relative to our mosaic given a particular hypothesis of camera pose?’. For the next measurement update and the reconstruction step, a particle mean pose is saved for each pixel.

We now turn to incrementally improving an estimate of the intensity mosaic. This takes two steps; pixel-wise incremental Extended Kalman Filter (EKF) estimation of the log gradient at each template pixel, and interleaved Poisson reconstruction to recover absolute log intensity. Each pixel of the gradient map has an independent gradient estimate and covariance matrix. Now, we want to improve a gradient estimate based on a new incoming event and a tracking result using the pixel-wise EKF. Assuming, based on the rapidity of events, that the gradient  $\mathbf{g}$  in the template and the camera velocity  $\mathbf{v}$  can be considered locally constant, we now say  $(\mathbf{g} \cdot \mathbf{v})\tau_c$  is the amount of log grey level change that has happened since the last event. Therefore, if we have an event camera where log

intensity change  $C$  should trigger an event, the brightness constancy tells us  $(\mathbf{g}^{(t)} \cdot \mathbf{v}^{(t)})\tau_c = \pm C$  which leads to define a measurement  $z^{(t)} = \frac{1}{\tau_c}$  and its measurement model  $h^{(t)} = \frac{\mathbf{g}^{(t)} \cdot \mathbf{v}^{(t)}}{C}$ . The gradient estimate and the uncertainty covariance matrix are then updated using the standard EKF equations. Essentially, each new event which lines up with a particular template pixel improves our gradient estimate in the direction parallel to the camera motion over the scene at that pixel while we learn nothing about the gradient in the direction perpendicular to the motion. Finally, we reconstruct the log intensity of the image whose gradients across the whole image domain are close to the estimated gradients in a least squares sense inspired by [2].

We conducted the spherical mosaicing reconstruction in both indoor and outdoor scenes as shown in Figure 2. Also, we show the potential for reconstructing high resolution and dynamic range scenes from very small camera motion as shown in Figure 3.



Figure 2: Spherical mosaicing for indoor and outdoor scenes. The overlaid boxes represent the field of view of the event camera.



Figure 3: (a) Comparison of a reconstructed high resolution image and a down sampled normal camera image; (b) comparison of a reconstructed high dynamic range image and a normal CCD camera image.

We believe these are breakthrough results, showing how joint sequential and global estimation permits the great benefits of an event camera to be applied to a real problem of mosaicing, and hopefully opening the door to similar approaches in dense 3D reconstruction and many other vision problems.

- [1] P. Lichtsteiner, C. Posch, and T. Delbruck. A  $128 \times 128$  120 dB 15  $\mu\text{s}$  Latency Asynchronous Temporal Contrast Vision Sensor. *IEEE Journal of Solid-State Circuits (JSSC)*, 43(2):566–576, 2008.
- [2] J. Tumblin, A. Agrawal, and R. Raskar. Why I want a Gradient Camera. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2005.