EMVS: Event-based Multi-View Stereo

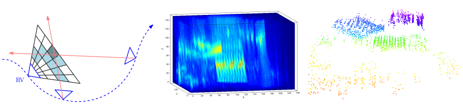
Henri Rebecq
rebecq@ifi.uzh.ch

Guillermo Gallego
guillermo.gallego@ifi.uzh.ch

Davide Scaramuzza
sdavide@ifi.uzh.ch

Robotics and Perception Group
University of Zurich
http://rpg.ifi.uzh.ch
Zurich, Switzerland

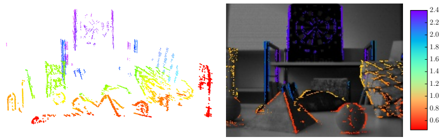
Event cameras are bio-inspired vision sensors that output pixel-level brightness changes instead of standard intensity frames. They offer significant advantages over standard cameras, namely a very high dynamic range, no motion blur, and a latency in the order of microseconds. However, because the output is composed of a sequence of asynchronous events rather than actual intensity images, traditional vision algorithms cannot be applied, so that a paradigm shift is needed.
We introduce the problem of Event-based Multi-View Stereo (EMVS) for event cameras and propose a solution to it. Unlike traditional MVS methods, which address the problem of estimating *dense* 3D structure from a set of known viewpoints, EMVS estimates *semi-dense* 3D structure from an event camera with known trajectory. Our EMVS solution elegantly exploits two inherent properties of an event camera: (*i*) its ability to respond to scene edges—which naturally provide semi-dense geometric information without any pre-processing operation—and (*ii*) the fact that it provides continuous measurements as the sensor moves. Despite its simplicity (it can be implemented in a few lines of code), our algorithm is able to produce accurate, semi-dense depth maps. We successfully validate our method on both synthetic and real data. Our method is computationally very efficient and runs in real-time on a CPU.



(a) (b) (c)

Figure 1: Events are back-projected into rays (a), which are counted in a ray density DSI (b), from which a semi-dense 3D reconstruction of the scene edges is extracted by detecting local maxima (c).

We solve the EMVS problem in a similar way to the Space-Sweep approach for MVS [2], showing how sparsity can be leveraged to estimate 3D structures without the need for explicit data association or photometric information.



(a) Front view. (b) Projection on a frame.

Figure 2: Reconstruction of a desk scene with objects of different shapes that cause occlusions. Depth is colored from red (close) to blue (far).

The Event-based Space-Sweep Method consists of three steps:

- Events are back-projected according to the viewpoint of the camera. Each event produces a viewing ray (Fig. 1(a)).
- Rays are counted on a discretized volume containing the 3D scene, yielding a DSI (Fig. 1(b)) that measures the spatial density of rays.
- Scene points are detected in the regions of high ray-density. The location and value of local maxima of the DSI provide the depth and confidence of reconstructed 3D points, respectively. The most confident values yield a semi-dense depth map or point cloud (Fig. 1(c))

Results. Our algorithm produces accurate 3D reconstructions (relative depth errors $< 5\%$) in the presence of high dynamic range (HDR) illumination and/or high-speed motions that cannot be handled by standard cameras, in spite of the low spatial resolution of the sensor [1] and the high amount of noise typical of event cameras.

Multimedia Material. A supplemental video for this work is available on the authors' webpage: http://rpg.ifi.uzh.ch

[1] C. Brandli et al. A 240x180 130dB 3us latency global shutter spatiotemporal vision sensor. *IEEE J. of Solid-State Circuits*, 2014.

[2] R. T. Collins. A space-sweep approach to true multi-image matching. In *IEEE CVPR*, 1996.