Live Demonstration: Real-time VI-SLAM with High-Resolution Event Camera

Gongyu Yang[1] Qilin Ye[1] Wanjun He[1] Lifeng Zhou[1]
Xinyu Chen[1] Lei Yu[1] Wen Yang[1] Shoushun Chen[2] Wei Li[3]
[1]Whuhan University, Whuhan, China
[2]CelePixel Technology Co.,Ltd
[3]Shanghai Baolong Automotive Corporation
{ygy1996 , ly.wd}@whu.edu.cn

Abstract

Event camera is bio-inspired vision sensors that output pixel-level brightness changes asynchronously. Compared to the conventional frame-based camera, it is with high dynamic range, low latency and high sensitivity, and thus can be exploited in SLAM to tackle the problem of occasions with high-speed camera moving and low-light scenes. In this demo, we implement the visual-inerial SLAM in real time with the recently released event camera, namely, CeleX-V. With the feature of high spatial resolution (1280×800) and low latency ($< 0.5\mu s$), the proposed method can provide frames with abundant textures and high time response, which leads to more stable tracking ability and better performance in SLAM system.

1. Introduction

Event camera has numerous advantages over standard cameras: a latency in the order of microseconds, low power consumption, and a very high dynamic range (130 dB compared to 60 dB of standard cameras) [3]. Consequently, it has great potential to many vision tasks especially for the case with very high-speed camera moving.

On the other hand, Simultaneous Localization and Mapping (SLAM) is such an application that can utilize the features of event camera. Inspired to the existing algorithms for vision-based SLAM, *direct or indirect* method, EVO [6] and Ultimate-SLAM [7] are proposed for event-based SLAM. And it has been shown that at least 85% improvements on localization precisions by introducing event-based sensors in addition to standard frames and IMU [7].

However, existing event-based SLAM algorithms are all based on event sensors with very low spatial resolution. Obviously, it makes insufficient textures when constructing image frames and thus features are not easy to stably tracking. In this demo, we exploit CeleX-V event camera with high spatial resolution (1280×800) and implement VI-SLAM with the fusion of event data and inertial measurements. Main features of CeleX-V are list in Tab. 1

Camera	DAVIS	ATIS	CeleX-V
Resolution	240×180	304×320	1280×800
Latency	$3\mu s$	$3.2\mu s$	$< 0.5\mu s$
Bandwidth	12MHz	30MHz	150MHz

Table 1. Parameters of cameras.

2. VI-SLAM with CeleX-V

The overview of our VI-SLAM pipeline is shown in Fig.1. Some important modules involved are inroduced in the rest of this section.

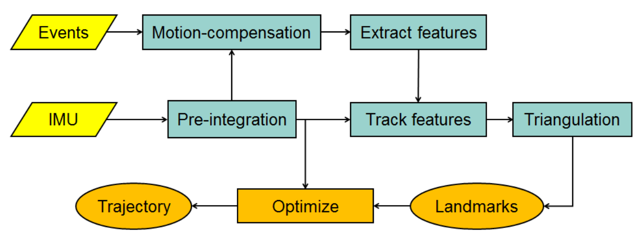

Figure 1. Overview of the pipeline.

2.1. Sensors

Event cameras have independent pixels that fire events at local relative brightness changes in continuous time. The CeleX-V embeds a 1280 × 800 pixels event camera with a 125Hz IMU. Events and IMU measurements are synchronized on hardware.

2.2. Motion compensation

Event frames are synthesized by fusing the events in a spatio-temporal window, and motion blur is generated when the motion is too fast. By using the transformation between the two event frames obtained through integration of the IMU measurements, part of the motion blur can be eliminated to enhance the quality of images. x_i is the pixel location of the last event frame,we can obtain the corrected event position in current event frame $x_i^{'}$.

$$x_i^{'} = K(T(Z(x_i)K^{-1}(x_i))) \tag{1}$$

where K is the event camera projection model obtain from intrinsic calibration, $Z(x_i)$ is the scene depth of pixel x_i ,and T the incremental transformation obtained from integration of the IMU measurements between two frames. [7]

2.3. Feature detection and tracking

Features are detected by the FAST corner detector on a motion-compensated event frame. In detail,we use NMS (Non-maximum suppression) method to filter extracted feature,ensuring the distance between any two features is more than 50 pixels. Due to the imaging characteristics of the event camera, using the traditional LK optical flow method directly will lead to trace missing or trace error. To this end, a 2D homography matrix H can be integrated from the IMU measurements.

$$H = K^{-1}RK \tag{2}$$

where R is the camera rotation by IMU measurements integration.[1] The last frame is warped through an affine warp, which making it more similar to the current frame. At the same time, the Normalized product correlation (NCC) is used for check the results of the features tracking, and the tracking point whose similarity is lower than a threshold is deleted.

2.4. Backend optimization

The parameters of the event frames in the sliding window are optimized by visual-inertial fusion. We base our backend implementation on VINS-MONO[2][5].

3. Demo illustration

This demo contains two main parts. In the first part, we give the parameters of the CeleX-V, compared with two popular event cameras, DAVIS and ATIS. At the same time, we give two video for comparison.The video at the top comes from the data set [4] which is generated by DAVIS. And the video at the bottom is generated by CeleX-V in our office. In the second part, we give two experimental results, which respectively represent the trajectory tracking results of small-scale and large-scale motion(as shown in Figure.2).

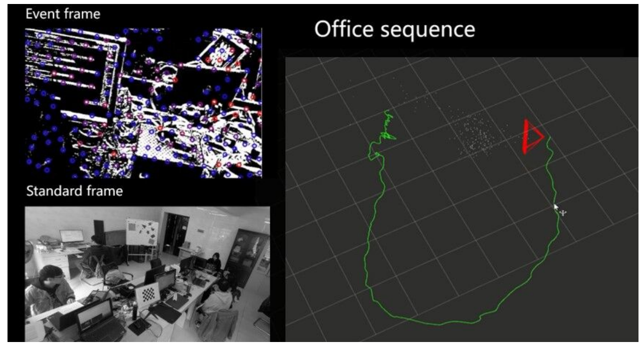

Figure 2. Our motion trajectory. Top Left: Event frame, Bottom Left: Standard frame, Right: motion trajectory.

4. Conclusion

We introduced the first hybrid pipeline that fuses high-resolution events and inertial measurements to yield robust and accurate motion trajectory estimation in SLAM system. Considering the special imaging effect of event camera we have adopted some strategies as mentioned above for more stable performance.

References

[1] Myung Hwangbo, Jun Sik Kim, and Takeo Kanade. Inertial-aided klt feature tracking for a moving camera. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2009.

[2] Stefan Leutenegger, Paul Furgale, Vincent Rabaud, Margarita Chli, Kurt Konolige, and Roland Siegwart. Keyframe-based visual-inertial slam using nonlinear optimization. *Proceedings of Robotis Science and Systems (RSS) 2013*, 2013.

[3] P. Lichtsteiner, C. Posch, and T. Delbruck. A 128 $\times$ 128 120 db 15 μs latency asynchronous temporal contrast vision sensor. *IEEE Journal of Solid-State Circuits*, 43(2):566–576, 2008.

[4] Elias Mueggler, Henri Rebecq, Guillermo Gallego, Tobi Delbruck, and Davide Scaramuzza. The event-camera dataset: Event-based data for pose estimation, visual odometry, and slam. *International Journal of Robotics Research*, 36(49):142–149, 2017.

[5] Tong Qin, Peiliang Li, and Shaojie Shen. Vins-mono: A robust and versatile monocular visual-inertial state estimator. *IEEE Transactions on Robotics*, 34(4):1004–1020, 2018.

[6] Henri Rebecq, Timo Horstschaefer, Guillermo Gallego, and Davide Scaramuzza. Evo: A geometric approach to event-based 6-dof parallel tracking and mapping in real-time. *IEEE Robotics and Automation Letters*, 2(2):593–600, 2017.

[7] Antoni Rosinol Vidal, Henri Rebecq, Timo Horstschaefer, and Davide Scaramuzza. Ultimate slam? combining events, images, and imu for robust visual slam in hdr and high-speed scenarios. *IEEE Robotics and Automation Letters*, 3(2):994–1001, 2018.