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

Gongyu Yang\textsuperscript{1} Qilin Ye\textsuperscript{1} Wanjun He\textsuperscript{1} Lifeng Zhou\textsuperscript{1}
Xinyu Chen\textsuperscript{1} Lei Yu\textsuperscript{1} Wen Yang\textsuperscript{1} Shoushun Chen\textsuperscript{2} Wei Li\textsuperscript{3}
\textsuperscript{1}Wuhan University, Wuhan, China \textsuperscript{2}CelePixel Technology Co., Ltd \textsuperscript{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-inertial SLAM in real time with the recently released event camera, namely, CeleX-V. With the feature of high spatial resolution (1280 \times 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 \times 800) and implement VI-SLAM with the fusion of event data and inertial measurements. Main features of CeleX-V are listed in Tab. 1

2. VI-SLAM with CeleX-V

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

![Figure 1. Overview of the pipeline.](image-url)

### Table 1. Parameters of cameras.

<table>
<thead>
<tr>
<th>Camera</th>
<th>DAVIS</th>
<th>ATIS</th>
<th>CeleX-V</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution</td>
<td>240\times180</td>
<td>304\times320</td>
<td>1280\times800</td>
</tr>
<tr>
<td>Latency</td>
<td>3\mu s</td>
<td>3.2\mu s</td>
<td>&lt; 0.5\mu s</td>
</tr>
<tr>
<td>Bandwidth</td>
<td>12MHz</td>
<td>30MHz</td>
<td>150MHz</td>
</tr>
</tbody>
</table>

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 \times 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 \( x'_i \).

\[
x'_i = K(T(Z(x_i)K^{-1}(x_i)))
\]

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.

\[
H = K^{-1}RK
\]

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 back-end 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).

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


