Event-based Visual Odometry: A Short Tutorial

Dr. Yi Zhou
eeyzhou@ust.hk

HKUST-DJI Joint Lab,
Dept. ECE at HKUST
About Event-based Cameras

Working principle:
◆ Asynchronous and independent pixels

Properties:
◆ High speed, low latency (~ 1 μs)
◆ High dynamic range (140 dB instead of 60 dB)
◆ Ultra-low power (mean: 1mW)
Challenges

Event streams cannot be fed directly to existing methods designed for standard cameras!

Question to Answer

“How to leverage the advantages of event-based cameras to solve a given task by optimally processing the event stream?”
Event-based VO

- **Introduction**
- A Review of Event-based VO
- ESVO System
- Conclusion

➢ About Event-based Cameras
➢ Challenges
➢ Talk’s Outline

- EVO [RAL 17]
- G. Gallego et. al [T-PAMI 18]
- H. Rebecq [BMVC 17]
- Ultimate SLAM [RAL 18]
- H. Kim [ECCV 16]
- ESVO [T-RO 21]
Outline

1. A literature review
2. An introduction to ESVO system
3. Some take-home messages
Review on Event-based Methods

- Event-based Depth Estimation (3D Reconstruction)

- Event-based Camera Pose Estimation
  [RSS 15, TPAMI 18, RAL 17, ICRA 19, IJCNN 11, BMVC 14, ICCP 17, ROBIO 12, ICVS 13, IROS 14]

- Event-based VO Systems
  [ECCV 16, RAL 17]
Event-based Mapping (3D Reconstruction)

- Instantaneous Stereo
  - Two-Step paradigm
    ① Finding eipolar matching
    ② Triangulation
  - SH. Ieng, et. al., Neuromorphic Event-Based Generalized Time-Based Stereovision, Front. Neurosci. 2018

- Temporal Stereo (monocular event camera!)
  - ① Require prior knowledge of the camera’s motion
  - ② Use occurred over a temporal window
  - G. Gallego, et. al., “A unifying contrast maximization framework for event cameras, with applications to motion, depth, and optical flow estimation,” CVPR 2018
Event-based Camera Pose Tracking

Motion and Scene Complexity: Simple -> Complex

### Planar Motion

### 3D Rotation

### 6-DoF Motion
- S. Bryner, et. al, “Event-based, direct camera tracking from a photometric 3D map using nonlinear optimization,” ICRA 2019
Event-based VO Systems


(Project page: https://sites.google.com/view/esvo-project-page/home)
Real-time 3D reconstruction and 6-DoF tracking with an event camera [ECCV 16]

Method Outline

Three interleaved probabilistic filters (EKFs)

- Filter 1: Tracks global 6-DoF camera motion
- Filter 2: Estimates the log intensity gradients in a keyframe image
- Filter 3: Estimates the inverse depths of a keyframe

Video courtesy:
https://www.youtube.com/watch?v=yHLyhdMSw7w&ab_channel=HanmeKim
EVO [RAL 17]

Pipeline Chart

- Event Stream
- Event Frame
- Pose Tracker
- Semi-Dense 3D Map
- 6 DOF Pose
- Create new keyframe?
  - Yes
    - Reset DSI
    - Extract Point Cloud
  - No
    - Update DSI

H. Rebecq, et. al, RAL’ 2017

Video courtesy:
https://www.youtube.com/watch?v=bYqD2qZIjxE&t=8s&ab_channel=UZRoboticsandPerceptionGroup
Core Problem of Event-based VO

VO Problem \[ p(x_t, m_t | z_t) \] Recursive State Estimation

Prediction \[ p(x_t, m_t | z_{t-1}) = \int \int p(x_t, m_t | x_{t-1}, m_{t-1}) \times p(x_{t-1}, m_{t-1} | z_{t-1}) dx_{t-1} dm_{t-1} \]

Correction \[ p(x_t, m_t | z_t) \propto p(z_t | x_t, m_t) \times p(x_t, m_t | z_{t-1}) \]

Implementation Perspective

Tracking Subproblem \[ p(x_t | z_t, m_t^*) \]

Mapping Subproblem \[ p(m_t | z_t, x_t^*) \]

Core Problem of State Estimation from a Methodology Perspective

- Data Association
- Measurement Model
Core Problem of Event-based VO

Filter 1: Tracks 6-DoF camera motion

\[
\begin{align*}
\dot{x} &= \pm C \\
\dot{h}_x &= I_l \left( p_{w}^{(t)} \right) - I_l \left( p_{w}^{(t-\tau_c)} \right),
\end{align*}
\]

where \( I_l \left( p_{w} \right) = (1-a-b) I_l (v_0) + a I_l (v_1) + b I_l (v_2) \).

Filter 2: Pixel-Wise EKF Based Gradient Estimation

\[
\begin{align*}
\dot{z}_g &= \pm \frac{C}{\tau_c} \\
\dot{h}_g &= \left( g(\hat{p}_k) \cdot m \right),
\end{align*}
\]

where \( m = \frac{p_k^{(t)} - p_k^{(t-\tau_c)}}{\tau_c} \).

Filter 3: Pixel-Wise EKF Based Inverse Depth Estimation

\[
\begin{align*}
\dot{z}_{\rho} &= \pm C \\
\dot{h}_{\rho} &= I_l \left( p_{w}^{(t)} \right) - I_l \left( p_{w}^{(t-\tau_c)} \right)
\end{align*}
\]

Core Problem of Event-based VO

How to Make A Difference?

- Can we find a novel X-metric information based on which the event-based data association is established?
- Is the monocular configuration the best choice? (How about stereo?)
- ...

A Brief Literature Review
➢ Core Problem: Data Association on Events
ESVO: Event-based Stereo Visual Odometry
Stereo Event-based Camera Rig
Our System

Stereo Events

Right

Left

Mapping

Tracking

3D Point Cloud and Camera Trajectory
Mapping
Time-Surface Map

\[ T(x, t) \doteq \exp \left( -\frac{t - t_{\text{last}}(x)}{\delta} \right) \]
Geometry

Illustration of the geometry of the proposed mapping method.
Problem Formulation

Objective function

$$\rho^* = \arg \min_{\rho} C(x, \rho, T_{\text{left}}(\cdot, t), T_{\text{right}}(\cdot, t), T_{t-\delta:t})$$

$$C(\cdot, \cdot) \doteq \sum_{x_{1,i} \in W_1, x_{2,i} \in W_2} \|\tau_{\text{left}}^t(x_{1,i}) - \tau_{\text{right}}^t(x_{2,i})\|^2_2$$
Probabilistic Fusion

Propagation

$p(D_{t-M})$ ...

Depth estim.

Propagation

$p(D_{t-1})$

Depth estim.

Propagation

$p(D_t)$

Depth estim.

Fusion

$p(D_t^*)$

Posterior

Fusion Strategy

- Reprojection
- Not Assigned
- Compatible
- Not Compatible
- Assign
- Fuse
- Replace or remain

Fuse $D_{t-1}$ to $D_t$

Events

t - $M$

t - 1

t

time

Histogram

Distribution

(a) simulation_3planes [48].

(b) upenn_indoor_flying1 [44].
Tracking
Exploiting Time Surfaces as Distance Fields

Time Surface $\tau(x)$

Anisotropic Distance Field

$\bar{\tau}(x) = 1 - \tau(x)$

Time Surface Negative $\bar{\tau}(x)$
3D–2D Registration

(a) Inverse depth map in the reference frame. (b) Registration on the negative time-surface map.

Objective Function
\[ \theta^* = \arg \min_{\theta} \sum_{x \in D_{\text{ref}}} \| \pi_{\text{left}}^{F_k}(W(x, \rho; \theta)) \|^2 \]
\[ W(x, \rho; \theta) \doteq \pi_{\text{left}}(T(\pi_{\text{ref}}^{-1}(x, \rho), G(\theta))) \]
\[ \theta \doteq [c_1, c_2, c_3, t_x, t_y, t_z]^T, \quad G(\theta) : \mathbb{R}^6 \rightarrow \text{SE}(3) \]

Forward Compositional LK Method
\[ F(\Delta \theta) \doteq \sum_{x \in D_{\text{ref}}} \| \pi_{\text{left}}^{F_k}(W(W(x, \rho; \Delta \theta); \theta)) \|_2^2 \]
\[ W(x, \rho; \theta) \leftarrow W(x, \rho; \theta) \circ W(x, \rho; \Delta \theta) \]
Objective Function

(a) Objective w.r.t $c_1$. (b) Objective w.r.t $c_2$. (c) Objective w.r.t $c_3$.

(d) Objective w.r.t $t_x$. (e) Objective w.r.t $t_y$. (f) Objective w.r.t $t_z$. 

Evaluation

TABLE II: Parameters of various stereo event-camera rigs used in the experiments.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Cameras</th>
<th>Resolution (pix)</th>
<th>Baseline (cm)</th>
<th>FOV (°)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zhou et al. (ECCV2018)</td>
<td>DAVIS240C</td>
<td>240 × 180</td>
<td>14.7</td>
<td>62.9</td>
</tr>
<tr>
<td>Zhu et al. (RAL2018)</td>
<td>DAVIS346</td>
<td>346 × 260</td>
<td>10.0</td>
<td>74.8</td>
</tr>
<tr>
<td>Mueggler et al. (IJRR2017)</td>
<td>Simulator</td>
<td>346 × 260</td>
<td>10.7</td>
<td>74.0</td>
</tr>
<tr>
<td>Ours</td>
<td>DAVIS346</td>
<td>346 × 260</td>
<td>7.5</td>
<td>66.5</td>
</tr>
</tbody>
</table>

Our stereo event camera rig set up.
Conclusion

Summary

1. Provide a brief literature review on event-based VO and point out the core problem in the design.


Take-Home Messages

1. Trade-off between latency and computation complexity. **Bash v.s. Event-by-Event**

2. Computational resource and power consumption.

Goal: compact and energy-efficient solution.
CVPR 2021
Event-based Vision Workshop