Comparing Representations in Tracking for Event Camera-based SLAM

Jianhao Jiao†, Huaiyang Huang†, Liang Li‡, Zhijian He†, Yilong Zhu†, Ming Liu

†The Hong Kong University of Science and Technology
‡The University of Hong Kong

CVPR 2021 Workshop on Event-based Vision
Motivation: Event Camera-only SLAM

Asynchronous events

3D map
Real-time 6DoF pose
Related Work: Event Camera-only SLAM (needs GPU)

Related Work: Event Camera-only SLAM (CPU-only)

EVO: monocular event camera-based VO

Related Work: Event Camera-only SLAM (CPU-only)

**EVO: monocular event camera-based VO**


**ESVO: stereo event camera-based VO**

Different Event Representations in Tracking

• Tracking Problem Formulation:
  \[ \sum_x \left[ I(W(x; p)) - T(x) \right]^2 \]

• Two event representations:
  1. Event Map (EM) in EVO: asynchronous output; fixed number of events; fast generation
Different Event Representations in Tracking

• Tracking Problem Formulation:

\[ \sum_x [I(W(x; p)) - T(x)]^2 \]

• Two event representations:
  1. Event Map (EM) in EVO: asynchronous output; fixed number of events; fast generation
  2. Time Surface Map (TS) in ESVO: synchronous output; implicit distance field for tracking

\[ I(x, t) \doteq \exp \left( - \frac{t - t_{\text{last}}(x)}{\delta} \right) \]
Different Event Representations in Tracking

• Tracking Problem Formulation:

\[ \sum_x [I(W(x; p)) - T(x)]^2 \]

• Two event representations:
  1. Event Map (EM) in EVO: asynchronous output; fixed number of events; fast generation
  2. Time Surface Map (TS) in ESVO: synchronous output; implicit distance field for tracking

Our interest: how do the different event representations influence the tracking performance?
Combine EM with TS with the degeneracy check

- **Degeneracy factor $\lambda$: the minimum eigenvalue of the Hessian matrix**

Experimental Results

1. *Simulated planar sequences + simulated 6DoF sequences* [1]
2. *RPG handheld 6DoF sequences* [2]
3. *UPenn UAV 6DoF sequences* [3]

---

Experimental Results: Figure Explanation

- Intensity image
- TS
- Inverse depth of the 3D map
- 3D map aligned on Representation (TS or EM)
- Real-time 3D map
- Estimated trajectories w.r.t. GT
Comparison on rpg_bin

Trajectories: Estimated w.r.t GT

TS

$EM_{2000}$ (unreliable)

TS + $EM_{4000}$
Comparison on upenn_indoor_flying3

Trajectories: Estimated w.r.t GT

TS (unreliable)  \quad EM_{4000}  \quad TS + EM_{4000}
## Quantitative Results

<table>
<thead>
<tr>
<th>Sequence</th>
<th>TS</th>
<th>EM&lt;sub&gt;2000&lt;/sub&gt;</th>
<th>EM&lt;sub&gt;3000&lt;/sub&gt;</th>
<th>EM&lt;sub&gt;4000&lt;/sub&gt;</th>
<th>EM&lt;sub&gt;5000&lt;/sub&gt;</th>
<th>TSEM&lt;sub&gt;4000&lt;/sub&gt; (λ&lt;sub&gt;th&lt;/sub&gt; = 31)</th>
</tr>
</thead>
<tbody>
<tr>
<td>simu_office_planar</td>
<td>4.7</td>
<td>4.0</td>
<td><strong>3.9</strong></td>
<td><strong>3.7</strong></td>
<td>4.1</td>
<td>4.9</td>
</tr>
<tr>
<td>simu_poster_planar</td>
<td>4.7</td>
<td><strong>3.7</strong></td>
<td><strong>4.3</strong></td>
<td>4.6</td>
<td>5.0</td>
<td>4.6</td>
</tr>
<tr>
<td>simu_checkerboard_planar</td>
<td>4.2</td>
<td>2.9</td>
<td><strong>2.2</strong></td>
<td><strong>2.3</strong></td>
<td>2.4</td>
<td>4.7</td>
</tr>
<tr>
<td>simu_office_6DoF</td>
<td><strong>9.1</strong></td>
<td>25.3</td>
<td>21.0</td>
<td>16.6</td>
<td>15.8</td>
<td>18.7</td>
</tr>
<tr>
<td>simu_poster_6DoF</td>
<td>18.2</td>
<td><strong>15.4</strong></td>
<td><strong>16.3</strong></td>
<td>16.8</td>
<td>17.4</td>
<td>17.3</td>
</tr>
<tr>
<td>simu_checkerboard_6DoF</td>
<td>23.0</td>
<td>17.0</td>
<td><strong>14.0</strong></td>
<td>15.1</td>
<td><strong>13.4</strong></td>
<td>28.1</td>
</tr>
<tr>
<td>rpg_bin_6DoF</td>
<td><strong>3.4</strong></td>
<td>22.4</td>
<td>16.6</td>
<td>8.0</td>
<td>14.1</td>
<td>3.8</td>
</tr>
<tr>
<td>rpg_box_6DoF</td>
<td><strong>6.5</strong></td>
<td><strong>5.3</strong></td>
<td>17.1</td>
<td>13.7</td>
<td>9.8</td>
<td>7.1</td>
</tr>
<tr>
<td>rpg_desk_6DoF</td>
<td>3.4</td>
<td>2.9</td>
<td>3.3</td>
<td>3.2</td>
<td><strong>2.9</strong></td>
<td>3.8</td>
</tr>
<tr>
<td>rpg_monitor_6DoF</td>
<td>7.2</td>
<td><strong>5.3</strong></td>
<td><strong>5.2</strong></td>
<td>7.4</td>
<td>7.3</td>
<td>7.0</td>
</tr>
<tr>
<td>upenn_indoor_flying1_6DoF</td>
<td>18.5</td>
<td>22.0</td>
<td>16.7</td>
<td><strong>16.0</strong></td>
<td>22.1</td>
<td><strong>14.8</strong></td>
</tr>
<tr>
<td>upenn_indoor_flying3_6DoF</td>
<td>20.9</td>
<td><strong>10.8</strong></td>
<td>11.9</td>
<td>14.0</td>
<td>15.0</td>
<td><strong>10.9</strong></td>
</tr>
</tbody>
</table>
Conclusion

• Extensive comparisons of two representations: event map and time surface map
• Enhanced tracker to make use of their complementary strengths
• Six tracker variations
• Indicate possible ways to improve the state-of-the-art (SOTA) methods.
Thank you!

Email: jijiao@connect.ust.hk
Page: https://gogojjh.github.io
Code: https://github.com/gogojjh/ESVO_extension