Neuronal models for spike-coded depth estimation from event camera data

Martin D. Pham
Department of Computer Science, University of Toronto

Motivation

- **Epipolar Stereo Depth Estimation:**
  - Given a point $X$ from the scene, captured on two rectified sensors as pixels $x_{\text{left}}$ and $x_{\text{right}}$, recover the depth $Z$ for that $X$.
  
- **Dynamic Vision Sensor (DVS):**
  - Every pixel is its own neuron with individual voltages (intensities). An 'ON' event occurs when the log intensity increases by a fixed threshold, an 'OFF' event occurs when it decreases. Each pixel thus codes its photon intensity as a pair of polarized spike-trains in time.
  
- **Spiking Neural Networks (SNN):**
  - A spiking neural network takes spike-trains as inputs, integrates the weighted spikes as a continuous value of voltage in time governed by a neuronal model, and applies a Heaviside activation function to the voltage state.
  - How does neuronal model affect results?

Changing Neuronal Model for Residual Blocks

- **We leave the architecture and learning unchanged except for the neuronal model within the residual bottleneck blocks by replacing the parametric leaky-integrate-and-fire (PLIF) neuronal model with voltage $V$:**

  $V[t] = V[t - 1] + \frac{1}{r} \left( X[t] - (V[t - 1] - V_{\text{reset}}) \right)$

  if $V > V_{\text{threshold}}$ then $V \leftarrow V_{\text{reset}}$

  where $V_{\text{reset}} := \text{reset voltage after spike}$

  $V_{\text{threshold}} := \text{voltage spike threshold}$

  $x[t] := \text{Sigmoid}(u) \times$ a learned parameter

- **Consider the quadratic integrate-and-fire (QIF) model:**

  $V[t] = V[t - 1] + \frac{1}{r} \left( X[t] + \alpha \cdot (V[t - 1] - V_{\text{reset}}(V[t - 1] - V_c)) \right)$

  if $V > V_{\text{threshold}}$ then $V \leftarrow V_{\text{reset}}$

  where $V_{\text{reset}} := \text{reset potential of membrane}$

  $V_c := \text{critical voltage threshold by short current pulse}$

  $\alpha := \text{membrane time constant}$

- **Both dynamical systems are Class I excitatory neuronal models (capable of firing low-frequency spikes when input is weak), but any Class I system describable by smooth ODEs may be transformed into the QIF form by a change of basis of voltage scale and constant current [3].**

Experimental Results

- **Reconstructed raw image**

- **Ground Truth from LIDAR [6]**

  **PLIF depth estimation**

  **QIF depth estimation**

  **The QIF model is able to perform similarly accurate to the PLIF, with a small decrease in the training time per epoch.**

Note that spike activities have been overlayed in order to visualize ON/OFF events.

<table>
<thead>
<tr>
<th></th>
<th>PLIF</th>
<th>QIF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average training time per epoch (seconds)</td>
<td>1.98</td>
<td>1.81</td>
</tr>
<tr>
<td>Testing Set Average Regression Loss</td>
<td>1.285</td>
<td>1.302</td>
</tr>
<tr>
<td>Testing Set Average Depth Error</td>
<td>0.191</td>
<td>0.189</td>
</tr>
</tbody>
</table>

References

[3] Izhikevich et. al, Which model to use for cortical spiking neurons?, IEEE Transactions on Neural Networks, 2004