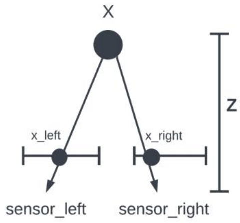
Neuronal models for spike-coded depth estimation from event camera data

Martin D. Pham

Department of Computer Science, University of Toronto

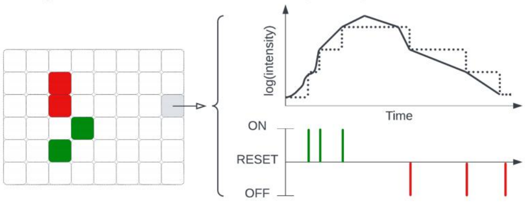
Motivation

- **Epipolar Stereo Depth Estimation:**



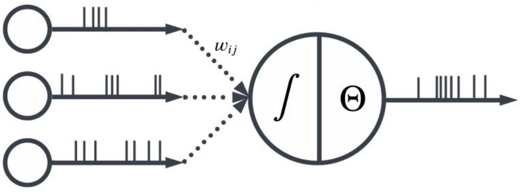
Given a point X from the scene, captured on two rectified sensors as pixels x_left and x_right, recover the depth Z for that X.

- **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):**

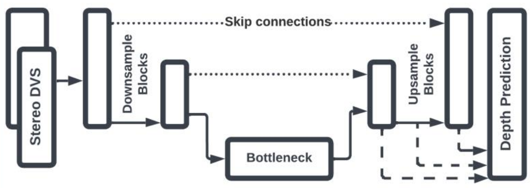


A spiking neural network takes spike-trains as inputs, integrates the weighted spikes as a continuous value of voltage in time governed by a neuronal model, and applies a Heaviside activation function to the voltage state.

- **How does neuronal model effect results?**

Related Work

- StereoSpike [4] is a U-Net (with residual bottleneck) convolutional SNN using surrogate gradient descent training, applying deep learning techniques to stereo DVS data [6].



- But chosen neuronal models lack biological features (i.e. spike coding capacity), are linear

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:

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

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

where

V_{reset} := reset voltage after spike

$X[t]$:= (integrated) input at time t

$V_{threshold}$:= voltage spike threshold

$\frac{1}{\tau} := Sigmoid(w)$, w a learned parameter

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

$$V[t] = V[t-1] + \frac{1}{\tau}\left(X[t] + a_0(V[t-1] - V_{rest})(V[t-1] - V_c)\right)$$

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

where

V_{rest} := resting potential of membrane

τ := membrane time constant

$V_{threshold}$:= neuron threshold voltage

$0 < a_0$:= quadratic term parameter

V_{reset} := neuron reset voltage

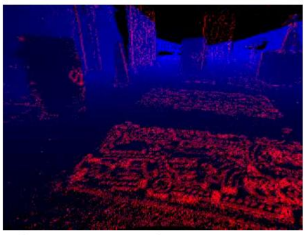
$X[t]$:= (integrated) input at time t

V_c := critical voltage threshold by short current pulse

- 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

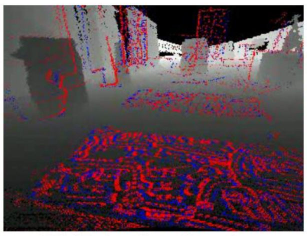
Reconstructed raw image



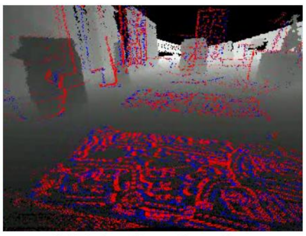
Ground Truth from LIDAR [6]

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

	PLIF	QIF
Average training time per epoch (seconds)	198	181
Testing Set Average Regression Loss	1.285	1.302
Testing Set Average Depth Error	0.191	0.189

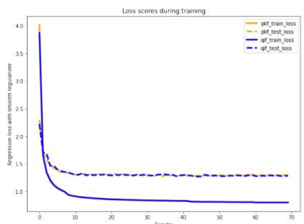


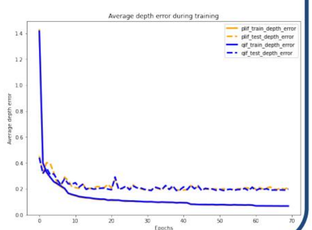
PLIF depth estimation



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.





References

[1] Furmonas et. al, Analytical Review of Event-Based Camera Depth Estimation Methods and Systems, Sensors, 2022

[2] Gallego et. al, Event-Based Vision: A Survey, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022

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

[4] Rançon et. al, StereoSpike: Depth Learning with a Spiking Neural Network, https://arxiv.org/abs/2109.13751, 2021

[5] Steffen et. al, Neuromorphic stereo vision: A survey of bio-inspired sensors and algorithms, Frontiers in Neurorobotics, 2022

[6] Zhu et. al, The Multivehicle Stereo Event Camera Dataset: An Event Camera Dataset for 3D Perception, IEEE Robotics and Automation Letters, 2018