# CSC2529 Project Proposal: Spiking neural networks for depth estimation

Martin D. Pham

November 17, 2022

## 1 Motivation

An important sensing feature of many cognitive systems is vision. Artificial intelligence systems taking inspiration from natural vision systems obtain state-of-the-art performance (e.g. depth estimation [PTB+22]) but lack biological plausibility: biological vision systems involve a heterarchy of physical, chemical, protein and genetic regulatory dynamics obviously unaccounted for in a frame-based model of vision. That is, they are not simulations of an *in vivo* biological process of vision perception. Spiking neural networks (SNN) are an area of research addressing this biological fidelity seeking more accurate models of neurons and action potential [HPGBF18, NCCC22]. SNNs model the dynamics of biological neural networks communicating action potentials via synapses as differential equations representing membrane voltage potential. Neuronal models can be categorized in two ways: pointwise (where the dendrites, somas and axons are treated as occupying the same point in space) or compartmental (where dendrites, somas, axons are treated separately and coupled via state variables in the neuronal system of equations). The connection of these neurons via synaptic models form a network topology. Different topologies and synaptic weights produce different neural representations and processing of an input field over time (e.g. light for vision systems, chemical neuroreceptors for olfactory, sound for auditory). SNNs differ from typical artificial neural networks (e.g. multilayer perceptrons, convolutional neural networks) because there is an assumed temporal dimension to neuronal dynamics [DF21]. They provide a membrane potential account of neural circuitry and therefore may be used to test theories of embedded cognition (i.e. cognition emerges from the interplay of an embrained body and embedded in an environment) such as neuroconstructivism [WMJ+07]. However, the relationship between structure (i.e. topologies) and function (i.e. neural activity) is not well understood in biological brains.

For vision, we may consider a simplified version of the human visual pathway: optical lens controlled by ciliary muscles to photosensitive retina neurons (where light is encoded into neuronal spikes) down optic nerve across chiasmus through lateral geniculate nucleus into striate cortex, as well as feedback throughout. [ZDL+20] presents a retina-like spiking neural network for image reconstruction. [RD20] makes use of known areas of the brain along the human vision pathway to construct a topologically analogous SNN network that models spatial and visual mental imagery. [BSGB+17] similarly organizes its neural architecture based on known anatomical structure, but specifically details the lateral geniculate nucleus, a subregion of the thalamus involved in vision, treating other neural populations as retinal neurons or interneurons. Such large scale network models may be useful for lesioning experimentation where lesions (e.g. reducing number of neurons in a population) at particular points along the pathway are known to produce particular patterned artifacts in the reconstructed image, and therefore may be compared to empirical evidence. This may be seen as a perturbation of model parameters to observe change in model behaviour (e.g. how is the reconstructed perceived visual field affected by reducing the number of neurons within a population group along the visual pathway?) [CDGM22]. Areas (represented by anatomical subnetworks/modules) such as the retina may be lesioned (e.g. some percentage of either cones and/or rods) to observe the degraded performance of image reconstruction.

Artificial photosensitive spiking retinal networks relate to a new kind of camera sensor called event-based cameras or dynamic vision sensors (DVS). See [GDO+22] for an extensive review of event-based vision problems and algorithms. Event cameras code visual information as discrete events (e.g. changes in light intensity) concurrently over a field of independent photosensitive neuronal models and require spiking algorithms to reconstruct and process the image. For this reason, SNNs are comportable with these event-based retinomorphic sensors that more closely capture the photosensitive action potential

dynamics of biological vision systems. However, hardware for event-based cameras is prohibitively expensive and so algorithms trained on preexisting datasets with ground truths is an alternative. The purpose of this work is to investigate the use of SNNs for depth estimation on event-based camera data. A review of event-based camera and spike-aware algorithms for depth estimation can be found in [FLB22].

## 2 Related work

The defining feature of SNNs is the spike encoding of information, and therefore we take selections from the Multivehicle Stereo Event Camera (MVSEC) dataset [ZT+18] which provides event-based data for 3D perception tasks. The work of [RCCM21] makes use of this dataset to do depth estimation using SNN-type neural processing in a UNet-like encoder-decoder architecture. For spiking data and networks, rather than a single pass of information as in a convolution, the network is always "on" and processing changes to the current neural representation as opposed to recording raw values in conventional camera systems. The exploitation of event-based representation is demonstrated in [RCI21] where two streams of event camera data are passed into an SNN architecture composed of two cooperative populations (one for coincidence, one for disparity), producing instantaneous stereo depth perception with real-world stimuli. A useful figure describing spiking stereo networks designed to solve the stereo correspondence problem for depth estimation can be found in Figure 1 of [OIBI17].

To approach biological fidelity, we consider the work of [Kas14] as an example of how to organize and train the neural activity response of neural populations to specific patterns of input stimuli. The work relies on a paradigm of SNN topologies called reservoir computing where a random population of recurrently connected neurons is instantiated and an unsupervised synaptic plasticity-based learning rule is applied before being read out by a supervised trained output layer. For depth estimation, the readout layer would be the depth map.

## 3 Project overview

The goal of the project is to implement reproduce the work of [RCCM21] and extend by comparing against a more biologically plausible SNN architecture that accounts for known anatomy. The main components of the project are

1. Spike encoding/decoding algorithm to produce single image from stereo dataset. The dataset is a series of spike trains across a pixel array and so some processing must be done to recover the scene.

2. Stereo correspondence problem solution using spiking stereo network architecture from existing literature.

3. Implement an SNN architecture analogous to known visual pathway anatomy (retina-chiasm-LGN-cortex) using reservoir computing framework. This network will be designed to solve the stereo correspondence problem to produce a depth map that may be compared with the spiking stereo network.

4. Degrade network parameters to observe/compare performance sensitivities.

## 4 Milestones, timeline & goals

Milestones:

1. Replicate results from [RCCM21] using their code and data: identify what parts of code apply to different required components. 1 week.
   See `https://github.com/urancon/StereoSpike`

2. Implement reservoir computing module for SNNs, use with other network architecture neural populations to simulate optic nerve-chiasmus-thalamic-cortical pathway. Network should also be solving the correspondence problem (in order to be comparable). This may be compared with

the UNet-like structure from the previous milestone under lesion settings. 1.5 weeks.
See https://nba.uth.tmc.edu/neuroscience/m/s2/chapter14.html

3. OPTIONAL (for symposium): Implement conventional (i.e. not event-based) depth estimation on Raspberry Pi hardware, find opportunities to use SNN-based algorithms (will require some kind of spike encoding process). ??? weeks.
See https://stereopi.com/blog/opencv-and-depth-map-stereopi-tutorial

Possible goals :

1. Basic understanding of approaches to stereoscopic depth estimation (i.e. conventional, AI, biologically faithful AI).
2. Develop working knowledge of event-based camera data and spike-aware algorithms for vision.
3. Model some retinal-thalamic-cortical pathway that can be extended with further subregion detailing. Justify configurations of each neural population based on known biophysics/functional neuroanatomy.
4. Evaluate Brian [SBG19] and Nengo [BBH+14] neural simulators for SNN reservoir computing. Compare technical details of implementation requirements and developer experience (Brian is targeted for neuroscientists, Nengo is targeted for computer scientists).

# References


[BBH+14] Trevor Bekolay, James Bergstra, Eric Hunsberger, Travis DeWolf, Terrence C Stewart, Daniel Rasmussen, Xuan Choo, Aaron Russell Voelker, and Chris Eliasmith. Nengo: a python tool for building large-scale functional brain models. *Frontiers in neuroinformatics*, 7:48, 2014.

[BSGB+17] Basabdatta Bhattacharya, Teresa Serrano-Gotarredona, Lorinc Balassa, Akash Bhattacharya, Alan Stokes, Andrew Rowley, Indar Sugiarto, and Steve Furber. A spiking neural network model of the lateral geniculate nucleus on the spinnaker machine. *Frontiers in Neuroscience*, 11:454, 08 2017.

[CDGM22] Nuno Calaim, Florian A Dehmelt, Pedro J Gonçalves, and Christian K Machens. The geometry of robustness in spiking neural networks. *Elife*, 11:e73276, 2022.

[DF21] Simon Davidson and Steve B Furber. Comparison of artificial and spiking neural networks on digital hardware. *Frontiers in Neuroscience*, 15:651141, 2021.

[FLB22] Justas Furmonas, John Liobe, and Vaidotas Barzdenas. Analytical review of event-based camera depth estimation methods and systems. *Sensors*, 22(3), 2022.

[GDO+22] Guillermo Gallego, Tobi Delbrck, Garrick Orchard, Chiara Bartolozzi, Brian Taba, Andrea Censi, Stefan Leutenegger, Andrew J. Davison, Jrg Conradt, Kostas Daniilidis, and Davide Scaramuzza. Event-based vision: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(1):154–180, 2022.

[HPGBF18] Michael Hopkins, Garibaldi Pineda-Garcia, Petruţ A Bogdan, and Steve B Furber. Spiking neural networks for computer vision. *Interface Focus*, 8(4):20180007, 2018.

[Kas14] Nikola K Kasabov. Neucube: A spiking neural network architecture for mapping, learning and understanding of spatio-temporal brain data. *Neural Networks*, 52:62–76, 2014.

[NCCC22] João D Nunes, Marcelo Carvalho, Diogo Carneiro, and Jaime S Cardoso. Spiking neural networks: A survey. *IEEE Access*, 10:60738–60764, 2022.

[OIBI17] Marc Osswald, Sio-Hoi Ieng, Ryad Benosman, and Giacomo Indiveri. A spiking neural network model of 3d perception for event-based neuromorphic stereo vision systems. *Scientific reports*, 7(1):1–12, 2017.

[PTB+22] Matteo Poggi, Fabio Tosi, Konstantinos Batsos, Philippos Mordohai, and Stefano Mattoccia. On the synergies between machine learning and binocular stereo for depth estimation from images: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(9):5314–5334, 2022.

[RCCM21] Ulysse Rançon, Javier Cuadrado-Anibarro, Benoit R. Cottereau, and Timothée Masquelier. Stereospike: Depth learning with a spiking neural network. *CoRR*, abs/2109.13751, 2021.

[RCI21] Nicoletta Risi, Enrico Calabrese, and Giacomo Indiveri. Instantaneous stereo depth estimation of real-world stimuli with a neuromorphic stereo-vision setup. In *2021 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 1–5, 2021.

[RD20] Sean Riley and Jim Davies. A spiking neural network model of spatial and visual mental imagery. *Cognitive Neurodynamics*, 14:239–251, 04 2020.

[SBG19] Marcel Stimberg, Romain Brette, and Dan FM Goodman. Brian 2, an intuitive and efficient neural simulator. *Elife*, 8:e47314, 2019.

[WMJ+07] Gert Westermann, Denis Mareschal, Mark H Johnson, Sylvain Sirois, Michael W Spratling, and Michael SC Thomas. Neuroconstructivism. *Developmental science*, 10(1):75–83, 2007.

[ZDL+20] Lin Zhu, Siwei Dong, Jianing Li, Tiejun Huang, and Yonghong Tian. Retina-like visual image reconstruction via spiking neural model. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1438–1446, 2020.

[ZT+18] Alex Zihao Zhu, Dinesh Thakur, Tolga zaslan, Bernd Pfrommer, Vijay Kumar, and Kostas Daniilidis. The multivehicle stereo event camera dataset: An event camera dataset for 3d perception. *IEEE Robotics and Automation Letters*, 3(3):2032–2039, 2018.