UTS UNIVERSITY OF TECHNOLOGY SYDNEY

DOTIE: Detecting Objects through Temporal Isolation of Events using a Spiking Architecture

Nagaraj 2022

This week…

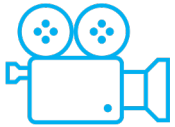
No maths, No code, Yes Engineering.

Just talk about the usefulness of SNN in object detection for Event Cameras.

The background story

Vision-based Autonomous navigation system

The answer →



Biologically inspired Event Camera

utilises →

Temporal Information

determines →

Moving Objects

- Fast
- Accurate
- Energy efficient

- Offers speed
- Energy efficient
- Robustness to lighting condition

01

Introduction and Summary

The Proposed Algorithm

A Spiking Neural Network

- Able to generates Spiking events

Separate events based on →

Speed of moving objects

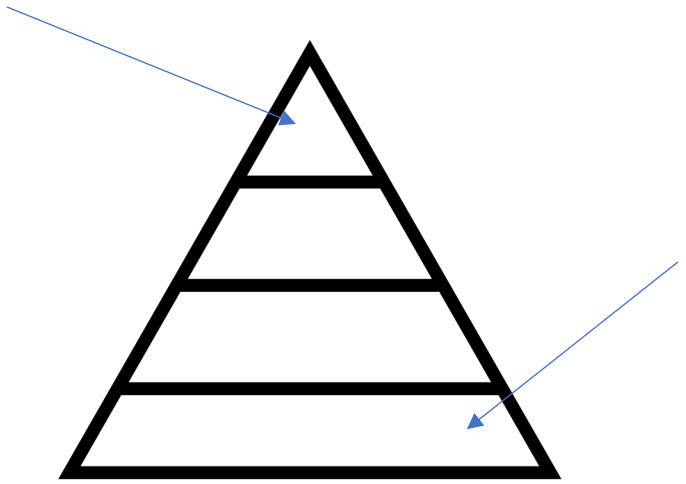
To determine →

Object boundaries

- Asynchronous
- Robust to camera noise
- Less energy overheads
- Surpass current event-based algorithm

Autonomous Navigation systems is an important research area

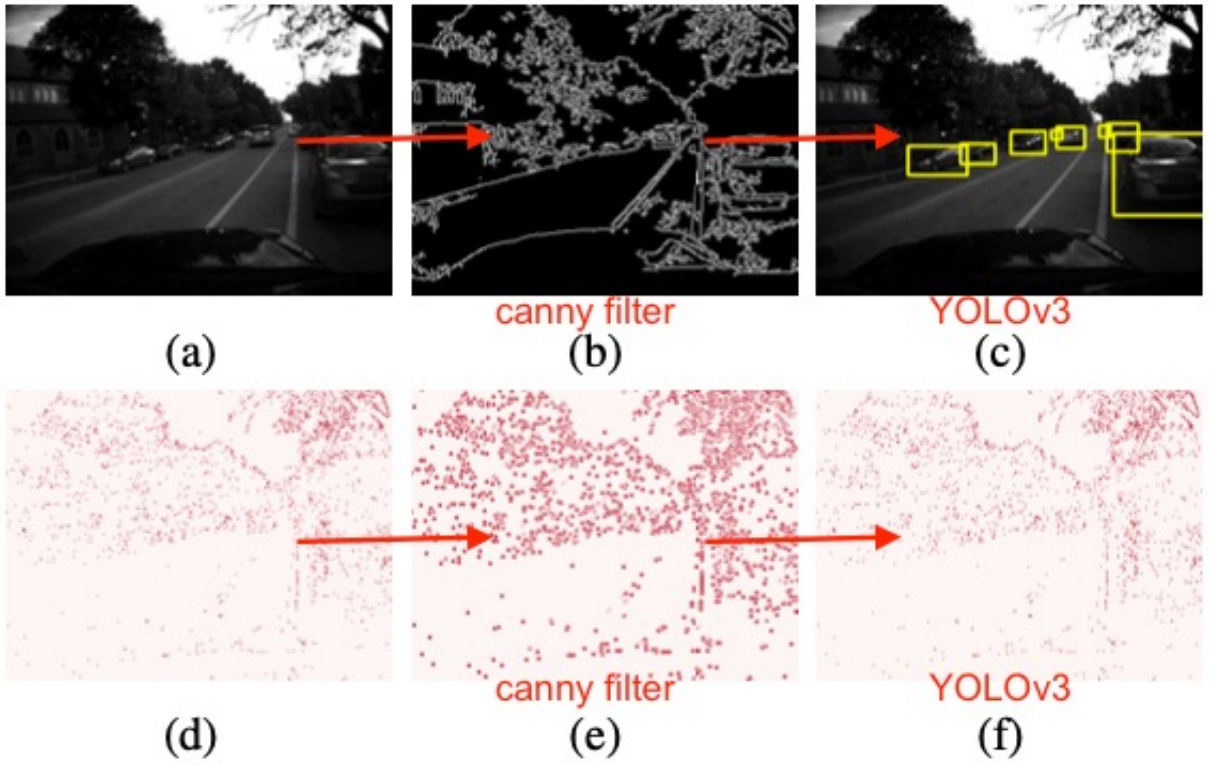
Level 5 being the highest (requires no human operator)



This paper works for level 1 and above (that is, the ability to detect and avoid obstacles)

Society of Automation Engineers (SAE) identified 6 levels of automation (0..5)

Event Camera Vs Frame Camera



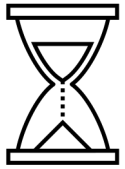
- Capture photometric features at fixed interval.
- Canny Filter and YOLOv3 perform for frame camera but fail for Event cameras.

- Event Camera captures light intensities for each pixel.
- Higher frequency
- Asynchronous.
- Operates in high speed

Temporal and Spatial analysis

1

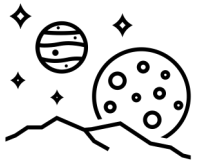
Leaky and Integrated Fire Model



Temporal Characteristics

2

Clustering



Spatial Characteristics

- The use of temporal information can differentiate objects via LIF.
- The output events (firing) generated when input events occur at a rate of higher frequency.
- Similar object have similar moving speed.

- Pixel of the same object are close together.
- Clustering in unsupervised. No labelling is required.
- Irrelevant to scenes.

02

Related work and Methodology

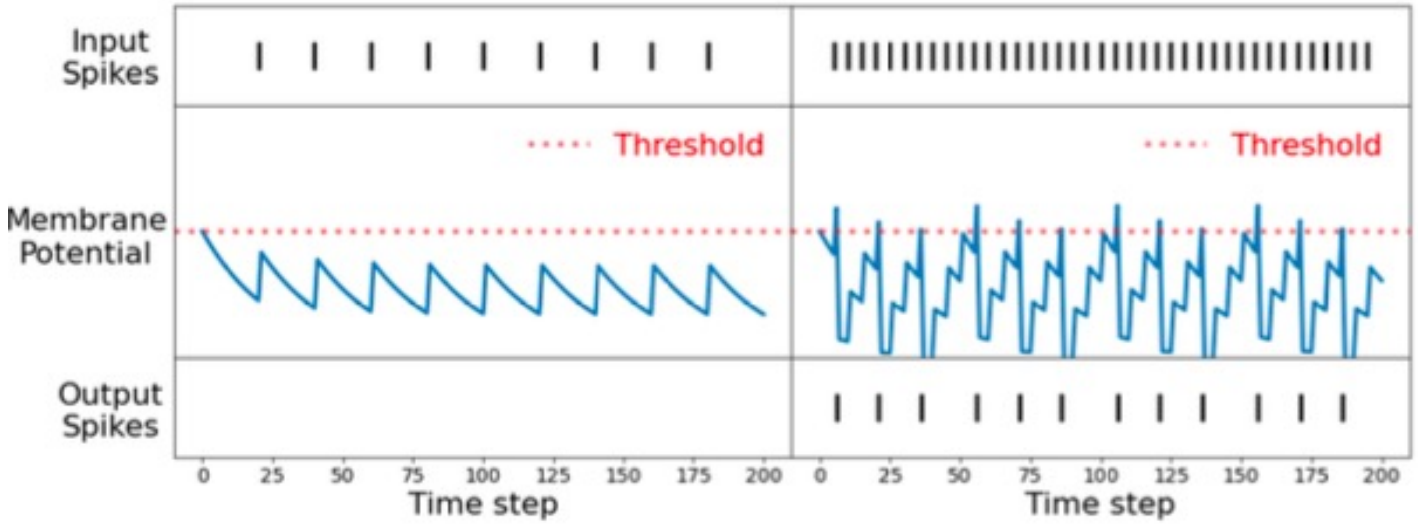
Related works

- There exists plenty algorithms for object detection for event camera such as [17], [19], [20]
  - Have too much computational Overheads
  - Are not asynchronous in nature.

- The proposed algo avoid Deep SNN, but a single layer SNN.
  - No accuracy degradation.

- The proposed algo uses clustering to generate boundaries.

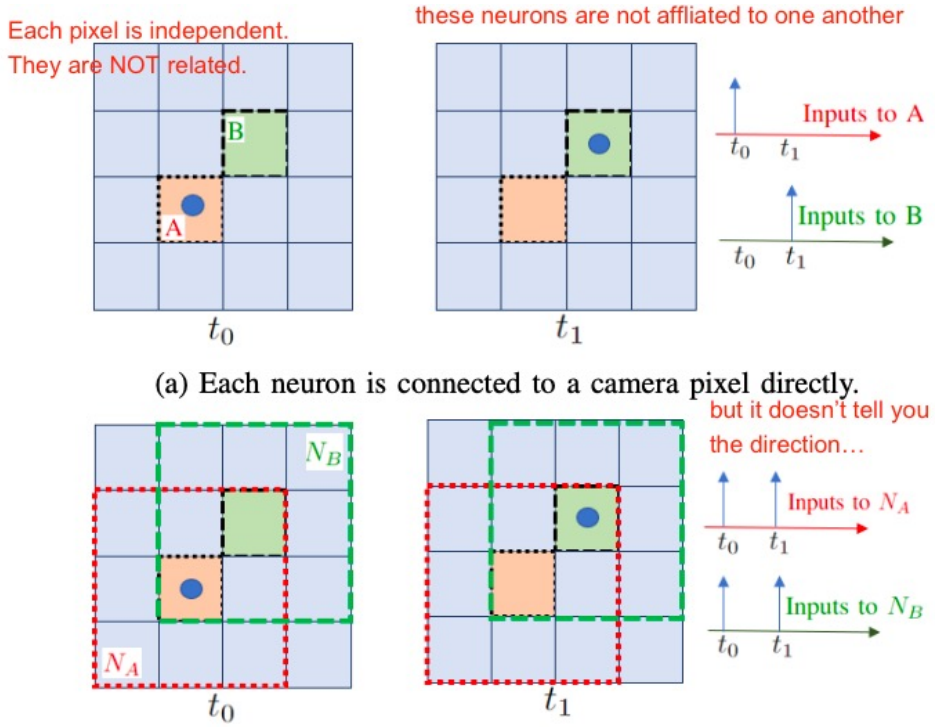
Methodology: Spikes to find objects



Nothing we didn't already know

- To goal using spike firing is to isolate objects.
- Faster moving objects generate more spikes

A spiking neuron is NOT enough



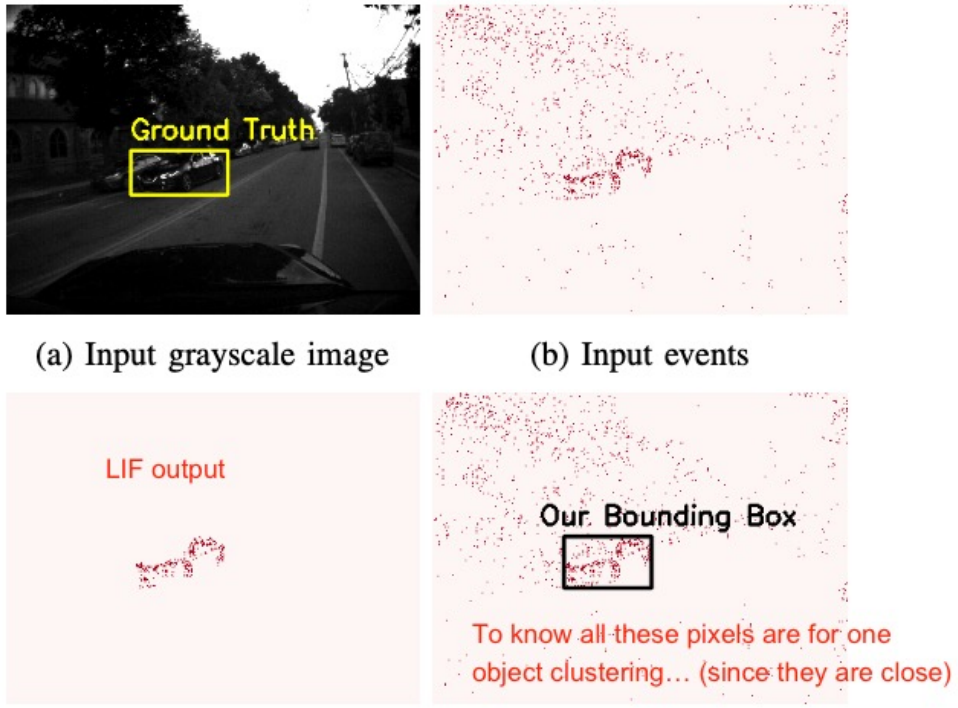
(a) Each neuron is connected to a camera pixel directly.

- The argument is that a pixel connected to a spiking neuron may only measure frequency, but Cannot measure speed
- A spiking neuron connected to a neighbour can identify fast moving objects.

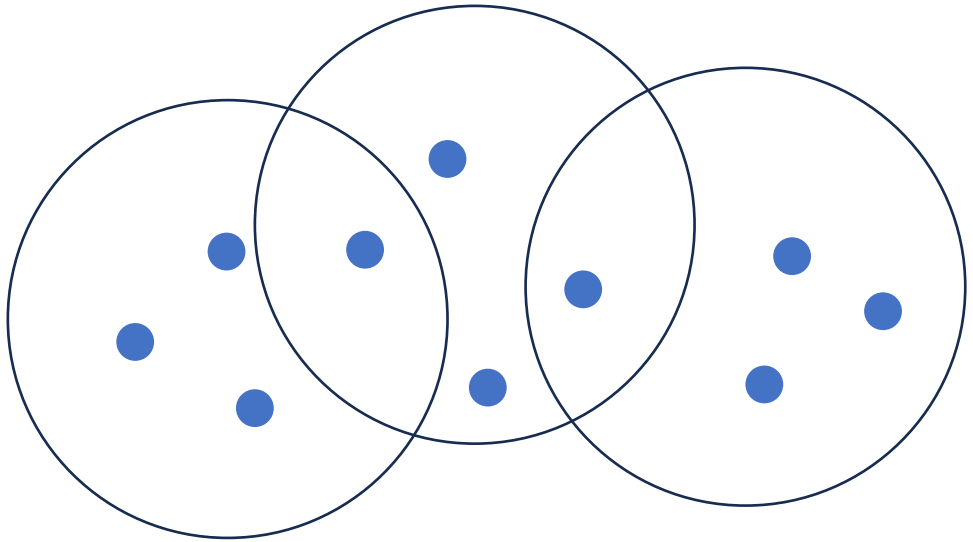
What happen if object didn't move?

- Normalised weights for each neighbourhood adds up to 1.
- Central pixel has a weight of 0.2

Clustering to “box” objects



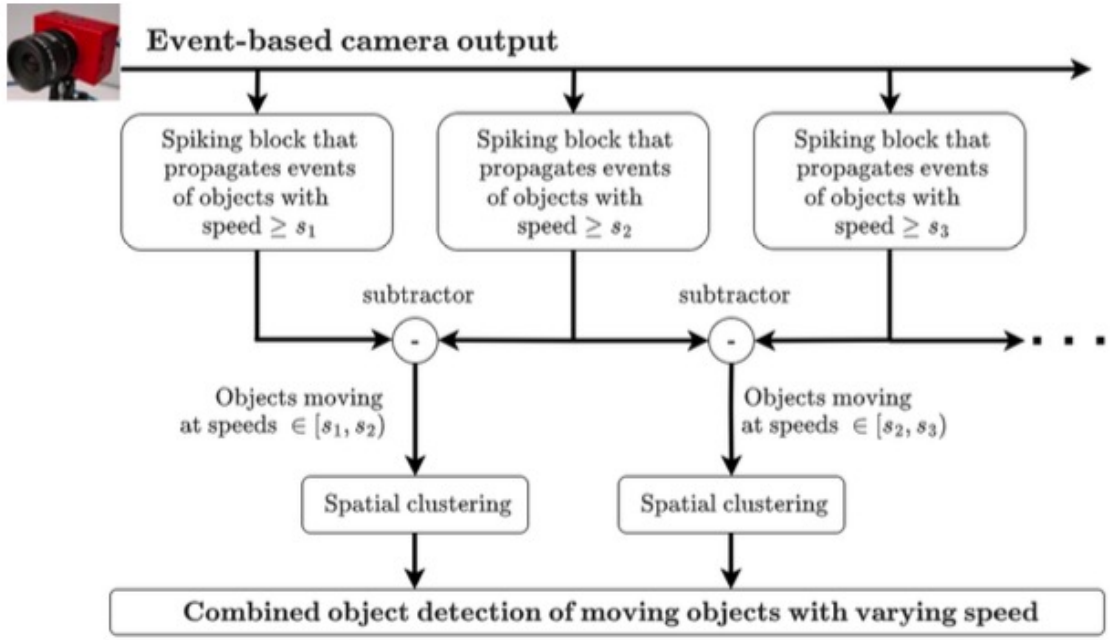
- An object would be made up of pixels close to each other.



Spatial Clustering

- We don’t care the size of the clusters.

Overall Architecture

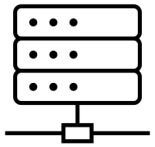


- Identify objects moving a different speed.
- Later each moving object to detect boundaries

03

Experiments and Results

Experiment

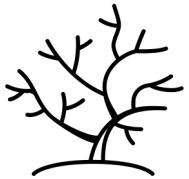


Data (MVSEC)

- Author claims that most algorithm have not been evaluated using any benchmark dataset since there aren't enough Event Camera output for reference. Author managed to use **MVSEC**.

SNN

- It has been successful. The background noise does not affect its performance since **SNN filters out noise.**



YOLOv3

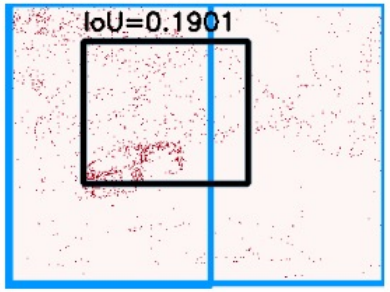
- Author used **YOLOv3** to generate box Boundaries as ground truth and compare their performance according to Interaction Over Union (IoU)

IoU

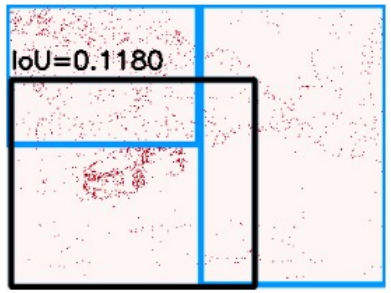
(a) Grayscale frame

(b) Our algorithm

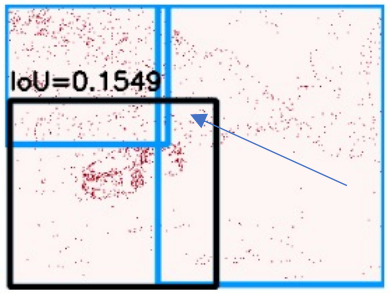
(c) Meanshift [39]



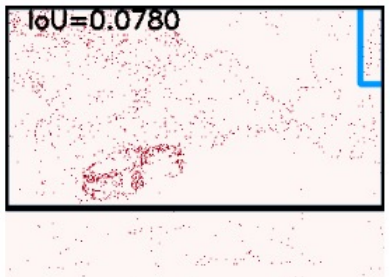
(d) GMM [19]



(e) K-Means [36]



(f) GSCE [40]



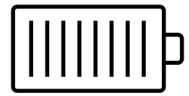
- IoU >= 0.5 (True Positive): Matching boundaries groud truth.
- IoU < 0.5 (False Negative): Cover some ground truth and beyond
- (False Positive): does NOT correspond to ground truth at all.

- The algorithm seems to respond very well.
- When multiple boundaries are generated. The cluster with the highest IoU is selected.

Final notes

- Energy and latency efficient.

- Spiking neural accumulate (AC) inputs. Unlike ANN which multiplies and accumulates (MAC) for every neuron.



- Energy consumption is estimated Eq. 2

The End

Any questions are welcome.