PROPHESEE METAVISION FOR MACHINES

Intelligent Signal Systems Laboratory

Introduction To Event Detection Cameras

Keigo Hirakawa
University of Dayton
Professor
khirakawa1@udayton.edu

Davide Migliore
Prophesee
Technical Business Developer
Senior Computer Vision Engineer
dmigliore@prophesee.ai

ICCV 2021 MONTREAL, CANADA

IEEE COMPUTER SOCIETY

Brief Introduction

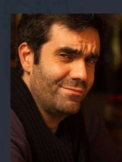
Instructor – Keigo Hirakawa

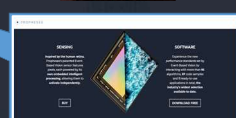
- University of Dayton, Professor
  - khirakawa1@udayton.edu
- Research interests – Computational Imaging
  - Event detection camera
  - Digital camera processing pipeline
  - Multispectral/hyperspectral imaging
  - Image/video forensics
  - Low-level computer vision
  - Statistical signal processing
- Intelligent Signal Systems Laboratory
  - issl.udayton.edu

Brief Introduction

Instructor – Davide Migliore

- Prophesee
  - Technical Business Developer
  - Director Sales & Marketing North America
  - dmigliore@prophesee.ai
- Research interests
  - Event Based Computer Vision
  - 3D reconstruction, SLAM, VIO
  - Neuromorphic processing





Welcome to the Short Course!

Goals

- Understand the key benefits and the characteristics of event detection cameras.
- Be able to interpret the event data and import event data into your work environment.
- Have a working knowledge of how event detection sensor data is processed.

Intended Audience

- Researchers (students/practitioners) with working knowledge of image processing and computer vision.

PROPHESEE METAVISION FOR MACHINES

Intelligent Signal Systems Laboratory

Disclaimers…

We are just a messenger…

- The event detection camera research is rich and diverse.
- This short course is neither complete nor comprehensive. Nor is it a survey.
- This short course is introductory. Every example is selected because of the insight we gain from it.
- Our own works appear as part of short course where it's contextually appropriate. It's not to self-promote.

The hurdle of entry is high!

- Event detection cameras are not the easiest thing to work with.
  <= *We hope this short course helps!*
- Many computational tools are still under development. (And always changing!)

PROPHESEE METAVISION FOR MACHINES

Intelligent Signal Systems Laboratory

Disclaimers…

We will focus more on the computational aspects of event detection camera.

- Will only briefly touch on the biological motivation for silicon retina.

We don't want it to turn into a short course about CNN architectures.

- There are already many resources about CNN.
- We want to focus on aspects that are specific to event detection cameras.

PROPHESEE METAVISION FOR MACHINES

Intelligent Signal Systems Laboratory

Outline of Tutorial

- Introduction
- Event Representation
- Computational Imaging
- Computer Vision
- Applications
- *Quick Start Guide*
- Conclusion and Q&A

PROPHESEE METAVISION FOR MACHINES

Intelligent Signal Systems Laboratory

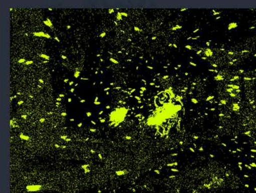
Introduction

What is a event detection camera?

Event @#&$*! Camera

- Event Camera
- Event Based Camera
- Event Based Vision Camera
- Event Detection Camera
- Neuromorphic Camera
- Contrast Detection Camera
- Silicon Retina
- Spiking Cameras

etc…



Prophesee VGA camera on bees

Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

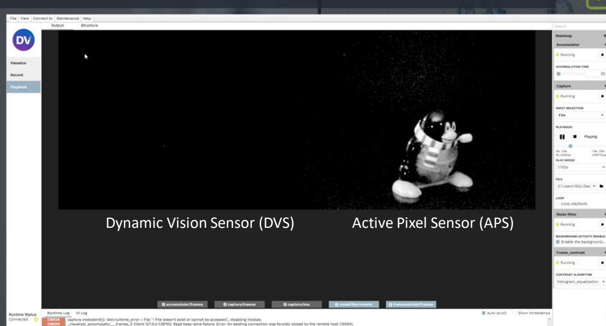
Event @#&$*! Pixel Sensor

- Event-Based Vision Sensor (EVS)
- Dynamic Vision Sensor (DVS)
- Asynchronous Time-Based Image Sensor (ATIS)
- Silicon Eye Vision Sensor (SEVS)

Prophesee Event-Based Vision Sensor (EVS)

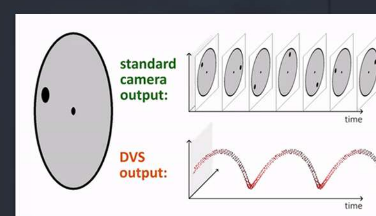
Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

Event Detection Camera



Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

Event Detection Camera

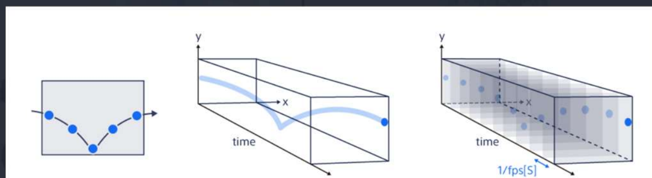


APS + Events

http://etinow.me/134
Almatrafi, M., Baldwin, R., Aizawa, K., Hirakawa, K., “Distance Surface for Event-Based Optical Flow,” IEEE Trans. Pattern Anal. Machine Intell. (TPAMI), 2020.

Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

Event Detection Camera



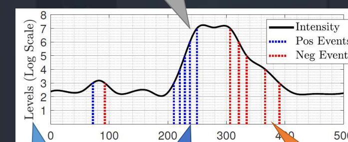
Trajectory of the ball
Event-based sensor
Frame-based sensor

13 https://www.sony-semicon.co.jp/e/products/IS/industry/product/evs.html#evs_vs_fvs

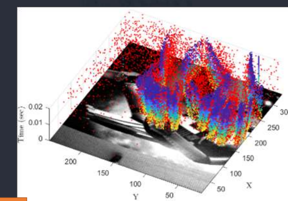
Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

Event Detection Camera

events generated at threshold



log intensity
positive polarity when increasing
negative polarity when decreasing



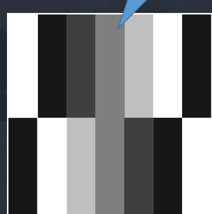
APS + Events

14 Almatrafi, M., Baldwin, R., Aizawa, K., Hirakawa, K., "Distance Surface for Event-Based Optical Flow," IEEE Trans. Pattern Anal. Machine Intell. (TPAMI), 2020.

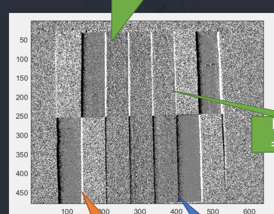
Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

Event Detection Camera

linear steps



low intensity => more events



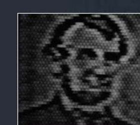
high intensity => less events
positive events
negative events

15

Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

Brief History of Event Detection Camera

- 1992 Mahowald
  - First "silicon retina"
- 2001 Boahen
  - Asynchronous spatial-temporal contrast
- 2003 Ruedi
  - Spatial contrast and direction. Frame based.
- 2005 Malik
  - Temporal frame-difference detection
- 2006 Lichtsteiner
  - Dynamic Vision Sensor (DVS), asynchronous temporal contrast
- 2008 Bassari
  - Binary spatial and temporal contrast
- 2010 Posch
  - Asynchronous time-based image sensor (ATIS)



1992 Mahowald

2006 Lichtsteiner

M. Mahowald, "VLSI analogs of neuronal visual processing: A synthesis of form and function," Ph.D. dissertation, California Institute of Technology, Pasadena, California, May 1992.

16 P. Lichtsteiner, C. Posch and T. Delbruck, "A 128 x 128 120 dB 15us Latency Asynchronous Temporal Contrast Vision Sensor," in *IEEE Journal of Solid-State Circuits*, vol. 43, no. 2, pp. 566-576, 2008.

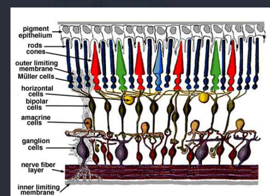
Retina Inspired Design

Photoreceptors
- Detect light, convert light into pulses
- Connect to bipolar cells via horizontal cells

Bipolar Cells, Amacrine Cells
- Lateral inhibition, feedback, long-range connections, normalization

Ganglion Cells
- Connected to brain by optic nerve
- Interpret incoming intensity patterns
- Patterns of spikes generated by ganglion cells are received by the brain

Human Retinal System

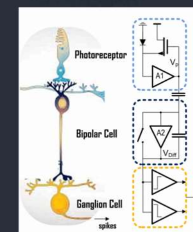


 https://embryology.med.unsw.edu.au/embryology

Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

Retina Inspired Design

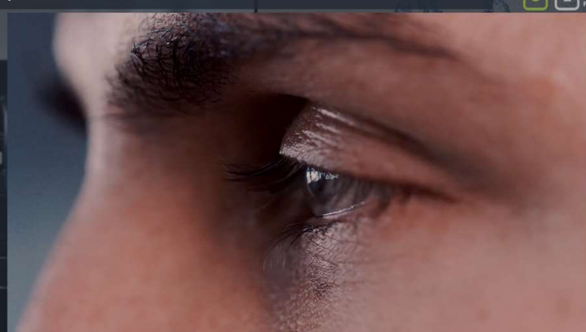
Event-Based Vision Sensor (EVS)
- EVS functions like a simplified Magno-cellucar transient pathway
  - Photoreceptive cells = logarithmic photodiode
  - Bipolar cell = differencing circuit
  - Ganglion cell = comparators



 Posch, C., Bio-inspired vision, J. of Instrumentation, 7 C01054, 2012. Bio-inspired explanation of the DVS and the ATIS.

Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

Biologically Inspired Visual Perception



Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

Event Detection Camera Today

Real-Time High Speed Imaging
- High speed imaging without high computational burden.
- Applications: Robotics, controls, tracking, etc…

Automotive Imaging
- Low latency, high dynamic range.
- Compliment slow APS sensors, environment noise mitigation.
- Applications: SLAM, navigation, obstacle avoidance, stereo vision, etc…

Object Recognition
- Recognition where motion is the defining characteristics
- Applications: Activity recognition, human pose estimation, robotics, etc…

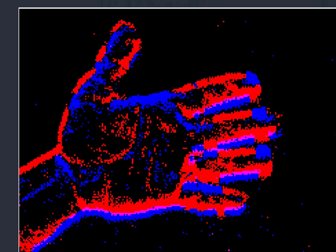
Low Power Sensing

Introduction

Event-Based Vision Sensor – Pixel Architecture

DAVIS Camera

APS (40hz max, synchronous)



DVS (800,000hz max, asynchronous)

 Almatrafi, M. M., Hirakawa, K., "DAVIS Camera Optical Flow," IEEE Trans. Comput. Imag. (TCI), 6:396-407, 2019.

Active Pixel Sensor (APS)

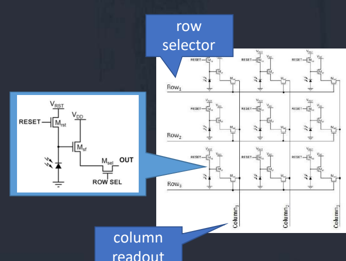
"Exposure"

- RESET=off, ROW SEL=on
- Current via Msf/Msel proportional to light intensity
- Columns read out in parallel
- Rows read out serially

Characteristics

- Measures intensity
- Synchronous ≡ "Frame"
- Fixed/controlled integration time

Conventional 3 transistor APS



 https://meroli.web.cern.ch/lecture_activepixelsensors.html

Active Pixel Sensor (APS)

Global Shutter

Rolling Shutter

 www.thinklucid.com

Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

PROPHESEE METAVISION FOR MACHINES

Intelligent Signal Systems Laboratory

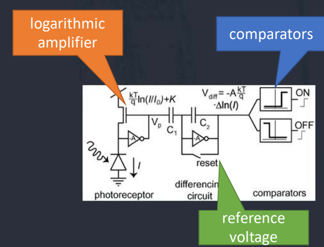
Event-Based Vision Sensor (EVS)

Logarithmic Photoreceptor
- Current/voltage proportional to log-intensity

Differencing Circuit
- Subtracts offset (reference voltage)

Comparators (1 bit A/D)
- ON event: Vdiff > threshold
- OFF event: Vdiff < -threshold

EVS circuit

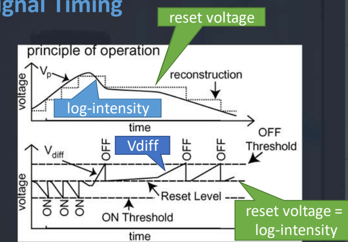


25 P. Lichtsteiner, C. Posch and T. Delbruck, "A 128 x 128 120 dB 15us Latency Asynchronous Temporal Contrast Vision Sensor," in *IEEE Journal of Solid-State Circuits*, vol. 43, no. 2, pp. 566-576, Feb. 2008.

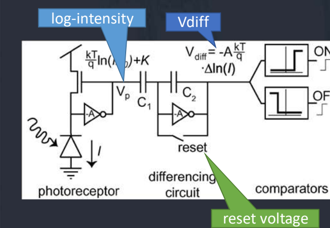
Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

Event Detection Forward Model

Signal Timing



EVS circuit



26 P. Lichtsteiner, C. Posch and T. Delbruck, "A 128 x 128 120 dB 15us Latency Asynchronous Temporal Contrast Vision Sensor," in *IEEE Journal of Solid-State Circuits*, vol. 43, no. 2, pp. 566-576, Feb. 2008.

Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

Event-Based Vision Sensor (EVS)

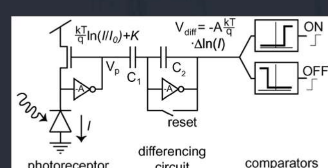
Complexity
- 26 transistors
- 3 capacitors

Power Consumption
- 10-24mW

Dynamic Range
- >120dB

Latency
- <15µs

EVS circuit



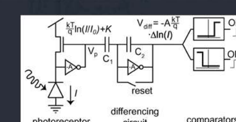
27 P. Lichtsteiner, C. Posch and T. Delbruck, "A 128 x 128 120 dB 15us Latency Asynchronous Temporal Contrast Vision Sensor," in *IEEE Journal of Solid-State Circuits*, vol. 43, no. 2, pp. 566-576, Feb. 2008.

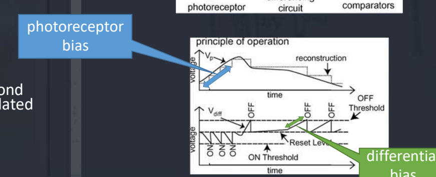
Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

Camera Parameters

Photoreceptor Bias
- Limits the current from the logarithmic photodiode
- Too high – responds to higher frequency electronic noise and produces lots of noise in dark scenes
- Too low – cannot see fast oscillations in illumination

Differential Bias
- Determines the speed at which the second stage adjusts to a change in the light-related signal
- Too fast – too many event per edge
- Too slow – only one event per edge





28 P. Lichtsteiner, C. Posch and T. Delbruck, "A 128 x 128 120 dB 15us Latency Asynchronous Temporal Contrast Vision Sensor," in *IEEE Journal of Solid-State Circuits*, vol. 43, no. 2, pp. 566-576, Feb. 2008.
https://docs.google.com/document/d/1qDReOT5-1ns-3fpOevm2fXw6idh0Yo0GBSu2Aey2SKg/edit

Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

Camera Parameters

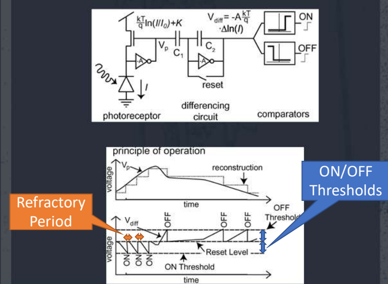
Thresholds For ON and OFF Events

- Change in illumination necessary to generate an event (i.e. contrast sensitivity)
- Too small – lots of noise
- Too large – only very large contrast changes generate events

Refractory Period

- Period after an event that the pixel does not respond to changes
- Too small – too much events generated
- Too large – attenuates high frequency signals

Contrast sensitivity ≈ (ON threshold)/(diff bias)



29 P. Lichtsteiner, C. Posch and T. Delbruck, "A 128 x 128 120 dB 15us Latency Asynchronous Temporal Contrast Vision Sensor," in *IEEE Journal of Solid-State Circuits*, vol. 43, no. 2, pp. 566-576, Feb. 2008.
https://docs.google.com/document/d/1qDReOT5-1ns-3fpOevm2fXw6idh0Yo0GBSu2Aey2SKg/edit

Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

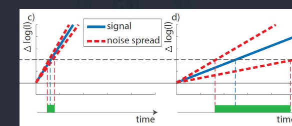
Time Stamp Property

Time Stamp

- Time since the "camera reset". ***Not*** since the beginning of recording the file.
- Time stamp can overflow and roll over!
- It is a good practice to reset the camera before taking a new sequence.

Time Stamp Ambiguity

- Event timing ambiguous if differential bias is set too slow

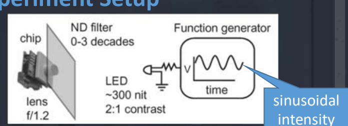


30

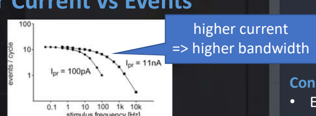
Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

Event Transfer Function

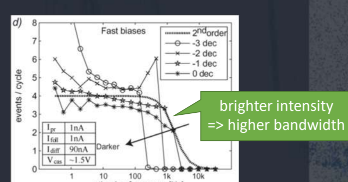
Experiment Setup



Amplifier Current vs Events



Intensity vs Events



Conclusion:

- Events cannot be generated on repeating edges.
- But latency is still low.

31 P. Lichtsteiner, C. Posch and T. Delbruck, "A 128 x 128 120 dB 15us Latency Asynchronous Temporal Contrast Vision Sensor," in *IEEE Journal of Solid-State Circuits*, vol. 43, no. 2, pp. 566-576, Feb. 2008.

Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

Address-Event Representation (AER)

Address Encoder (18 bits)

- Row events → 9 bit address
- Column events → 9 bit address

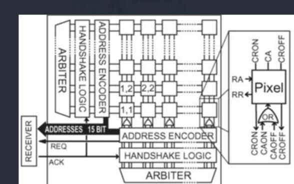
Event Polarity (1 bit)

- 0=OFF, 1=ON

Time-Stamp (32 bits)

- System clock counter
- 100kHz-1MHz resolution

AER circuit



32 P. Lichtsteiner, C. Posch and T. Delbruck, "A 128 x 128 120 dB 15us Latency Asynchronous Temporal Contrast Vision Sensor," in *IEEE Journal of Solid-State Circuits*, vol. 43, no. 2, pp. 566-576, Feb. 2008.

Introduction

Sensor Technology and Roadmap

Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

Camera Suppliers

Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

Camera Comparisons

	Supplier	iniVation			Prophesee				Samsung			CelePixel		Insightness
	Camera model	DVS128	DAVIS240	DAVIS346	ATIS	Gen3 CD	Gen3 ATIS	Gen 4 CD	DVS-Gen2	DVS-Gen3	DVS-Gen4	CeleX-IV	CeleX-V	Rino 3
Sensor specifications	Year, Reference	2008 [2]	2014 [4]	2017	2011 [3]	2017 [67]	2017 [67]	2020 [68]	2017 [5]	2018 [69]	2020 [39]	2017 [70]	2019 [71]	2018 [72]
	Resolution (pixels)	128 × 128	240 × 180	346 × 260	304 × 240	640 × 480	480 × 360	1280 × 720	640 × 480	640 × 480	1280 × 960	768 × 640	1280 × 800	320 × 262
	Latency (μs)	12μs @ 1klux	12μs @ 1klux	20	3	40 - 200	40 - 200	20 - 150	65 - 410	50	150	10	8	125μs @ 10lux
	Dynamic range (dB)	120	120	120	143	> 120	> 120	> 124	90	90	100	90	120	> 100
	Min. contrast sensitivity (%)	17	11	14.3 - 22.5	13	12	12	11	9	15	20	30	10	15
	Power consumption (mW)	23	5 - 14	10 - 170	50 - 175	36 - 95	25 - 87	32 - 84	27 - 50	40	130	-	400	20-70
	Chip size (mm^2)	6.3 × 6	5 × 5	8 × 6	9.9 × 8.2	9.6 × 7.2	9.6 × 7.2	6.22 × 3.5	8 × 5.8	8 × 5.8	8.4 × 7.6	15.5 × 15.8	14.3 × 11.6	5.3 × 5.3
	Pixel size (μm^2)	40 × 40	18.5 × 18.5	18.5 × 18.5	30 × 30	15 × 15	20 × 20	4.86 × 4.86	9 × 9	9 × 9	4.95 × 4.95	18 × 18	9.8 × 9.8	13 × 13
	Fill factor (%)	8.1	22	22	20	25	20	> 77	11	12	22	8.5	8	22
	Supply voltage (V)	3.3	1.8 & 3.3	1.8 & 3.3	1.8 & 3.3	1.8	1.8	1.1 & 2.5	1.2 & 2.8	1.2 & 2.8		1.8 & 3.3	1.2 & 2.5	1.8 & 3.3
	Stationary noise (ev/pix/s) at 25C	0.05	0.1	0.1	-	0.1	0.1	0.1	0.03	0.03		0.15	0.2	0.1
	CMOS technology (nm)	350	180	180	180	180	180	90	90	90	65/28	180	65	180
		2P4M	1P6M MIM	1P6M MIM	1P6M	1P6M CIS	1P6M CIS	BI CIS	1P5M BSI			1P6M CIS	CIS	1P6M CIS
	Grayscale output	no	yes	yes	yes	no	yes	no	no	no	no	yes	yes	yes
	Grayscale dynamic range (dB)	NA	55	56.7	130	NA	> 100	NA	NA	NA	NA	90	120	50
	Max. frame rate (fps)	NA	35	40	NA	NA	NA	NA	NA	NA	NA	50	100	30
Camera	Max. Bandwidth (Meps)	1	12	12	-	66	66	1066	300	600	1200	200	140	20
	Interface	USB 2	USB 2	USB 3		USB 3	USB 3	USB 3	USB 2	USB 3	USB 3			USB 2
	IMU output	no	1 kHz	1 kHz	no	1 kHz	1 kHz	no	no	1 kHz	no	no	no	1 kHz

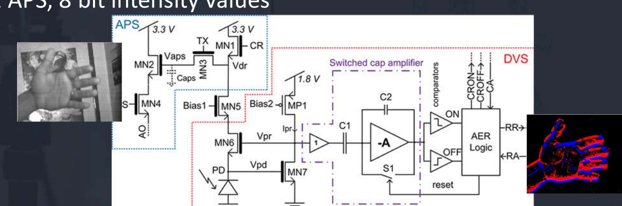
Gallego, G., Delbruck, T., Orchard, G., Bartolozzi, C., Taba, B., Censi, A., Leutenegger, S., Davison, A., Conradt, J., Daniilidis, K., Scaramuzza, D., "Event-based Vision: A Survey," IEEE Trans. Pattern Anal. Machine Intell. (TPAMI), 2020.

Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

Dynamic + Active Pixel Vision Sensor (DAVIS)

Dynamic and Active Pixel Vision Sensor (DAVIS)

- One photodiode shared by APS and DVS readout circuit
- 40Hz APS, 8 bit intensity values

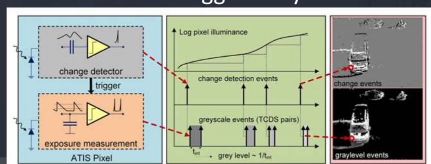


Brandli, C., Berner, R., Yang, M., Liu, S.-C., Delbruck, T., A 240x180 130 dB 3 μs Latency Global Shutter Spatiotemporal Vision Sensor, IEEE J. Solid-State Circuits, 49(10):2333-2341, 2014.

Asynchronous Time-based Image Sensor (ATIS)

Asynchronous Time-based Image Sensor (ATIS)

- Exposure measurement:
  - Measures times to charge a capacitor
  - Inversely proportional to intensity
- Exposure measurements are triggered by contrast events

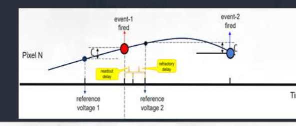


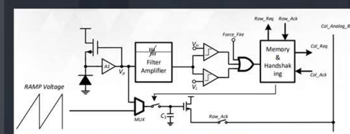
Posch, C., Serrano-Gotarredona, T., Linares-Barranco, B., Delbruck, T., Retinomorphic Event-Based Vision Sensors: Bioinspired Cameras With Spiking Output, Proc. IEEE (2014), 102(10):1470-1484.

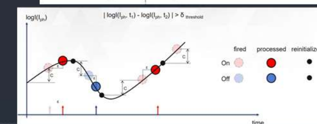
CeleX Pixel

In-Pixel Time Step

- Log voltage can be exported => intensity measurement
- RAMP voltage = analog global time step
  - Reduce latency
  - Readout delay







Shoushun Chen, "Development of Event-based Sensor and Applications," CVPRW Event-based vision, 2021.

Color Event Detection Camera

Color Event Detection Camera

- Color filter array = spatial multiplexing of pixel-sized color filters
- Demosaicking = interpolation of missing color components

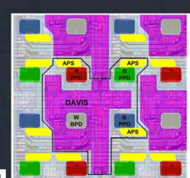
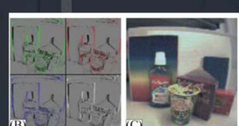


Figure 1 Pixel array arrangement. Yellow=APS area. Purple=DAVIS area. One RGBW unit is circled in black.

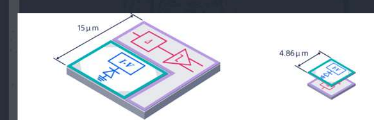


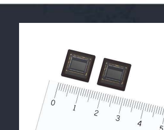
Moeys, D. P., Corradi, F., Li, C., Bamford, S. A., Longinotti, L., Voigt, F. F., Berry, S., Taverni, G., Helmchen, F., Delbruck, T., A Sensitive Dynamic and Active Pixel Vision Sensor for Color or Neural Imaging Applications, IEEE Trans. Biomed. Circuits Syst. 12(1):123-136 2018.
C. Li et al., "Design of an RGBW color VGA rolling and global shutter dynamic and active-pixel vision sensor," 2015 IEEE International Symposium on Circuits and Systems (ISCAS), Lisbon, 2015, pp. 718-721.

Stacked Event-Based Vision Sensor

Stacked Event-Based Vision Sensor

- Conventional technology has photodetector and read out circuit on the same layer.
- By stacking pixel chip (upper layer) and logic chip (lower layer), light efficiency is increased.





https://www.sony-semicon.co.jp/e/products/IS/industry/technology/evs.html

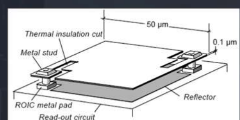
PROPHESEE METAVISION FOR MACHINES

Intelligent Signal Systems Laboratory PI: Keigo Hirakawa, University of Dayton

Longwave Infrared Event Detection Camera

LWIR Event Detection Camera

- Microbolometer measures temperature variation stemming from absorption of thermal infrared radiation
- EVS-ROIC is in CMOS



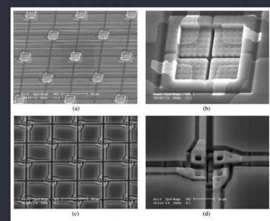


Fig. 2. (a) ROIC after deposition of aluminum reflector, (b) bolometer contact pads of four pixels, (c) microbolometer array on top of the ROIC, and (d) contact sites of four bolometer elements.

Posch, C., Matolin, D., Wohlgenannt, R., Maier, T., Litzenberger, M., A Microbolometer Asynchronous Dynamic Vision Sensor for LWIR, IEEE Sensors Journal, 9(6):654-664, 2009.

Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

PROPHESEE METAVISION FOR MACHINES

Intelligent Signal Systems Laboratory PI: Keigo Hirakawa, University of Dayton

Roadmap for Event Detection Cameras

Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

PROPHESEE METAVISION FOR MACHINES

Intelligent Signal Systems Laboratory PI: Keigo Hirakawa, University of Dayton

Roadmap for Event Detection Cameras

DARPA FENCE Program

- Fast Event-based Neuromorphic Camera and Electronics
- Intelligent Event-Based Imagers
- Event-based infrared camera

https://www.darpa.mil/news-events/2021-07-02

Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

PROPHESEE METAVISION FOR MACHINES

Intelligent Signal Systems Laboratory PI: Keigo Hirakawa, University of Dayton

Introduction

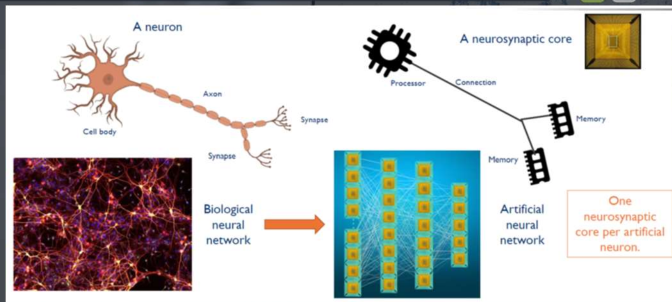
Neuromorphic Processing

Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

PROPHESEE METAVISION FOR MACHINES

Intelligent Signal Systems Laboratory

Neuromorphic Computing



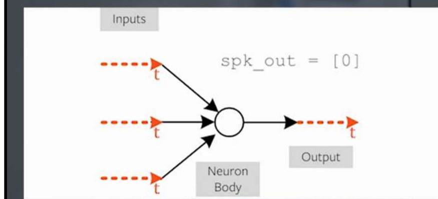
- Credit: Yole

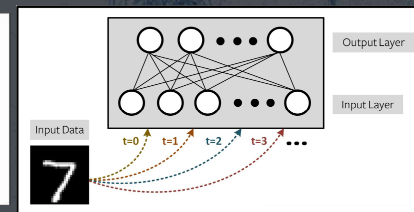
Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

PROPHESEE METAVISION FOR MACHINES

Intelligent Signal Systems Laboratory

Neuromorphic Computing





- Credit: SNNTorch - https://snntorch.readthedocs.io/en/latest/index.html

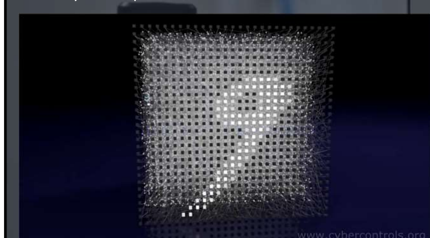
Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

PROPHESEE METAVISION FOR MACHINES

Intelligent Signal Systems Laboratory

Learning from Events

Multilayer Perceptron



Spiking Neural Network

- Credit: Denis Dmitriev for Cyber Controls - https://www.youtube.com/watch?v=3JQ3hYko51Y

Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

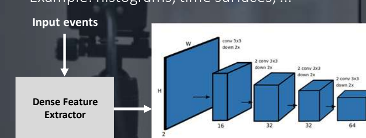
PROPHESEE METAVISION FOR MACHINES

Intelligent Signal Systems Laboratory

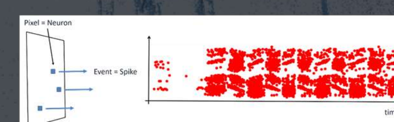
Learning from Events

TWO MAIN APPROACHES

- Dense
  - Example: histograms, time surfaces, …



- Sparse
  - Example: Spiking Neural Networks



⊕ Can leverage Frame-based Algorithm and Hardware

⊖ Latency, Power Consumption, Lose Sparsity

⊕ Low latency, Low Computation

⊖ Limited in Size and Resolution, Limited by Hardware

PROPHESEE METAVISION FOR MACHINES

Intelligent Signal Systems Laboratory PI: Keigo Hirakawa, University of Dayton

Learning from Events

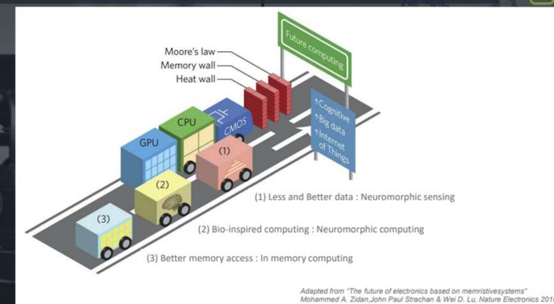
- Credit GML - https://www.graimatterlabs.ai/

Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

PROPHESEE METAVISION FOR MACHINES

Intelligent Signal Systems Laboratory PI: Keigo Hirakawa, University of Dayton

Neuromorphic Computing



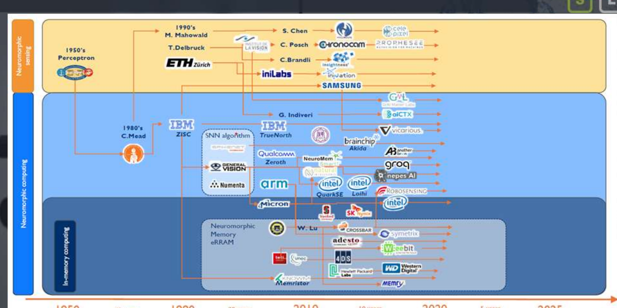
- Credit: Yole

Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

PROPHESEE METAVISION FOR MACHINES

Intelligent Signal Systems Laboratory PI: Keigo Hirakawa, University of Dayton

Ecosystem Timeline



- Credit: Yole

Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

PROPHESEE METAVISION FOR MACHINES

Intelligent Signal Systems Laboratory PI: Keigo Hirakawa, University of Dayton

Event Representation

Event Detection Forward Model And Simulation Methods

Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

PROPHESEE METAVISION FOR MACHINES

Intelligent Signal Systems Laboratory

Event Detection Forward Model

Logarithmic Photoreceptor

$J(t) = \log(I(t))$

image

log voltage

Differencing Circuit

$J(t) - J(t_{i-1})$

Comparator

- Event time stamp

$t_i = \arg\min_t \{t > t_{i-1} \mid |J(t) - J(t_{i-1})| > \epsilon\}$

previous event time stamp

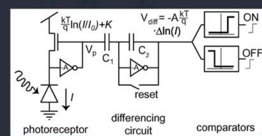
threshold

- Polarity

$p_i = sign(J(t) - J(t_{i-1}))$

increasing? decreasing?

EVS circuit



53

P. Lichtsteiner, C. Posch and T. Delbruck, "A 128 x 128 120 dB 15us Latency Asynchronous Temporal Contrast Vision Sensor," in *IEEE Journal of Solid-State Circuits*, vol. 43, no. 2, pp. 566-576, Feb. 2008.

Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

PROPHESEE METAVISION FOR MACHINES

Intelligent Signal Systems Laboratory

Event Detection Forward Model

Global-order event representation

$e^i = \{e_x^i, e_y^i, e_t^i, e_p^i\}$

i-th event on sensor

i-th event time

i-th event location

i-th event polarity

Pixel-order event representation

$e_i(x,y) = \{t_i(x,y), p_i(x,y)\}$

time of i-th event at pixel (x,y)

i-th event at pixel (x,y)

polarity of i-th event at pixel (x,y)

54

Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

PROPHESEE METAVISION FOR MACHINES

Intelligent Signal Systems Laboratory

Event Detection Forward Model

Why logarithmic photoreceptor?

log intensity image

$J(x,y,t) = \log(I(x,y,t))$

image

- Photometry:

$I(x,y,t) = L \times R(x,y,t)$

illumination

reflectance

- Differencing Circuit:

$$\begin{aligned} J(x,y,t) - J(x,y,t_{i-1}) &= \log I(x,y,t) - \log I(x,y,t_{i-1}) \\ &= \log L + \log R(x,y,t) - \log L - \log R(x,y,t_{i-1}) \\ &= \log R(x,y,t) - \log R(x,y,t_{i-1}) \end{aligned}$$

- Conclusion: events are *illumination invariant*

55

Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

PROPHESEE METAVISION FOR MACHINES

Intelligent Signal Systems Laboratory

Event Detection Forward Model

How often does an event occur?

- Differencing Circuit

$J(t) - J(t_{i-1}) = (t - t_{i-1}) \cdot J_t(t)$

temporal derivative

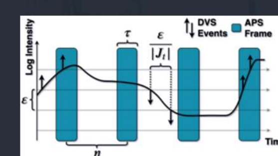
time difference

- Comparator

$|J(t) - J(t_{i-1})| = (t - t_{i-1}) \cdot |J_t(t)| > \epsilon$

- Conclusion: Rate of event generation is

$\Delta_t^{-1} = \dfrac{J_t(t)}{\epsilon}$



56

Baldwin, R., Mohammed Almatrafi, Vijayan Asari, and Keigo Hirakawa. "Event Probability Mask (EPM) and Event Denoising Convolutional Neural Network (EDnCNN) for Neuromorphic Cameras." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1701-1710. 2020.

Event Detection Forward Model

How often does an event occur?

- Event Generation Rate

$$\Delta_t^{-1} = \frac{J_t(x,y,t)}{\epsilon}$$

- Optical Flow Equation (OFE)

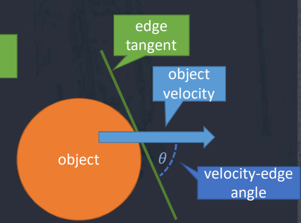
$$J_t(x,y,t) \approx -V_x \cdot J_x(x,y,t) - V_y \cdot J_y(x,y,t)$$
$$= -|\vec{V}| \cdot |\nabla J| \cdot \sin\theta$$

velocity

spatial gradient

velocity-edge angle

- **Conclusion:** Events generated when velocity and edge orientations are orthogonal.



Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

Video-Based Event Simulator

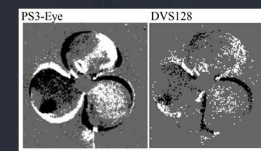
APS video => Events

- Frame differeincing to determine log-intensity change

$$J_t(x,y,t) \approx \frac{J(x,y,t+\Delta_t) - J(x,y,t)}{\Delta_t}$$

- Frame interpolation to upsample frame rate
- Model the temporal spike time decay

- Pros: APS ground truth available
- Cons: No accurate model of noise
- Cons: Motion less than one pixel per frame.
- Cons: APS blurs motion, clips high dynamic range scenes



Katz, M. L., Nikolic, K., Delbruck, T. “Live demonstration: Behavioural emulation of event-based vision sensors,” IEEE Int. Symp. Circuits & Systems (ISCAS), 2012.
Kaiser, J., Tieck, J. C. V., Hubschneider, C., Wolf, P., Weber, M., Hoff, M., Friedrich., A., Wojtasik, K., Roennau, A., Kohlhaas, R., Dillmann, R., Zoellner, M., “Towards a framework for end-to-end control of a simulated vehicle with spiking neural networks,” IEEE Int. Conf. on Sim., Modeling, and Program.for Auto. Robots, 2016.
Bi, Y. and Andreopoulos, Y., “PIX2NVS: Parameterized conversion of pixel-domain video frames to neuromorphic vision streams,” IEEE Int. Conf. Image Processing 2017.
Delbruck, T., Hu, Y., He, Z., “V2E: From video frames to realistic DVS event camera streams,” arXiv, 2020.

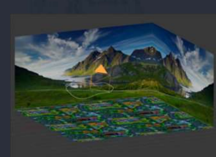
Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

Forward Model-Based Event Simulator

3D Graphics Model => Events

- Generate synthetic high speed frames
- Use frame differencing to simulate events

- Pros: 3D graphics ground truth available
- Cons: No accurate model of noise
- Cons: Photorealism of scenes



E. Mueggler, H. Rebecq, G. Gallego, T. Delbruck, D. Scaramuzza, “The Event-Camera Dataset and Simulator: Event-based Data for Pose Estimation, Visual Odometry, and SLAM,” Int. J. Robotics Research, 36:2, pp. 142-149, 2017.
W. Li, S. Saeedi, J. McCormac, R. Clark, D. Tzoumanikas, Q. Ye, Y. Huang, R. Tang, S. Leutenegger, “Interiornet: Mega-scale multi-sensor photo-realistic indoor scenes dataset,” British Machine Vis. Conf. (BMVC), 2018.
H. Rebecq, D. Gehrig, D. Scaramuzza, “ESIM: an Open Event Camera Simulator,” Conf. on Robot Learning (CoRL), 2018.

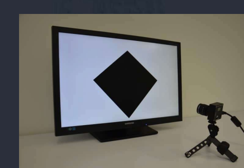
Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

Event Camera + Monitor

Event Camera + Monitor

- Camera observes a video played out on monitor

- Pros: The event data is real
- Pros: Real sensor noise represented
- Pros: Easy to calibrate coordinates
- Cons: 59Hz/60Hz flickering of the screen is very observable in sensor data



E. Mueggler, B. Huber, and D. Scaramuzza, “Event based, 6-DOF Pose Tracking for High-Speed Maneuvers,” IEEE, 2014.
G. Orchard, A. Jayawant, G. K. Cohen, et al., “Converting static image datasets to spiking neuromorphic datasets using saccades,” Frontiers in Neuroscience, vol. 9, no. NOV, pp. 1–15, 2015.
T. Serrano-Gotarredona and B. Linares-Barranco, “Poker-DVS and MNIST-DVS. Their History, How They Were Made, and Other Details,” Frontiers in Neuroscience, vol. 9, no. December, pp. 1–10, 2015.

Controlled Motion

Controlled Motion

- Camera on motorized platform
- Static fronto parallel planar scene

- Pros: The event data is real
- Pros: Real sensor noise represented
- Cons: Scene-camera registration errors
- Cons: Flat scenes only

G. Orchard, A. Jayawant, G. K. Cohen, et al., "Converting static image datasets to spiking neuromorphic datasets using saccades," Frontiers in Neuroscience, vol. 9, no. NOV, pp. 1–15, 2015.
E. Mueggler, H. Rebecq, G. Gallego, T. Delbruck, D. Scaramuzza, "The Event-Camera Dataset and Simulator: Event-based Data for Pose Estimation, Visual Odometry, and SLAM," Int. J. Robotics Research, 36:2, pp. 142-149, 2017.

Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

Event Camera + Gimbal + IMU

Event Prediction

- DAVIS cameras has APS, EVS, IMU
- Optical Flow Equation (OFE)

$$J_t(x,y,t) \approx -V_x \cdot J_x(x,y,t) - V_y \cdot J_y(x,y,t)$$

From IMU

From IMU

From APS

From APS

- Pros: The event data is real
- Pros: Real sensor noise represented
- Pros: Real scenes, not restricted to flat
- Pros: Can handle fast motion (since single frame)
- Cons: Restricted to static scene
- Cons: Restricted to rotation (not translation)
- Cons: Events only near APS frames

Gimbal + IMU

- Sensor restricted to rotation
- Use gyroscope
- Instantaneous pixel velocity recoverable

Baldwin, R., Mohammed Almatrafi, Vijayan Asari, and Keigo Hirakawa. "Event Probability Mask (EPM) and Event Denoising Convolutional Neural Network (EDnCNN) for Neuromorphic Cameras." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1701-1710. 2020.

Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

Event Representation

Event Representation in Machine Vision

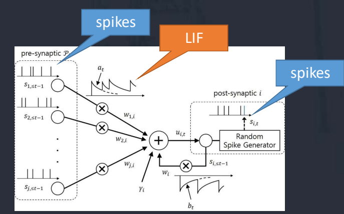
Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

Spiking Neural Network

Spiking Neural Networks (SNN)

- Bio-inspired architecture
- Natively asynchronous and spiking (event) form
- Leaky integrate-and-fire (LIF) neuron models
- Well-matched for event detection cameras

- Pros: High accuracy
- Pros: Very low power, fast throughput
- Cons: Very difficult to train, non-optimal solutions
- Cons: Not human interpretable
- Cons: Suffers from "spike starvation" in very deep networks.



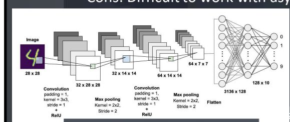
H. Jang, O. Simeone, B. Gardner and A. Gruning, "An Introduction to Probabilistic Spiking Neural Networks: Probabilistic Models, Learning Rules, and Applications," in *IEEE Signal Processing Magazine*, vol. 36, no. 6, pp. 64-77, Nov. 2019, doi: 10.1109/MSP.2019.2935234.

Convolutional Neural Network

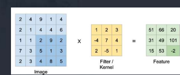
Convolutional Neural Networks (CNN)

- Natively synchronous with frame data input
- Successive convolution, activation, and pooling
- Well-matched for APS cameras

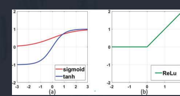
- Pros: High flexibility
- Pros: Easier to train, though time consuming
- Cons: Not human interpretable
- Cons: Computationally complex
- Cons: Difficult to work with asynchronous data



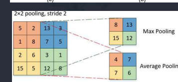
conv



activation



pooling

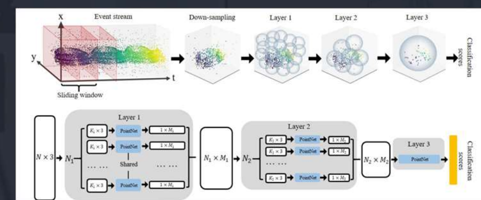


://towardsdatascience.com/convolution-neural-networks-a-beginners-guide-implementing-a-mnist-hand-written-digit-8aa60330d022
https://www.intechopen.com/books/green-electronics/optimizing-of-convolutional-neural-network-accelerator

Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

3D Event Representation

- A network to learn 3D geometry directly from 3D point cloud.
- Event camera adaptation of PointNet
  - PointNet: C. R. Qi (2017) 3D classification and segmentation



Q. Wang, Y. Zhang, J. Yuan and Y. Lu, "Space-Time Event Clouds for Gesture Recognition: From RGB Cameras to Event Cameras," 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), Waikoloa Village, HI, USA, 2019, pp. 1826-1835.

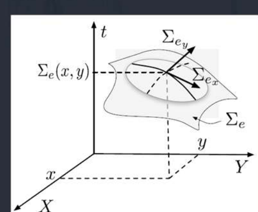
Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

Event Manifold

Local Plane Fitting

- Manifold learned from event cloud
- Regularization used to overcome noise
- Spatial gradient = motion orientation
- Temporal slope = velocity

Surface of Active Events

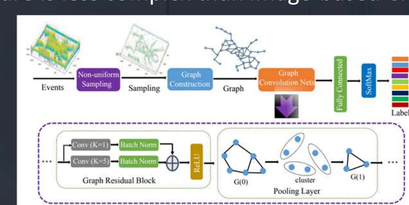


R. Benosman, C. Clercq, X. Lagorce, S.-H. Ieng, and C. Bartolozzi, "Event-based visual flow," IEEE transactions on neural networks and learning systems, vol. 25, no. 2, pp. 407–417, 2014.

Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

Graph Based Object

- Represent events as graphs
- Use residual graph CNN
- Network structure is less complex than image-based CNN



Y. Bi, A. Chadha, A. Abbas, E. Bourtsoulatze, and Y. Andreopoulos, "Graph-based object classication for neuromorphic vision sensing," in Proceedings of the IEEE International Conference on Computer Vision, 2019, pp. 491-501.

Asynchronous Data in ML

- Event data is 3D: 1D continuous time + 2D discrete space
- ML on 3D data is challenging (need more data)
- ML on continuous data is challenging (data representation)
- Preprocessing step to "encode" or "represent" 3D data efficiently
  - ⇒ But this encoding is "lossy"

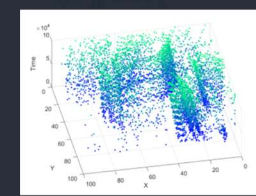
Asynchronous => Synchronous

Synchronous Representations

- 2D discrete space + 1D continuous time
  - ⇒ too difficult
- Need preprocessing step to represent 3D data efficiently
- Projection to 2D/3D discrete signal is lossy

Proxy Frame

- Capture spatial-temporal scene features in proxy synchronous frame-like representation
- Higher "frame rate"
  - (Arguably) closer to asynchronous data
  - Computational overhead is very high
  - Wasteful usage of sparse data



Temporal/Event Windowing

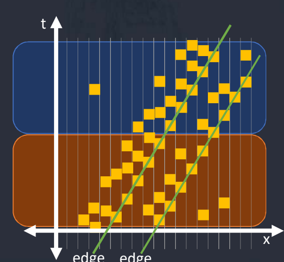
Temporal Windowing

- Partition by fixed time interval

Event Windowing

- Partition by fixed number of events

- Short temporal window…
  - (Arguably) closer to asynchronous data
  - Computational overhead is very high
  - Wasteful usage of sparse data

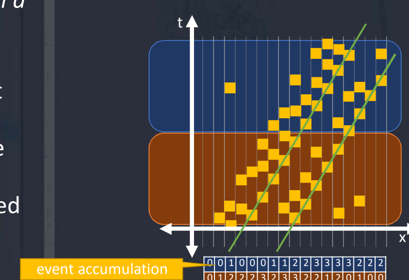


Event Accumulation

How do we represents events in a window?

Event Accumulation

- Count the events occurring at each pixel
- Pros: decently robust to noise and missing events
- Cons: multiple edges combined together
- Cons: Timing information lost



A. I. Maqueda, A. Loquercio, G. Gallego, N. Garcia, and D. Scaramuzza, "Event-based vision meets deep learning on steering prediction for self-driving cars," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 5419- 5427.
A. Z. Zhu, L. Yuan, K. Chaney, and K. Daniilidis, "Ev-flownet: Self-supervised optical flow estimation for event-based cameras," Robotics: Science and Systems 2018.

PROPHESEE METAVISION FOR MACHINES
Intelligent Signal Systems Laboratory

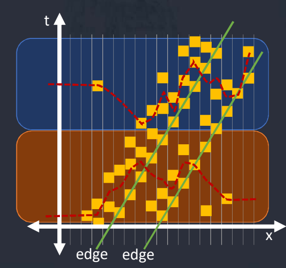
Time Surface

How do we represents events in a window?

Mean/Median "Time Surface"

- Represent average time stamp as an image intensity
- a.k.a. *Surface of Active Events (SAE)*

- Pros: Timing preserved
- Cons: Multiple events and edges are combined



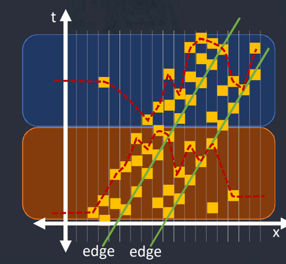
 A. Sironi, M. Brambilla, N. Bourdis, X. Lagorce, and R. Benosman, "HATS: Histograms of averaged time surfaces for robust event-based object classification," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 1731-1740.

Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

Hierarchy Of Time Surface (HOTS)

Hierarchy Of Time Surface (HOTS)

- In some biological phenomena, most recent events carry greatest influence.
  - ⇒ Represent time-stamp of most recent event as image intensity
- Pros: Preserves some event times
- Cons: Timing of last events are noisy



 Lagorce, X., Orchard, G., Gallupi, F., Shi, B., Benosman, R., "HOTS: A Hierarchy Of event-based Time-Surfaces for pattern recognition," IEEE Trans. Pattern Anal. Machine Intell. (TPAMI), 39(7):1346-1359, 2017.

Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

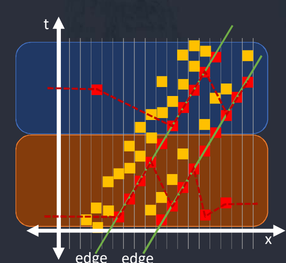
Filtered Surface of Active Events (FSAE)

Filtered Surface of Active Events (FSAE)

- One edge yields multiple events
- First event corresponds to the timing of edge arrival
- Filter out events that immediately follow another event.

$$\{t_i | (t_i - t_{i-1} > \tau)\}$$

- Pros: Improved edge fidelity
- Pros: Reduced event representation
- Cons: Not robust to noise



 I. Alzugaray and M. Chli, "Asynchronous corner detection and tracking for event cameras in real time," IEEE Robotics and Automation Letters, vol. 3, no. 4, pp. 3177-3184, 2018.

Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

Inceptive Event Time Surface (IETS)

Events classified into three types:

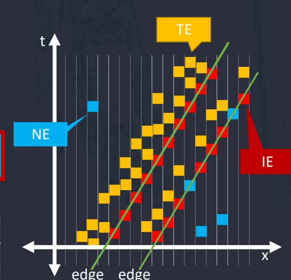
- Inceptive Events (IE)
  - First event in a group of events
  - Corresponds to edge arrival timing
- Trailing Events (TE)
  - Events following IE
  - Corresponds to edge magnitude
- Noisy Events (NE)
  - Temporally isolated events

Follows another event

Long time since last event

Followed by another event

	$t_i - t_{i-1} \le \tau$	$t_i - t_{i-1} > \tau$
$t_{i+1} - t_i \le \tau$	Trailing Event	Inceptive Event
$t_{i+1} - t_i > \tau$	Trailing Event	Noisy Event



 Baldwin R.W., Almatrafi M., Kaufman J.R., Asari V., Hirakawa K., "Inceptive Event Time-Surfaces for Object Classification Using Neuromorphic Cameras," Int. Conf. on Image Analysis and Recognition (ICIAR), 2019.

Inceptive Event Time Surface (IETS)

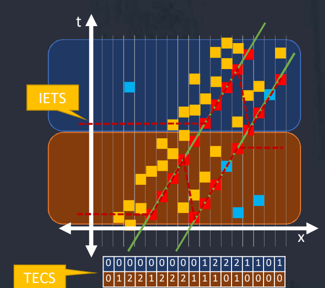
Inceptive Event Time Surface (IETS)

- IE corresponds to the timing of edge arrival
- IETE encodes time of edge arrivals as image intensity.

Trailing Event Count Surface (TECS)

- TE corresponds to edge magnitude
- Timing of TE not meaningful
- Associate TE count with the IE surface

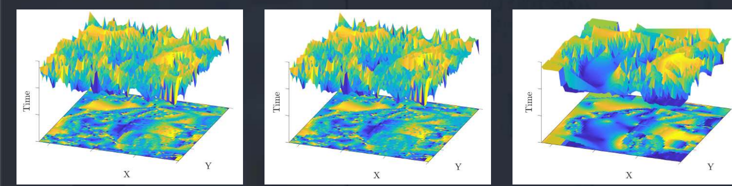
- Pros: Improved edge fidelity
- Pros: Robustness to noise
- Pros: Data reduction by 85%
- Cons: Filters out slow moving edges



 Baldwin R.W., Almatrafi M., Kaufman J.R., Asari V., Hirakawa K., "Inceptive Event Time-Surfaces for Object Classification Using Neuromorphic Cameras," Int. Conf. on Image Analysis and Recognition (ICIAR), 2019.

Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

Time Surface



Averaged Time Surface | FSAE Time Surface | IE Time Surface

Removes 34% of N-Cars events | Removes 85% of N-Cars events

 Baldwin R.W., Almatrafi M., Kaufman J.R., Asari V., Hirakawa K., "Inceptive Event Time-Surfaces for Object Classification Using Neuromorphic Cameras," Int. Conf. on Image Analysis and Recognition (ICIAR), 2019.

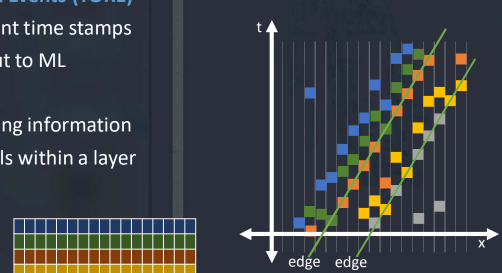
Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

Time-Ordered Recent Events (TORE)

Time-Ordered Recent Events (TORE)

- FIFO of N most recent time stamps
- Use N layers as input to ML

- Pros: Preserves timing information
- Cons: Adjacent pixels within a layer have mixed edges.



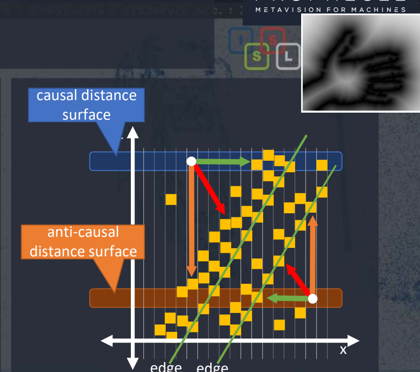
 Baldwin, R., Mohammed Almatrafi, Vijayan Asari, and Keigo Hirakawa. "Time-Ordered Recent Event (TORE) Volumes for Event Cameras," https://arxiv.org/abs/2103.06108.

Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

Distance Surface

Distance Surface

- Compute distance transform
  - 2D spatial
  - 1D temporal
  - 2D+1D spatial-temporal
- Can be causal or anti-causal
  - 1D temporal causal distance surface = HOTS
  - 1D temporal anti-causal distance surface ≈ FSAE TS

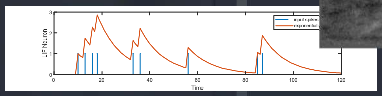


 Almatrafi, M., Baldwin, R., Aizawa, K., Hirakawa, K., "Distance Surface for Event-Based Optical Flow," IEEE Trans. Pattern Anal. Machine Intell. (TPAMI), 2020.

Exponential Surface

Leaky Integrate-and-Fire (LIF)

- Neurons that introduce "memory"
- Models exponentially decaying membrane potentials
- LIF Neuron models used extensively in spiking neural networks
- Idea: Sample LIF response to events in regular grid.

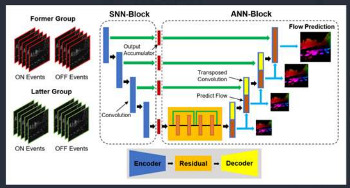


D. Gehrig, A. Loquercio, K. Derpanis and D. Scaramuzza, "End-to-End Learning of Representations for Asynchronous Event-Based Data," *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, Korea (South), 2019, pp. 5632-5642.
C. Scheerlinck, N. Barnes, and R. Mahony, "Asynchronous spatial image convolutions for event cameras," IEEE Robotics and Automation Letters,4(2):816–822, 2019.

SNN-CNN Hybrid

SNN-CNN Hybrid

- SNN is efficient at handling asynchronous events.
- Deep SNN suffers from data starvation
- Solution: U-Net structure with SNN encoder and CNN decoder.
- Still requires lossy asynchronous-to-synchronous data mapping



Chankyu Lee, Adarsh Kumar Kosta, Alex Zihao Zhu, Kenneth Chaney, Kostas Daniilidis, Kaushik Roy, "Spike-FlowNet: Event-based Optical Flow Estimation with Energy-Efficient Hybrid Neural Networks," Arxiv https://arxiv.org/abs/2003.06696, July 21, 2020.

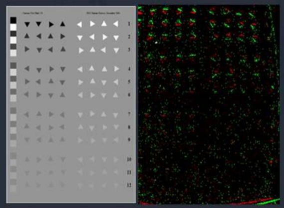
Event Representation

Event Filtering

Event Filtering

Goals of Event Filtering

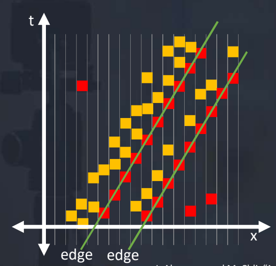
- Simplification: Represent scene features with less data
- Noise Robustness: Filter out events not pertinent to scene
- Data Reduction: Do more with less data



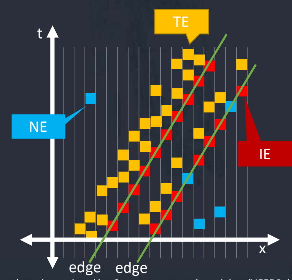
Baldwin, R., Mohammed Almatrafi, Vijayan Asari, and Keigo Hirakawa. "Time-Ordered Recent Event (TORE) Volumes for Event Cameras," https://arxiv.org/abs/2103.06108.

Event Filtering

Filtered Surface of Active Events



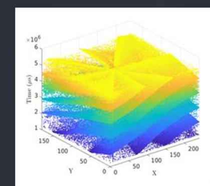
Inceptive Event Filtering



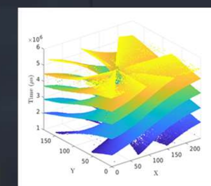
85 I. Alzugaray and M. Chli, "Asynchronous corner detection and tracking for event cameras in real time," IEEE Robo. & Auto. Letters, vol. 3, no. 4, pp. 3177-3184, 2018.
Baldwin R.W., Almatrafi M., Kaufman J.R., Asari V., Hirakawa K., "Inceptive Event Time-Surfaces for Object Classification Using Neuromorphic Cameras," ICIAR, 2019.

Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

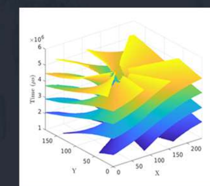
Inceptive Event + Local Plane Fitting



Raw Senor Events



Inceptive Events Filter



IE Filter + Local Plane Fitting

86 R. Benosman, C. Clercq, X. Lagorce, S.-H. Ieng, and C. Bartolozzi, "Event-based visual flow," IEEE transactions on neural networks and learning systems, vol. 25, no. 2, pp. 407–417, 2014.
Baldwin R.W., Almatrafi M., Kaufman J.R., Asari V., Hirakawa K., "Inceptive Event Time-Surfaces for Object Classification Using Neuromorphic Cameras," ICIAR, 2019.

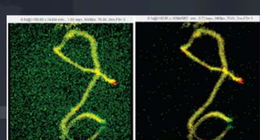
Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

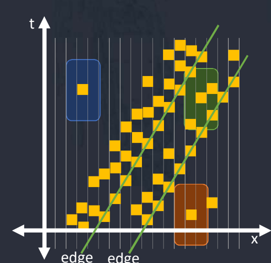
Background Activity Filter (BAF)

Background Activity Filter (BAF)

- Idea: Noise occurs in spatial-temporal isolation.
- Filter out an event if the spatially neighboring pixels have not generated

$$\left\{ t_i(x,y) \middle| \begin{array}{c} \exists t_{i'}, (\Delta_x, \Delta_y) \in W \quad \ni \\ t_i(x,y) - t_{i'}(x+\Delta_x, y+\Delta_y) < \tau \end{array} \right\}$$





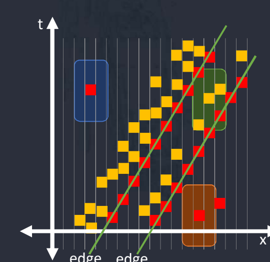
Tobi Delbruck. "Frame-free dynamic digital vision." InPro-ceedings of Intl. Symp. on Secure-Life Electronics, AdvancedElectronics for Quality Life and Society, pages 21–26, 2008

Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

Nearest Neighbor Filtering (NNF)

Nearest Neighbor Filtering (NNF)

- First stage: FSAE (temporal filtering)
- Second stage: BAF (spatial filtering)

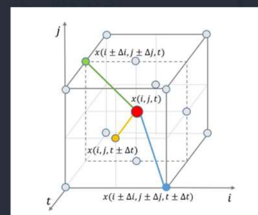


88 Dominguez-Morales, M., Jimenez-Fernandez, A., Paz, R., López-Torres, M. R., Cerezuela-Escudero, E., Linares-Barranco, A., et al. (2011). "An approach to distance estimation with stereo vision using address-event-representation. Neural information processing," International Conference on Neural Information Processing (Berlin; Heidelberg: Springer), Vol. 7062, 190–198. doi: 10.1007/978-3-642-24955-6_23

Graph-Based Cost Minimization

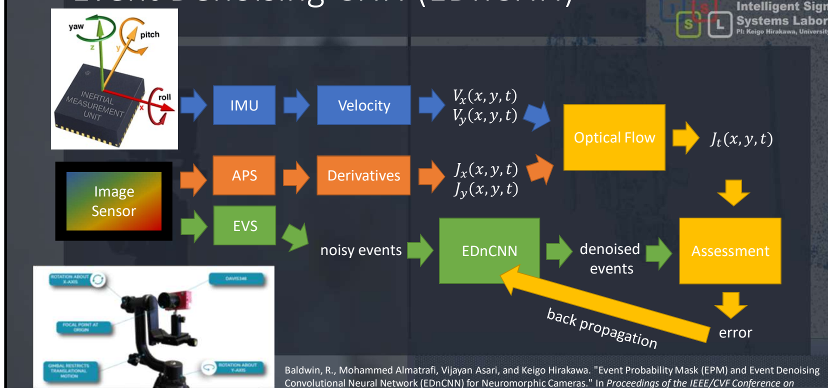
Cost function

- X=events due to signal
- N=noisy events
- $Y = X + N$ observation
- Penalize
  - $Y \neq X$
  - X not similar to neighbors
- $\hat{X} = \arg\min_{X\in\{0,1\}} ||X - Y||_0 + \lambda \left|\left|X - X_{neighbor}\right|\right|_0$



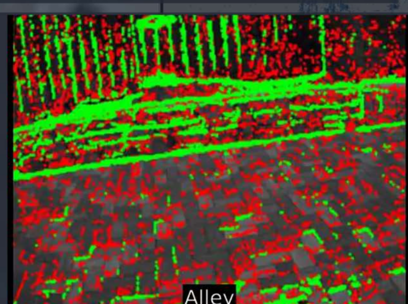
J. Wu, C. Ma, L. Li, W. Dong and G. Shi, "Probabilistic Undirected Graph Based Denoising Method for Dynamic Vision Sensor," in IEEE Transactions on Multimedia, 2020.

Event Denoising CNN (EDnCNN)



Baldwin, R., Mohammed Almatrafi, Vijayan Asari, and Keigo Hirakawa. "Event Probability Mask (EPM) and Event Denoising Convolutional Neural Network (EDnCNN) for Neuromorphic Cameras." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1701-1710. 2020.

Event Denoising CNN (EDnCNN)



Baldwin, R., Mohammed Almatrafi, Vijayan Asari, and Keigo Hirakawa. "Event Probability Mask (EPM) and Event Denoising Convolutional Neural Network (EDnCNN) for Neuromorphic Cameras." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1701-1710. 2020.

Filtering Result

Baldwin, R., Mohammed Almatrafi, Vijayan Asari, and Keigo Hirakawa. "Event Probability Mask (EPM) and Event Denoising Convolutional Neural Network (EDnCNN) for Neuromorphic Cameras." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1701-1710. 2020.

Denoising Evaluation

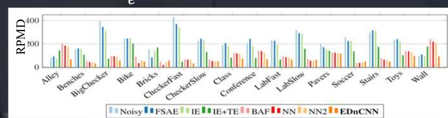
Relative Plausibility Measure of Denoising (RPMD)

- Compute $J_t(x,y,t)$ from APS+IMU

  => estimate the rate of event generated at $J_t(x,y,t)$

  => estimate probability of event occurring within time $[t, t+\Delta_t)$

- $RPMD = -\frac{1}{N} \log \frac{\Pr[e_{denoised}]}{\arg\max_e \Pr[e]}$ (Smaller is better)



Baldwin, R., Mohammed Almatrafi, Vijayan Asari, and Keigo Hirakawa. "Event Probability Mask (EPM) and Event Denoising Convolutional Neural Network (EDnCNN) for Neuromorphic Cameras." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1701-1710. 2020.

Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

Spatial-Temporal Convolutional Filter

Exploit relationship between spatially neighboring events

- Edge detection
- Edge orientation extraction
- Blurring
- Sharpening
- "Hand Crafted" Feature Extraction
- Signal estimation, signal denoising

Challenges

- Events describe temporal change to intensity

  => events are temporally "highpass"

- Goal: spatially filter temporally highpass intensity signal

Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

Spatial-Temporal Convolutional Filter

Accumulated Events

- $J(x,y,t) = J(x,y,t_0) + \tau\,(p_1 + \cdots + p_N)$

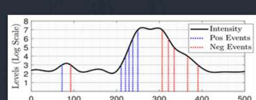
unknown initial value

threshold value

events indicate thresholds were crossed

Challenges

- We don't have $J(x,y,t_0)$
- Only have samples at $t = t_0, t_1, t_2, \ldots$



Mohammed Almatrafi, and Keigo Hirakawa. "DAViS Camera Optical Flow." IEEE Transactions on Computational Imaging, 2019.

Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

Spatial-Temporal Convolutional Filter

***"Give up"* on temporal lowpass signal**

=> apply temporal highpass filter $G(t) = \frac{d}{dt}\, e^{-\lambda t} u(t)$

$$J_{HP}(x,y,t) = G(t) \star J(x,y,t)$$
$$= G(t) \star J(x,y,0) + G(t) \star \tau\,(p_1(x,y) + \cdots + p_n(x,y))$$

Filtered events

Goes to zero

unknown initial value

In temporal Laplace transform

$$\mathcal{J}_{HP}(x,y,s) = \tau \cdot \frac{p_1 e^{-t_1 s} + \cdots + p_1 e^{-t_i s}}{s+\lambda}$$

C. Scheerlinck, N. Barnes, and R. Mahony, "Asynchronous spatial image convolutions for event cameras," IEEE Robotics and Automation Letters,4(2):816–822, 2019.

Spatial-Temporal Convolutional Filter

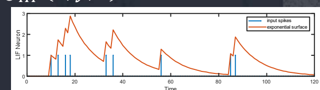
filtered signal

convolution kernel

Filter J_{HP}

- Space-Time domain: $K(x,y,t) = H(x,y) \star J_{HP}(x,y,t)$
- Space-Laplace domain: $\mathcal{K}(x,y,s) = H(x,y) \star \mathcal{J}_{HP}(x,y,s)$

Inverse Laplace leads to an efficient algorithm



- t_p is the previous event time stamp
- (e_x, e_y, e_t, e_p) is the new incoming event

Just need to keep track of last event time

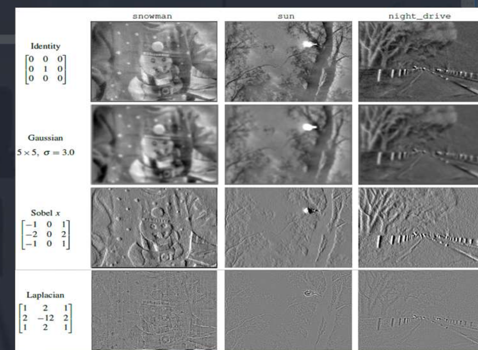
- No event: $K(x,y,t) = e^{-\lambda(t-t_p)} K(x,y,t_p)$
- New event update:

Shifted filter kernel

$$K(x,y,e_t) = e^{-\lambda(e_t-t_p)} K(x,y,t_p) + \tau\, e_p H(x-e_x, y-e_y)$$

 C. Scheerlinck, N. Barnes, and R. Mahony, "Asynchronous spatial image convolutions for event cameras," IEEE Robotics and Automation Letters,4(2):816–822, 2019.

Spatial-Temporal Convolutional Filter



 C. Scheerlinck, N. Barnes, and R. Mahony, "Asynchronous spatial image convolutions for event cameras," IEEE Robotics and Automation Letters,4(2):816–822, 2019.

Computational Imaging

Optical Flow

Optical Flow Equation

velocity

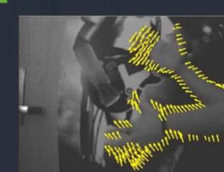
Optical Flow Equation (OFE)

spatial gradient

$$I_t(x,y,t) \approx -V_x \cdot I_x(x,y,t) - V_y \cdot I_y(x,y,t) = -|\vec{V}| \cdot |\nabla I| \cdot \sin\theta$$

temporal derivative

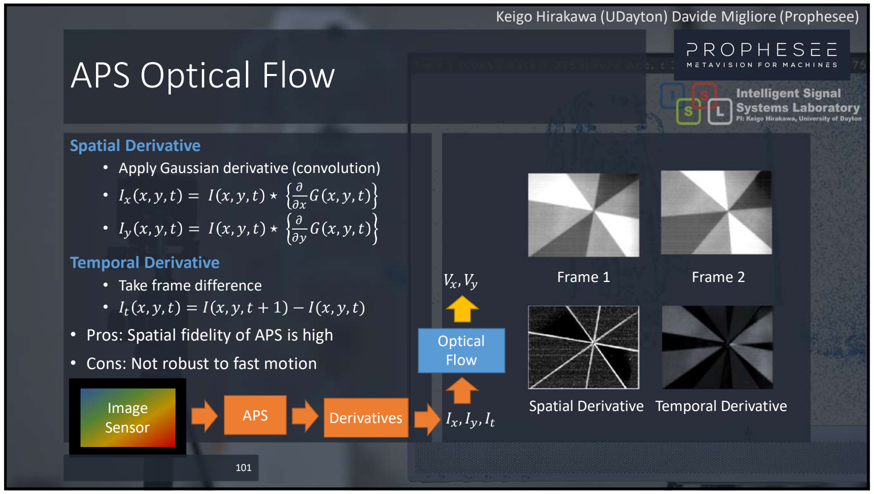
velocity-edge angle

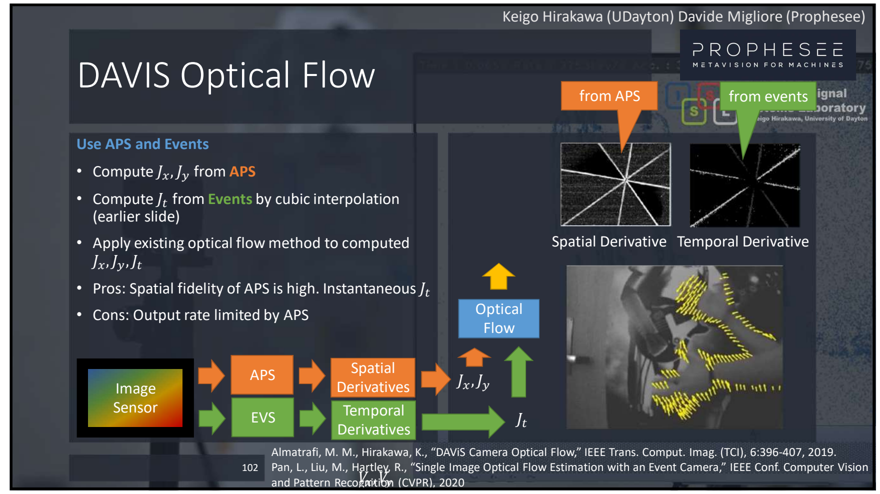


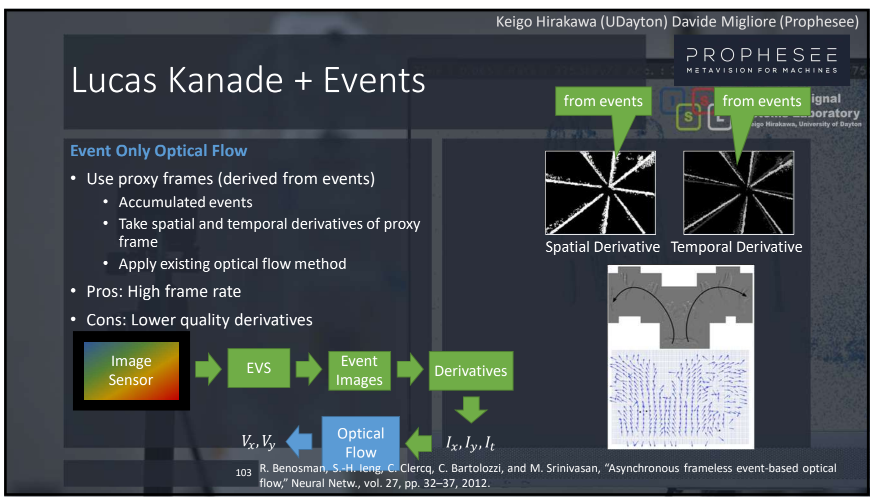
- Goal: Estimate velocity V_x, V_y

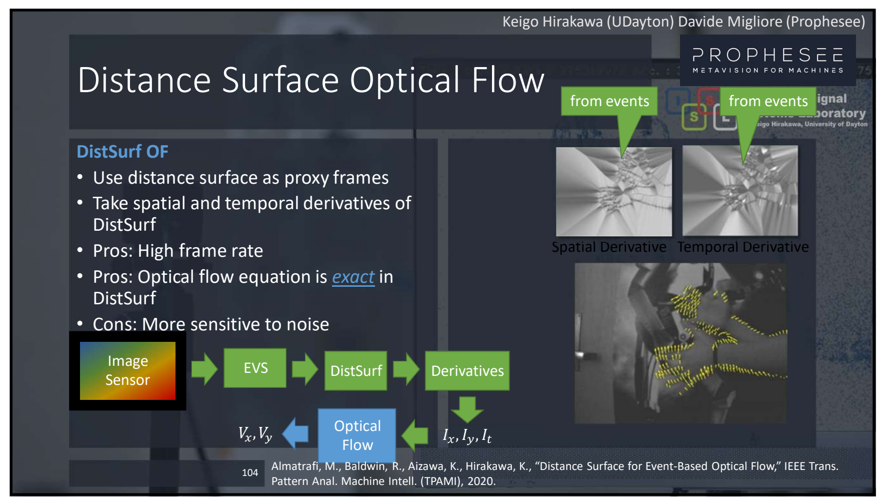
Challenges:

- How do you compute I_x, I_y, I_t from events?
- *Aperture problem:* one equation, two unknowns


Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)
PROPHESEE
METAVISION FOR MACHINES
Intelligent Signal Systems Laboratory
APS Optical Flow
Spatial Derivative
Apply Gaussian derivative (convolution)
$I_x(x,y,t) = I(x,y,t) \star \left\{\frac{\partial}{\partial x}G(x,y,t)\right\}$
$I_y(x,y,t) = I(x,y,t) \star \left\{\frac{\partial}{\partial y}G(x,y,t)\right\}$
Temporal Derivative
Take frame difference
$I_t(x,y,t) = I(x,y,t+1) - I(x,y,t)$
Pros: Spatial fidelity of APS is high
Cons: Not robust to fast motion
Image Sensor
APS
Derivatives
I_x, I_y, I_t
Optical Flow
V_x, V_y
Frame 1
Frame 2
Spatial Derivative
Temporal Derivative
101


Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)
PROPHESEE
METAVISION FOR MACHINES
DAVIS Optical Flow
from APS
from events
Use APS and Events
Compute J_x, J_y from APS
Compute J_t from Events by cubic interpolation (earlier slide)
Apply existing optical flow method to computed J_x, J_y, J_t
Pros: Spatial fidelity of APS is high. Instantaneous J_t
Cons: Output rate limited by APS
Spatial Derivative
Temporal Derivative
Optical Flow
Image Sensor
APS
EVS
Spatial Derivatives
Temporal Derivatives
J_x, J_y
J_t
102
Almatrafi, M. M., Hirakawa, K., "DAVIS Camera Optical Flow," IEEE Trans. Comput. Imag. (TCI), 6:396-407, 2019.
Pan, L., Liu, M., Hartley, R., "Single Image Optical Flow Estimation with an Event Camera," IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2020


Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)
PROPHESEE
METAVISION FOR MACHINES
Lucas Kanade + Events
from events
from events
Event Only Optical Flow
Use proxy frames (derived from events)
Accumulated events
Take spatial and temporal derivatives of proxy frame
Apply existing optical flow method
Pros: High frame rate
Cons: Lower quality derivatives
Spatial Derivative
Temporal Derivative
Image Sensor
EVS
Event Images
Derivatives
I_x, I_y, I_t
Optical Flow
V_x, V_y
103
R. Benosman, S.-H. Ieng, C. Clercq, C. Bartolozzi, and M. Srinivasan, "Asynchronous frameless event-based optical flow," Neural Netw., vol. 27, pp. 32–37, 2012.


Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)
PROPHESEE
METAVISION FOR MACHINES
Distance Surface Optical Flow
from events
from events
DistSurf OF
Use distance surface as proxy frames
Take spatial and temporal derivatives of DistSurf
Pros: High frame rate
Pros: Optical flow equation is exact in DistSurf
Cons: More sensitive to noise
Spatial Derivative
Temporal Derivative
Image Sensor
EVS
DistSurf
Derivatives
I_x, I_y, I_t
Optical Flow
V_x, V_y
104
Almatrafi, M., Baldwin, R., Aizawa, K., Hirakawa, K., "Distance Surface for Event-Based Optical Flow," IEEE Trans. Pattern Anal. Machine Intell. (TPAMI), 2020.

Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

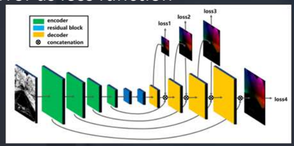
PROPHESEE METAVISION FOR MACHINES

Intelligent Signal Systems Laboratory

Learning Based Event Optical Flow

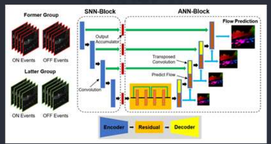
CNN Based Optical Flow

- Accumulated events as input
- Auto-encoder CNN structure
- During training, use APS photometric error as loss function



SNN-CNN Hybrid

- U-Net structure with SNN encoder and CNN decoder.
- Still requires lossy asynchronous-to-synchronous data mapping



Zhu, A., Yuan, L., Chaney, K., Daniilidis, K., “EV-FlowNet: Self-Supervised Optical Flow Estimation for Event-based Cameras,” Robotics: Science and Systems (RSS), 2018.
Chankyu Lee, Adarsh Kumar Kosta, Alex Zihao Zhu, Kenneth Chaney, Kostas Daniilidis, Kaushik Roy, "Spike-FlowNet: Event-based Optical Flow Estimation with Energy-Efficient Hybrid Neural Networks," Arxiv https://arxiv.org/abs/2003.06696, July 21, 2020.

Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

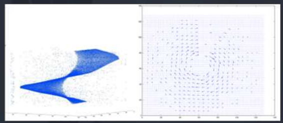
PROPHESEE METAVISION FOR MACHINES

Intelligent Signal Systems Laboratory

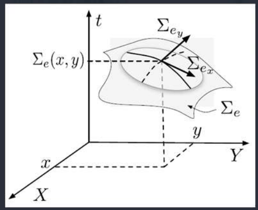
Event Manifold

Local Plane Fitting

- Manifold learned from event cloud
- Regularization used to overcome noise
- Spatial gradient = motion orientation
- Temporal slope = velocity



Surface of Active Events



R. Benosman, C. Clercq, X. Lagorce, S.-H. Ieng, and C. Bartolozzi, “Event-based visual flow,” IEEE transactions on neural networks and learning systems, vol. 25, no. 2, pp. 407–417, 2014.

Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

PROPHESEE METAVISION FOR MACHINES

Intelligent Signal Systems Laboratory

Computational Imaging

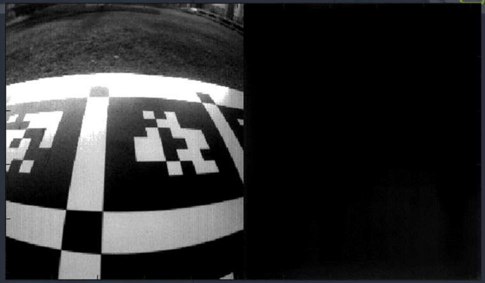
High Speed Frame Reconstruction

Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

PROPHESEE METAVISION FOR MACHINES

Intelligent Signal Systems Laboratory

Frame Reconstruction



APS Frames Frames reconstructed from events

Baldwin, R., Mohammed Almatrafi, Vijayan Asari, and Keigo Hirakawa. "Time-Ordered Recent Event (TORE) Volumes for Event Cameras," https://arxiv.org/abs/2103.06108.

Frame Reconstruction

Goals of Frame Reconstruction

- Estimate image intensity values from events
- Output images that are interpretable

Benefits to Frame Reconstruction

- Faster frame rate than APS
- Higher dynamic range than APS
- Empirical proof that events capture rich spatial information beyond edges
- Apply intensity-based feature extraction, tracking, object classification methods

Disadvantages to Frame Reconstruction

- Dense representation of sparse events => increased computational burden
- Not ideal for real-time high-speed applications
- APS does this better in many/most cases

Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

Event Accumulation

Accumulated Events

unknown initial value

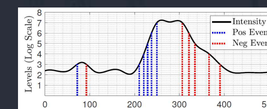
events indicate thresholds were crossed

- $J(x,y,t_1) = J(x,y,t_0) + \tau\, p_1$
- $J(x,y,t_2) = J(x,y,t_0) + \tau\,(p_1 + p_2)$
- $J(x,y,t_3) = J(x,y,t_0) + \tau\,(p_1 + p_2 + p_3)$

threshold values

Challenges

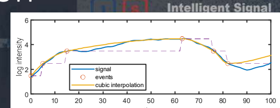
- We don't have $J(x,y,t_0)$
- Only have samples at $t = t_0, t_1, t_2, \ldots$



Mohammed Almatrafi, and Keigo Hirakawa. "DAViS Camera Optical Flow." IEEE Transactions on Computational Imaging, 2019.

Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

Temporal Derivative Interpolation



Cubic Interpolation

- $\hat{J}(x,y,t) = \sum_{k=0}^{3} c_k t^k$

coefficients

- Known points

event polarity

known pixel values

$$\begin{pmatrix} J(x,y,t_0) \\ J(x,y,t_1) \\ J(x,y,t_2) \\ J(x,y,t_3) \end{pmatrix} = J(x,y,t_0) + \begin{pmatrix} 0 \\ p_1 \\ p_1+p_2 \\ p_1+p_2+p_3 \end{pmatrix} = \begin{pmatrix} 1 & t_0^1 & t_0^2 & t_0^3 \\ 1 & t_1^1 & t_1^2 & t_1^3 \\ 1 & t_2^1 & t_2^2 & t_2^3 \\ 1 & t_3^1 & t_3^2 & t_3^3 \end{pmatrix} \begin{pmatrix} c_0 \\ c_1 \\ c_2 \\ c_3 \end{pmatrix}$$

coefficients

event times

- Solve for coefficients

coefficients

$$\begin{pmatrix} c_0 \\ c_1 \\ c_2 \\ c_3 \end{pmatrix} = \begin{pmatrix} 1 & t_0^1 & t_0^2 & t_0^3 \\ 1 & t_1^1 & t_1^2 & t_1^3 \\ 1 & t_2^1 & t_2^2 & t_2^3 \\ 1 & t_3^1 & t_3^2 & t_3^3 \end{pmatrix}^{-1} \begin{pmatrix} J(x,y,t_0) \\ J(x,y,t_1) \\ J(x,y,t_2) \\ J(x,y,t_3) \end{pmatrix} = J(x,y,t_0) \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + \tau \begin{pmatrix} 1 & t_0^1 & t_0^2 & t_0^3 \\ 1 & t_1^1 & t_1^2 & t_1^3 \\ 1 & t_2^1 & t_2^2 & t_2^3 \\ 1 & t_3^1 & t_3^2 & t_3^3 \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ p_1 \\ p_1+p_2 \\ p_1+p_2+p_3 \end{pmatrix}$$

- **Significance:** $(c_1, c_2, c_3)^T$ solvable without $J(x,y,t_0)$. But c_0 not solvable.

Mohammed Almatrafi, and Keigo Hirakawa. "DAViS Camera Optical Flow." IEEE Transactions on Computational Imaging, 2019.

Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

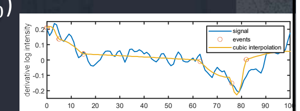
Temporal Derivative Interpolation

Significance: $(c_1, c_2, c_3)^T$ solvable without $J(x,y,t_0)$

Cubic interpolation

- $\hat{J}(x,y,t) = \sum_{k=0}^{3} c_k t^k$
- $\hat{J}_t(x,y,t) = \sum_{k=1}^{3} k c_k t^{k-1}$ (i.e. don't need c_0)

take a derivative



Conclusion

- Temporal derivative can be interpolated from events only

Mohammed Almatrafi, and Keigo Hirakawa. "DAViS Camera Optical Flow." IEEE Transactions on Computational Imaging, 2019.

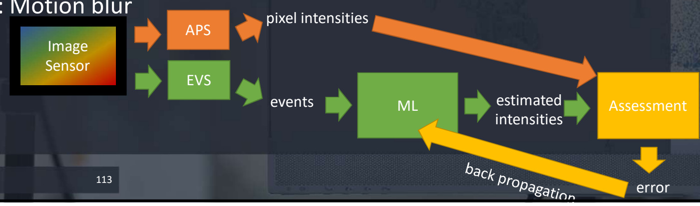
PROPHESEE METAVISION FOR MACHINES

Intelligent Signal Systems Laboratory

APS As Ground Truth

Using APS As Ground Truth

- Pros: Real sensor noise in EVS data
- Pros: Dynamic scenes (foreground/background motion)
- Cons: Limited dynamic range of APS
- Cons: Motion blur



PROPHESEE METAVISION FOR MACHINES

Intelligent Signal Systems Laboratory

Frame Estimation

Use events only to reconstruct frames

- EVS does not capture temporal low frequency data
- EVS does not capture spatial low frequency data
- Penalty term will be dominated by spatial-temporal low frequency.
  - $\sum_{x,y,t} |I(x,y,t) - \hat{I}(x,y,t)|^2$
- Idea: apply highpass filter before computing penalty
  - $\sum_{x,y,t} |I_{HP}(x,y,t) - \hat{I}_{HP}(x,y,t)|^2$

Baldwin, R., Mohammed Almatrafi, Vijayan Asari, and Keigo Hirakawa. "Time-Ordered Recent Event (TORE) Volumes for Event Cameras," https://arxiv.org/abs/2103.06108.

PROPHESEE METAVISION FOR MACHINES

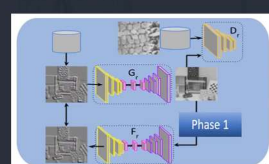
Intelligent Signal Systems Laboratory

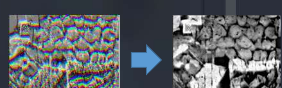
Unsupervised Adversarial Learning

Unsupervised Adversarial Learning

- Gr network predicts APS from events
- Fr network predicts events from APS
- Dr network discriminates APS

- Also designed to perform super resolution.





Wang, Lin, Tae-Kyun Kim, and Kuk-Jin Yoon. "EventSR: From Asynchronous Events to Image Reconstruction, Restoration, and Super-Resolution via End-to-End Adversarial Learning." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 8315-8325. 2020.

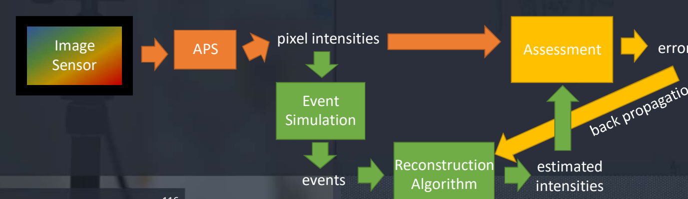
PROPHESEE METAVISION FOR MACHINES

Intelligent Signal Systems Laboratory

Simulation-Based Reconstruction

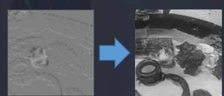
Simulation-Based Ground Truth

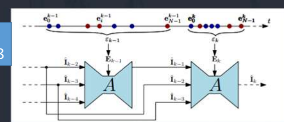
- Pros: Dynamic scenes (foreground/background motion)
- Pros: High fidelity ground truth data
- Cons: No accurate model of noise



Past Frame + New Events

- Inputs to network are…
  - Past 3 reconstructed frames
  - N new events
  - => Promotes consistency between frames
- Used event simulator (ESIM)
  - Ground truth available
  - Full Reference loss function (LPIPS)





Rebecq, Gehrig, Scaramuzza 2018

Zhang, Isola, Efros, Shechtman, Wang 2018

Rebecq, Henri, René Ranftl, Vladlen Koltun, and Davide Scaramuzza. "Events-to-video: Bringing modern computer vision to event cameras." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3857-3866. 2019.

Rebecq, Henri, René Ranftl, Vladlen Koltun, and Davide Scaramuzza.. "High Speed and High Dynamic Range Video with an Event Camera." IEEE Trans. Pattern Anal. Mach. Intell, 2019.

Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

Computational Imaging

High Speed Frame Interpolation and Image Deblurring

Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

APS + Events = High Speed Imaging

Deblurring APS using events

unknown initial value

threshold values

$$J(x,y,t) \approx J(x,y,t_0) + \tau \sum_{i:t_0<t_i<t} p_i$$

event polarity

sum all events between t_0 and t

- APS exposed for Δ_t seconds:

$$APS(x,y) = \int_{t_0}^{t_0+\Delta_t} I(x,y,t)dt = \int_{t_0}^{t_0+\Delta_t} e^{J(x,y,t_0)+\tau\sum_{i:t_0<t_i<t} p_i} dt$$
$$= e^{J(x,y,t_0)} \int_{t_0}^{t_0+\Delta_t} e^{\tau\sum_{i:t_0<t_i<t} p_i} dt$$

blurring

deblurred initial value

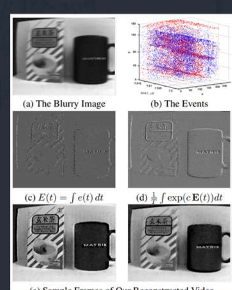
- Solve for $J(x,y,t_0)$

$$J(x,y,t_0) = \log APS(x,y) - \log \int_{t_0}^{t_0+\Delta_t} e^{\tau\sum_{i:t_0<t_i<t} p_i} dt$$

APS + Events = High Speed Imaging

- $J(x,y,t) = J(x,y,t_0) + \tau \sum_{i:t_0<t_i<t} p_i$

deblurred initial value

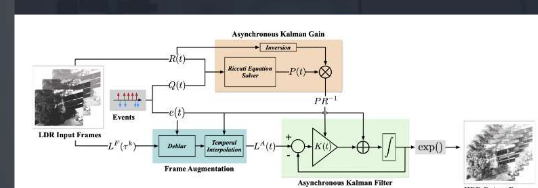


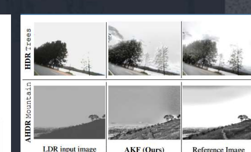
Pan, L., Scheerlinck, C., Yu, X., Hartley, R., Liu, M., Dai, Y., "Bringing a Blurry Frame Alive at High Frame-Rate with an Event Camera," IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2019.

Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

APS + Events = HDR imaging

- Asyncronous Kalman-Bucy filter to reconstruct HDR from LDR frames + Events
  - A frame augmentation to remove blur and increase temporal resolution
  - A Kalman gain estimated pixel-by-pixel that integrates uncertainty models
  - An asynchronous Kalman filter to fuse augmented frames with events streams

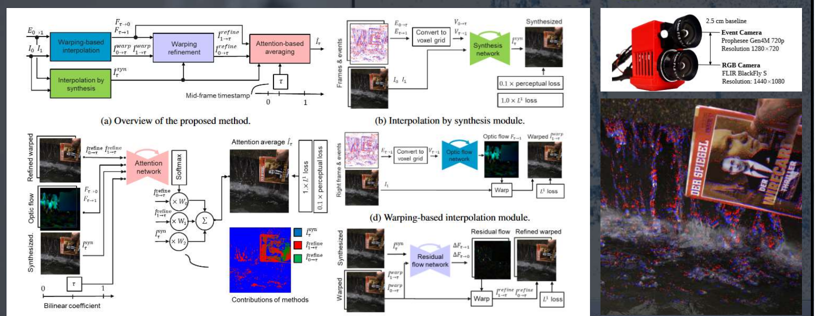




Ziwei Wang, Yonhon Ng, Cedric Scheerlinck and Robert Mahony, IEEE Int. Conf. Computer Vision (ICCV), 2021

Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

Frame + Event-based fusion



S. Tulyakov, D. Gehrig, S. Georgoulis, J. Erbach, M. Gehrig, Y. Li, D. Scaramuzza, Time Lens: Event-based Video Frame Interpolation IEEE Conference on Computer Vision and Pattern Recognition, 2021.

Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

Computer Vision

Event-Based Feature Extraction

Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

Feature Extraction

Feature Extraction

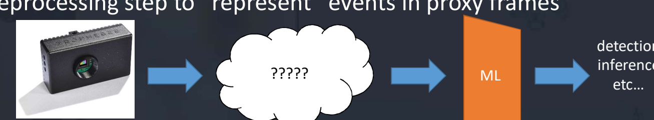
- Feature Detector
  - Identify a primitive spatial intensity pattern (corner, edge, blob, shapes, etc.)
- Feature Descriptor
  - Discriminative vector characterizing the detected intensity pattern
- Feature Matching
  - Identify the occurrence of same features in multiple images/frames by finding similar descriptors
- Match Acceptance Testing
  - Hypothesis testing to reject false matches
  - Photometric Testing: Adaptive thresholding on matched descriptors
  - Geometric Testing: Reject outliers based on homography

Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

Event-Based Feature Matches

Challenges to Event Detection Cameras

- Asynchronous events
- Existing feature extraction techniques rely on intensity patterns
- Events are sensitive to object velocity and camera motion
  - ⇒ Can we achieve invariance or robustness to velocity?
- Preprocessing step to “represent” events in proxy frames



Event Camera

preprocessing

eHarris Corner Detector

Event-Based Harris Corner Detector

- Encode events as binary spatial pattern (last N events)

$$B(x,y)=\begin{cases}1 & \exists\, event\ at\ (x,y)\\ 0 & \nexists\, event\ at\ (x,y)\end{cases}$$

- Compute its spatial derivatives

$$B_x(x,y)=\frac{\partial}{\partial x}B(x,y) \qquad B_y(x,y)=\frac{\partial}{\partial y}B(x,y)$$

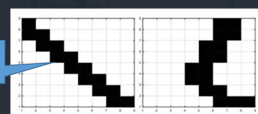
- Compute the intensity-based Harris corner detector

$$M(e_x,e_y)=\sum_{(x,y)} g(x-e_x,y-e_y)\begin{bmatrix}B_x^2 & B_xB_y\\ B_xB_y & B_y^2\end{bmatrix}$$

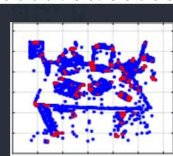
$$R(e_x,e_y)=\det M-k\,trace(M)^2$$

- Corner if $R(e_x,e_y)>\tau$.

1 if event present



Harris & Stephens 1988



V. Vasco, A. Glover, and C. Bartolozzi. "Fast event-based harris cornerdetection exploiting the advantages of event-driven cameras." In2016IEEE/RSJ International Conference on Intelligent Robots and Systems(IROS), pages 4144–4149. IEEE, 2016

Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

Continuous Harris Event Corners (CHEC)

- Apply spatial-temporal filter

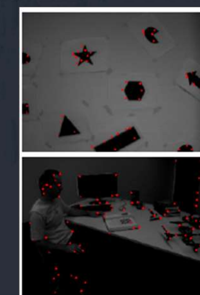
$$K(x,y,e_t)=e^{-\lambda(e_t-t_p))}K(x,y,t_p)+\tau\, e_pH(x-e_x,y-e_y)$$

See earlier slides

- K is Harris corner detector kernels.

Event Camera → Asynchronous Spatial-Temporal Filter → Harris Corner Detector →

Harris & Stephens 1988



C. Scheerlinck, N. Barnes, and R. Mahony, "Asynchronous spatial image convolutions for event cameras," IEEE Robotics and Automation Letters,4(2):816–822, 2019.

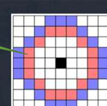
Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

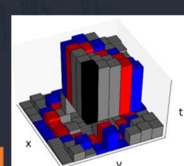
Event-based Features from Accelerated Segment Test (eFAST)

eFAST

- Construct a time surface
- Two concentric Bresenham circles centered at new event
  - Radii 3 and 4
- Find longest contiguous arc with more recent time stamps than the rest
  - If inner arc length is between 3 and 6
  - If outer arc length is between 4 and 8
  - $\Rightarrow$ this is a corner

Bresenham Circles





FAST: Rosten & Drummond 2006

E. Mueggler, C. Bartolozzi, and D. Scaramuzza. "Fast event-based corner detection." In BMVC, 2017.

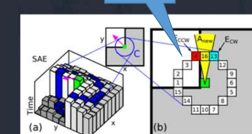
Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

Arc*

Arc*

- Improvement on eFast
- Compute FSAE
- Bresenham circles (radii 3 and 4)
- Longest arc lengths [3,6] and [4,8]
  - $\Rightarrow$ This is a corner
- Longest arc length [10,13] and [12,16]
  - $\Rightarrow$ This is also a corner

corner



I. Alzugaray and M. Chli. "Asynchronous Corner Detection and Tracking for Event Cameras in Real Time." IEEE Robotics and Automation Letters, 2018.

Fast and Asynchronous Corner Detector

PROPHESEE METAVISION FOR MACHINES

Intelligent Signal Systems Laboratory

Fast and Asynchronous Harris Corner Detector (FA-Harris)

Event Camera

Alzugaray & Chli, 2018

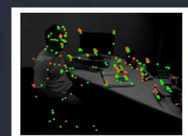
FSAE

eFAST

Mueggler, Bartolozzi, Scaramuzza, 2017

eHarris

Vasco, Glover, Bartolozzi, 2016



129 R. Li, D. Shi, Y. Zhang, K. Li and R. Li, "FA-Harris: A Fast and Asynchronous Corner Detector for Event Cameras," 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Macau, China, 2019, pp. 6223-6229.

Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

Distribution Aware Retinal Transform (DART)

PROPHESEE METAVISION FOR MACHINES

Intelligent Signal Systems Laboratory

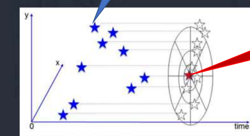
Distribution Aware Retinal Transform (DART)

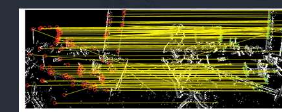
Belongie, Malik, & Puzicha, 2002

- DART is a descriptor
- Partition region around event in log-polar grids (these are histogram bins)
- Make a weighted histogram
  - Use N last events
  - Record bilinear weights to neighboring bins

past N events

new event





(b) Indoor scene with scale and rotation changes.

130 Ramesh, Bharath, Hong Yang, Garrick Michael Orchard, Ngoc Anh Le Thi, Shihao Zhang, and Cheng Xiang. "Dart: distribution aware retinal transform for event-based cameras." IEEE transactions on pattern analysis and machine intelligence (2019).

Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

Distance Transform Descriptor

PROPHESEE METAVISION FOR MACHINES

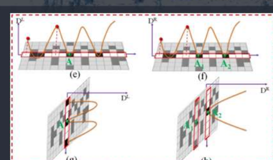
Intelligent Signal Systems Laboratory

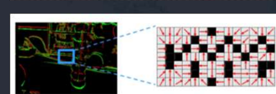
Local Context Aware Descriptor

- Compute 1D horizontal distance transform
- Compute 1D vertical distance transform

Signed Vector Distance Transform

- For each pixel, identify the closest event.
- Form a vector field pointing to the direction to the closest event.
- Quantize vector field by its sign.





131 Zou, D., Shi, F., Liu, W., Li, J., Wang, Q., Park, P.K.J., Shi, C.-W., Roh, Y.J., Ryu, H., Robust Dense Depth Map Estimation from Sparse DVS Stereos, British Machine Vision Conf. (BMVC), 2017

Zou, Dongqing, Ping Guo, Qiang Wang, Xiaotao Wang, Guangqi Shao, Feng Shi, Jia Li, and Paul-KJ Park. "Context-aware event-driven stereo matching." In 2016 IEEE International Conference on Image Processing (ICIP), pp. 1076-1080. IEEE, 2016.

Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

Event-Blob Feature Extraction

PROPHESEE METAVISION FOR MACHINES

Intelligent Signal Systems Laboratory

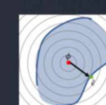
EventDistBlob Detector

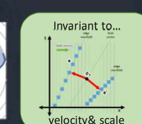
- Blob boundaries denoted by events.
- Blob centers are farthest from these events
- Distance transform

$$D(x,y,t) = \min_e \sqrt{(e_x - x)^2 + (e_y - y)^2 + (e_t - t)^2}$$

- Event Blob Detection = spatial local maximum of $D(x,y,t)$

$$\Phi = \left\{(x,y,t) \middle| \begin{matrix} D(x,y,t) > D(x+\kappa, y+\epsilon, t) \\ \forall \kappa^2 + \epsilon^2 < \tau \end{matrix} \right\}$$





132 J. Raffoul, M. Almatrafi, K. Hirakawa. "EventDistBlob: EventDistBlob: Event-Based Blob Feature Extraction using Asynchronous Distance Transform." Arxiv, 2021.

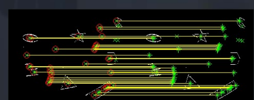
Event-Blob Feature Extraction

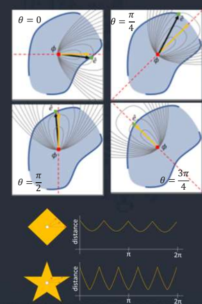
EventDistBlob Descriptor

- Polar coordinates centered at blob center
- Directionally biased distance transform

$$D_\theta(x,y,t) = \min_e \sqrt{(e_x-x)^2+(e_y-y)^2+(e_t-t)^2+(e_\theta-\theta)^2}$$

- Robust to noise and variations





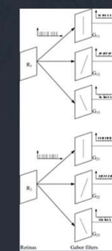
133 J. Raffoul, M. Almatrafi, K. Hirakawa. "EventDistBlob: EventDistBlob: Event-Based Blob Feature Extraction using Asynchronous Distance Transform." Arxiv, 2021.

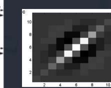
Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

Gabor Filter Edge Descriptor

Gabor Filter Edge Descriptor

- Gabor filters are sensitive to edges in predefined orientations.
- Gabor filter responses describes edge orientation.
- Direct application of Gabor filters to edge pixels yield event edge orientation.

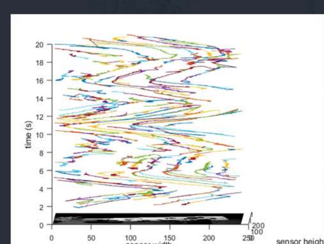




134 Camunas-Mesa, Luis Alejandro, Teresa Serrano-Gotarredona, Sio Hoi Ieng, Ryad Benjamin Benosman, and Bernabe Linares-Barranco. "On the use of orientation filters for 3D reconstruction in event-driven stereo vision." Frontiers in neuroscience 8 (2014): 48.

Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

Event-Based Tracking Example



135 J. Raffoul, M. Almatrafi, K. Hirakawa. "EventDistBlob: EventDistBlob: Event-Based Blob Feature Extraction using Asynchronous Distance Transform." Arxiv, 2021.

Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

Computer Vision

Event-Based Egomotion, Visual Odometry and SLAM

136

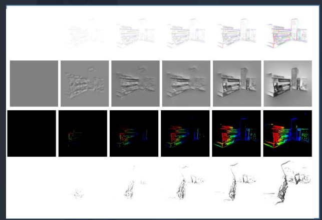
Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

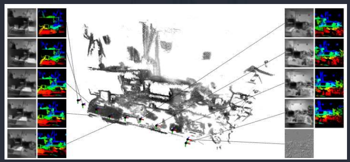
PROPHESEE METAVISION FOR MACHINES

Intelligent Signal Systems Laboratory

From Events to Frames to 3D

- EKF to estimate 6DoF camera pose with Inverse Depth
- Reconstruction of image-like log intensity





137 Kim, H., Leutenegger, S., Davison, A.J., Real-Time 3D Reconstruction and 6-DoF Tracking with an Event Camera, European Conference on Computer Vision (ECCV), 2016, pp. 349-364

Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

PROPHESEE METAVISION FOR MACHINES

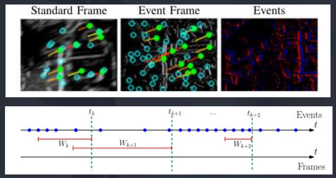
Intelligent Signal Systems Laboratory

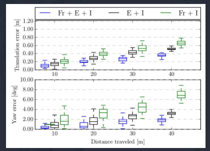
Event+Frames+IMU

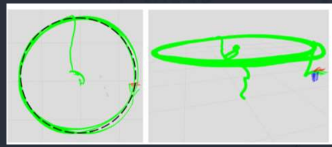
- Visual-Inertial Fusion with Nonlinear optimization

$$J = \sum_{i=0}^{1}\sum_{k=1}^{K}\sum_{j\in\mathcal{J}(i,k)} \mathbf{e}^{i,j,k^T}\mathbf{W}_r^{i,j,k}\mathbf{e}^{i,j,k} + \sum_{k=1}^{K-1}\mathbf{e}_s^{k^T}\mathbf{W}_s^k\mathbf{e}_s^k$$

$$\mathbf{e}_r^{i,j,k} = \mathbf{z}^{i,j,k} - \pi_i\left(\mathbf{T}_{C_i}^k\mathbf{T}_{SW}^k 1^{i,j}\right)$$







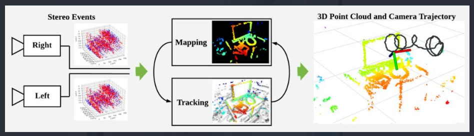
138 T. Rosinol Vidal, H.Rebecq, T. Horstschaefer, D. Scaramuzza
Ultimate SLAM? Combining Events, Images, and IMU for Robust Visual SLAM in HDR and High Speed Scenarios
IEEE Robotics and Automation Letters (RA-L), 2018.

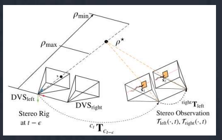
Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

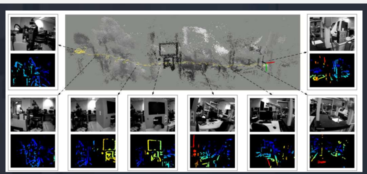
PROPHESEE METAVISION FOR MACHINES

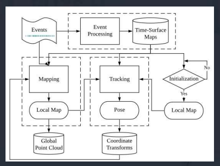
Intelligent Signal Systems Laboratory

Event Based Stereo Visual Odometry









139 Yi Zhou, Guillermo Gallego, Henri Rebecq, Laurent Kneip, Hongdong Li, and Davide Scaramuzza. "Semi-dense 3D reconstruction with a stereo event camera." In Proceedings of the European Conference on Computer Vision (ECCV), pp. 235-251. 2018.

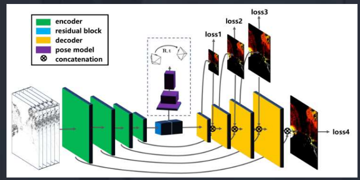
Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

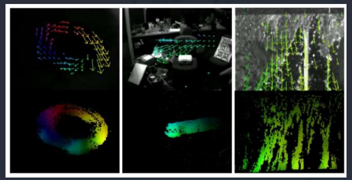
PROPHESEE METAVISION FOR MACHINES

Intelligent Signal Systems Laboratory

Unsupervised Learning

- Learning motion from events stream
  - Input as discretized volume and NN to predict motion
  - This motion to remove blur – loss on motion compensated events volume representation
  - 2 Networks: one to predict optical flow and one to predict egomotion and depths

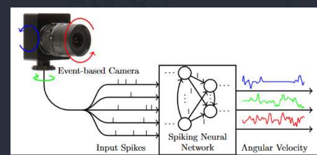


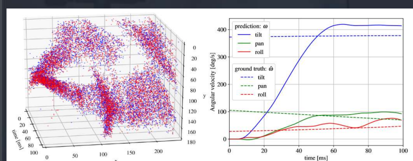


140 Alex Zihao Zhu, Liangzhe Yuan, Kenneth Chaney, Kostas Daniilidis, CVPR 2019,
Unsupervised Event-Based Learning of Optical Flow, Depth, and Egomotion.

Event + SSN for Angular velocity

- Convolutional feedforward SSN
  - 5 conv layers + pooling fully connected
  - Global Average Spike Pooling (GASP)



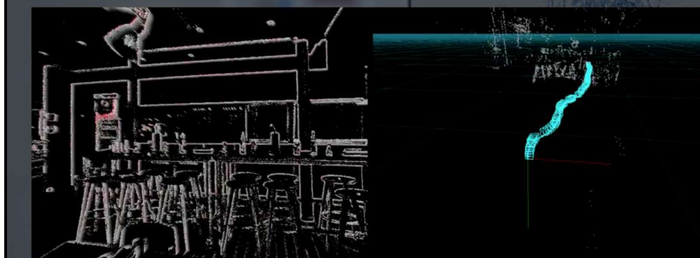


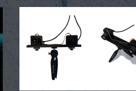
HYPERPARAMETERS OF THE SPIKING NEURAL NETWORK ARCHITECTURE.

Layer-type	Conv 1	Conv 2	Conv 3	Conv 4	Conv 5	Fully connected
Kernel size	3×3	3×3	3×3	3×3	3×3	-
Channels	16	32	64	128	256	-
Stride	2	2	2	2	1	-
$\tau_s,\ \tau_r$ [ms]	2, 1	2, 1	4, 4	4, 4	4, 4	8, -

Gehrig, M., Shrestha, S. B., Mouritzen, D., Scaramuzza, D., Event-Based Angular Velocity Regression with Spiking Networks, IEEE Int. Conf. Robotics and Automation (ICRA), 2020

ICP on 3D points clouds





Prophesee demo IROS 2018

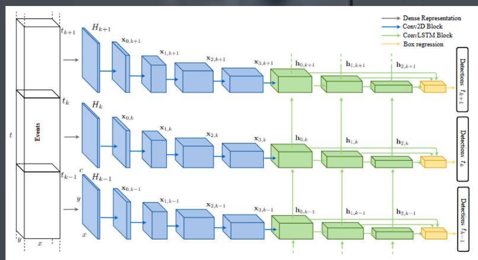
Workshop competition on Stereo

Computer Vision

Event-Based Object Detection and Classification

Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

LSTM with Events



	1Mpx Detection Dataset			Gen1 Detection Dataset	
	mAP	runtime (ms)	params (M)	mAP	runtime (ms)
MatrixLSTM [50]	-	-	-	0.31*	-
SparseConv [13]	-	-	-	0.15	-
Events-RetinaNet	0.18	44.05	32.8	0.34	18.29
E2Vid-RetinaNet	0.25	840.66	43.5	0.27	263.28
RED (ours)	**0.43**	**39.33**	**24.1**	0.40	**16.70**
Gray-RetinaNet	**0.43**	41.43	32.8	**0.44**	17.35

* Provided by the authors, using a pretrained YOLOv3.

145

Perot, E., de Tournemire, P., Nitti, D., Masci, J., Sironi, A.,
Learning to Detect Objects with a 1 Megapixel Event Camera,
Advances in Neural Information Processing Systems 33 (NeurIPS), 2020.

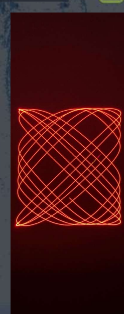
Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

Applications

146

Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

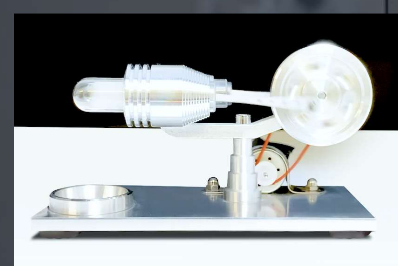
Ultra slow motion



147

Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

Industrial Automation



148

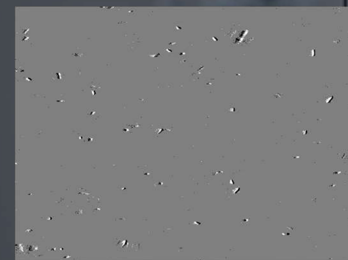
Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

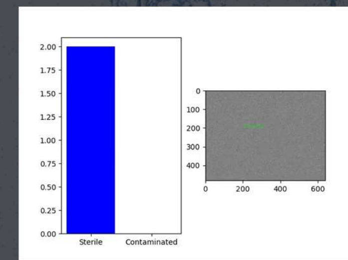
Same algorithm different use case

- Credit: IQT blog - https://www.iqt.org/exploring-and-experimenting-with-event-cameras/

Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

Biology





- Credit: Cambridge Consultants – PureSentry®

Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

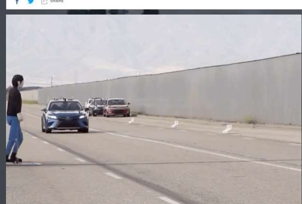
Automotive

THE VERGE

TECH \ TRANSPORTATION \ CARS

Cars with high-tech safety systems are still really bad at not running people over

AAA brought receipts



« Vehicles struck the dummy pedestrians (...) 60 % of the time (...)in daylight hours at speeds of 20 mph. »

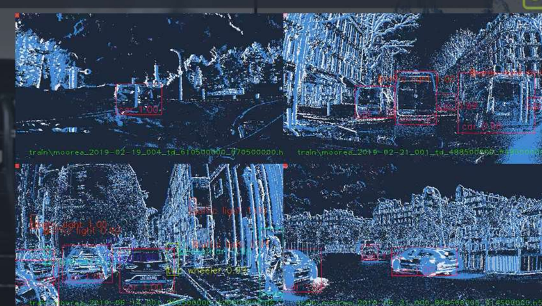
« With a child-sized version, the results got much, much worse: a collision occurred 89% of the time. »

« None of the cars tested were able to detect an adult pedestrian at night. »

151

Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

Automotive



- Credit: "Learning to Detect Objects with a 1 Megapixel Event Camera" – E. Perot et al. – to appear at NeurIPS 2020

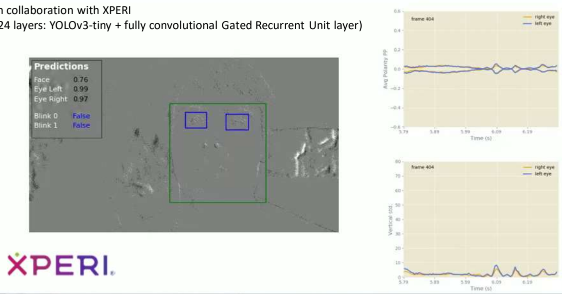
152

PROPHESEE
METAVISION FOR MACHINES

Intelligent Signal Systems Laboratory
PI: Keigo Hirakawa, University of Dayton

Automotive

In collaboration with XPERI
(24 layers: YOLOv3-tiny + fully convolutional Gated Recurrent Unit layer)



- Credit: Xperi

Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

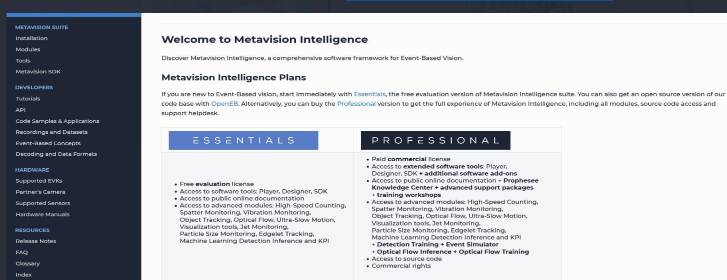
"Quick Start Guide"

Hardware Demonstration

Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

Introduction

- All the material is available here: https://docs.prophesee.ai



METAVISION SUITE
Installation
Modules
Tools
Metavision SDK

DEVELOPERS
Tutorials
API
Code Samples & Applications
Recordings and Datasets
Event-Based Concepts
Decoding and Data Formats

HARDWARE
Supported EVKs
Partner's Camera
Supported Sensors
Hardware Manuals

RESOURCES
Release Notes
FAQ
Glossary
Index

Welcome to Metavision Intelligence

Discover Metavision Intelligence, a comprehensive software framework for Event-Based Vision.

Metavision Intelligence Plans

If you are new to Event-Based vision, start immediately with Essentials, the free evaluation version of Metavision Intelligence suite. You can also get an open source version of our code base with OpenEB. Alternatively, you can buy the Professional version to get the full experience of Metavision Intelligence, including all modules, source code access and support helpdesk.

ESSENTIALS

- Free **evaluation** license
- Access to software tools: Player, Designer, SDK
- Access to public online documentation
- Access to advanced modules: High-Speed Counting, Spatter Monitoring, Vibration Monitoring, Object Tracking, Optical Flow, Ultra-Slow Motion, Visualization tools, Jet Monitoring, Particle Size Monitoring, Edgelet Tracking, Machine Learning Detection Inference and KPI

PROFESSIONAL

- Paid **commercial** license
- Access to **extended software tools**: Player, Designer, SDK **+ additional software add-ons**
- Access to public online documentation + **Prophesee Knowledge Center + advanced support packages + training workshops**
- Access to advanced modules: High-Speed Counting, Spatter Monitoring, Vibration Monitoring, Object Tracking, Optical Flow, Ultra-Slow Motion, Visualization tools, Jet Monitoring, Particle Size Monitoring, Edgelet Tracking, Machine Learning Detection Inference and KPI **+ Detection Training + Event Simulator + Optical Flow Inference + Optical Flow Training**
- Access to source code
- Commercial rights

Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

Let's introduce the sensor

CHARACTERISTICS HD

Supplier	PROPHESEE	Case	Aluminium	MECHANIC	
Year	2021	USB	Type C	EVK dimensions	102mm x 58mm x 42mm
Resolution (px)	1280 x 720	Trigger In	MCX	Weight	260g (excl. optic)
Latency (µs)	220	Sync In	MCX	OPTIC	
Dynamic Range (dB)	>110	Sync/Trig Out	MCX	D-FOV	81.5°
Nominal contrast treshold (%)	25	Camera power (W)	7.5	MOUNT	Foctek 5mm CS or S
Pixel size (µm)	4.86 x 4.86	DC in supply	12V 3A 2.1mm jack	ADD. INFO	
Camera Max. Bandwidth (Mbps)	1066			Power	DC in for non type C host

Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

PROPHESEE METAVISION FOR MACHINES

Intelligent Signal Systems Laboratory

Let's introduce the sensor

Live Demo

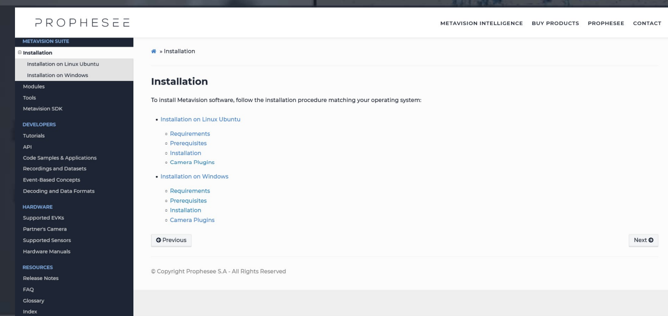
157

Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

PROPHESEE METAVISION FOR MACHINES

Intelligent Signal Systems Laboratory

Install the software



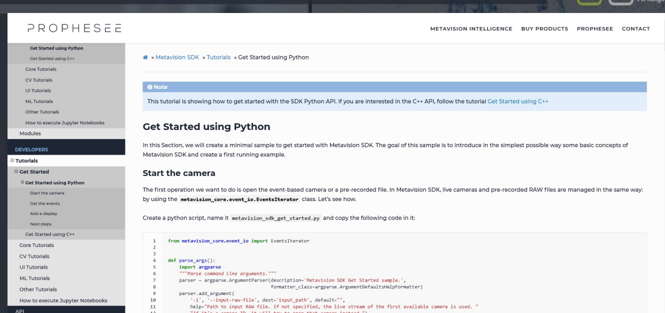
158

Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

PROPHESEE METAVISION FOR MACHINES

Intelligent Signal Systems Laboratory

Use the Python library



159

Keigo Hirakawa (UDayton) Davide Migliore (Prophesee)

PROPHESEE METAVISION FOR MACHINES

Intelligent Signal Systems Laboratory

Use the Python library

Live Demo

160

Conclusion

Conclusion

This short course covered fundamentals of event detection cameras

- EVS hardware, data acquisition, camera parameterization
- EVS forward models, simulation methods
- Event representation, event filtering
- Low-level event processing tasks
  - Event feature extraction
  - Frame reconstruction
  - Optical flow
- Applications

Resources

There is so much more!

Survey Paper

- Gallego, G., Delbruck, T., Orchard, G., Bartolozzi, C., Taba, B., Censi, A., Leutenegger, S., Davison, A., Conradt, J., Daniilidis, K., Scaramuzza, D., "Event-based Vision: A Survey," IEEE Trans. Pattern Anal. Machine Intell. (TPAMI), 2020.

Comprehensive Paper Archive

- https://github.com/uzh-rpg/event-based_vision_resources

Tutorial and Courses

- Event-Based Robot Vision (TU Berlin):
  https://sites.google.com/view/guillermogallego/teaching/bio-inspired-computer-vision
  https://www.youtube.com/channel/UCuulEatPpBZQkQ_kSoeZtCQ/videos
- Bio-inspired Computer (TU Berlin):
  https://sites.google.com/view/guillermogallego/teaching/bio-inspired-computer-vision

Resources

Where to buy an Event-Camera?

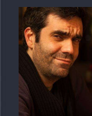
- Prophesee: https://www.prophesee.ai/event-based-evaluation-kits/
- Framos: https://www.framos.com/en/product-catalog/sensors/event-based-sensors
- Other: https://github.com/uzh-rpg/event-based_vision_resources#devices

People to Thank

- *You!* For attending this Short Course
- Computer Vision Foundation (CVF)
- IEEE Computer Society
- Organization Committee of **ICCV 2021**
- Students and Collaborators at the **University of Dayton**
  - Wes Baldwin, Mohammad Almatrafi, Joseph Raffoul, Luc Tinch, Osama Alsattam, Abdulmajeed Alyazidi, Abigail Wolf, Ryan Goetemoeller, Vijayan Asari, Sid Gunasekaran
- Collaborators
  - Scott McCloskey (Kitware), Shannon Brooks-lehnert

Thank You!

Keigo Hirakawa
University of Dayton
Professor
khirakawa1@udayton.edu



Davide Migliore
Prophesee
Technical Business Developer
Senior Computer Vision Engineer
dmigliore@prophesee.ai

Questions and Discussions

- What potential new applications are there for event detection cameras?
  - **Pros:** fast, high dynamic range, low power, bio-inspired, sparse
- Smart/intelligent pixels
  - Event detection cameras won't replace APS camera.
  - But when does event camera make more sense than APS?
- Machine Learning + Event Data
  - How best to represent event data for CNN?
  - What CNN architecture is appropriate for event data?
- Sensor Fusion
  - Event detection camera + APS
  - Event detection camera + Lidar, radar, ultrasound, etc…
  - Event detection camera + ???