Technische Universität Berlin

EINSTEIN CENTER Digital Future

Event-based Robot Vision

Prof. Dr. Guillermo Gallego
Chair: Robotic Interactive Perception

guillermo.gallego@tu-berlin.de

http://www.guillermogallego.es

Optical Flow

What is Optical Flow?

- A technical topic with > 40 years of research
- It is the **apparent motion** of pixels on the image plane. Find it without knowledge about scene geometry or motion
- What is a “flow field”?

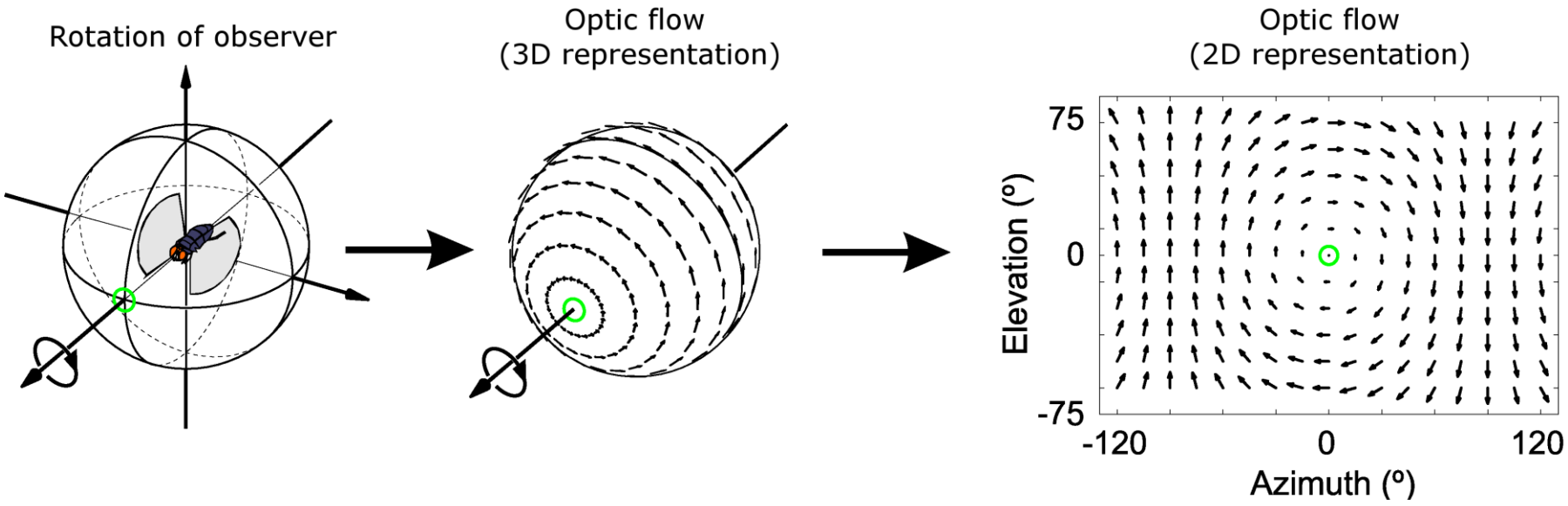
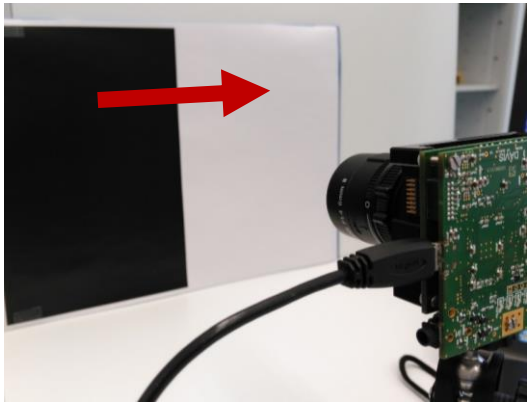


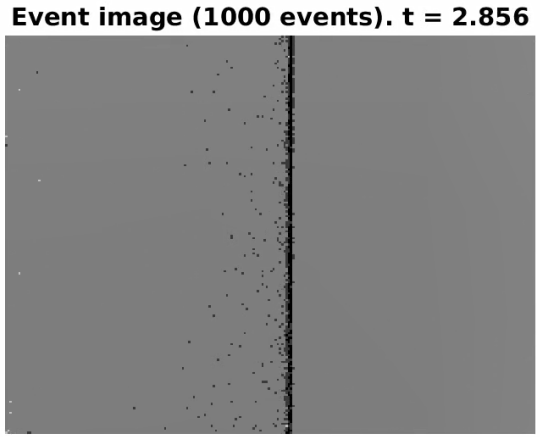
Image from: https://en.wikipedia.org/wiki/Optical_flow

A moving edge

Consider the motion of an edge, viewed by an event-based camera:

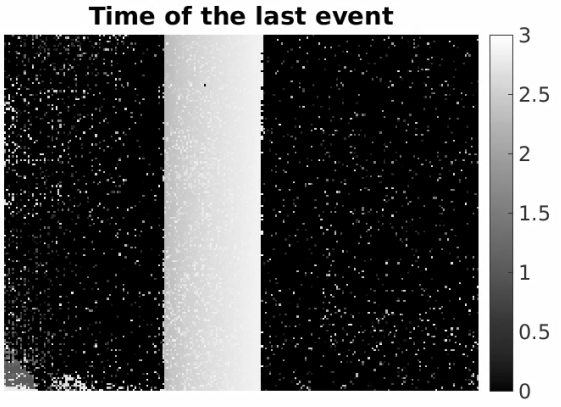


White pixels become black



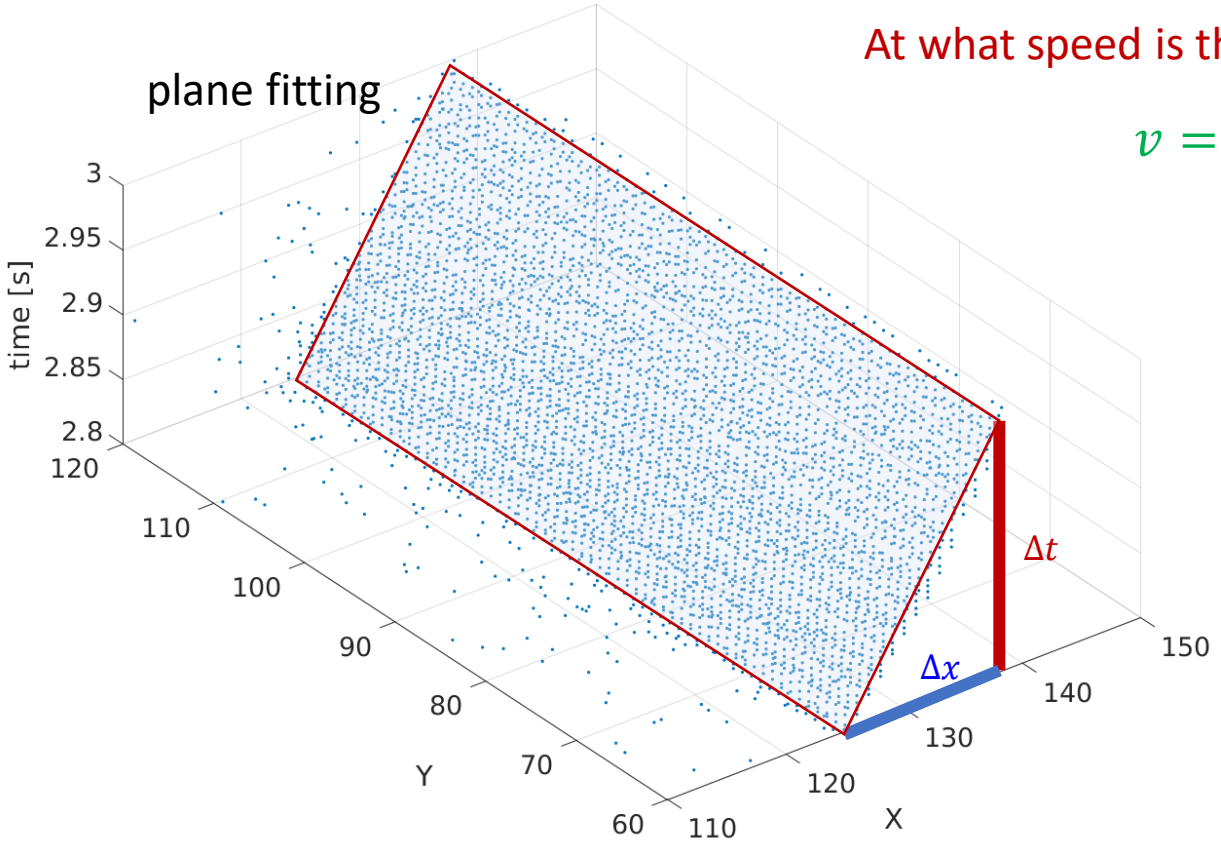
→ brightness decrease

→ negative events (in black color)



A moving edge

The same edge, visualized in space-time
Events are represented by dots (point “cloud")

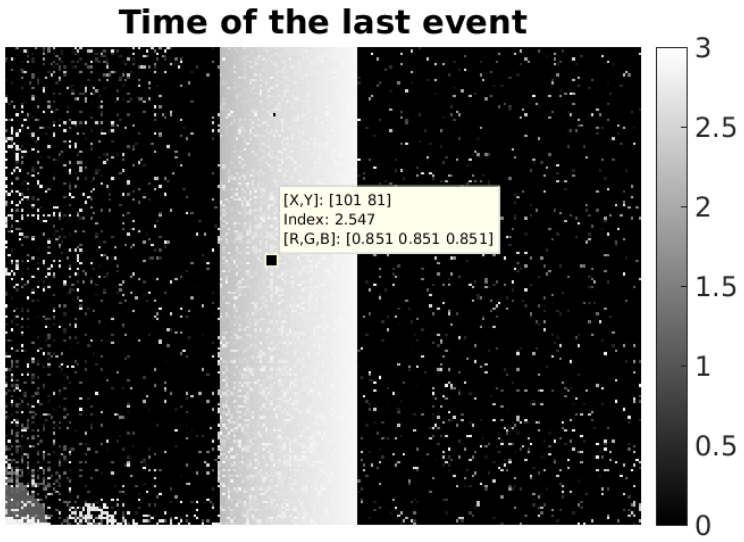


At what speed is the edge moving?

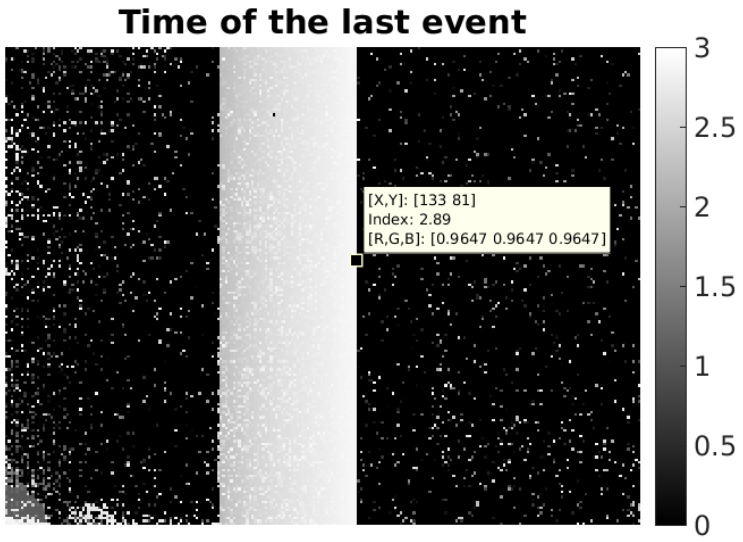
$$v = \frac{\Delta x}{\Delta t}$$

A moving edge

At t = 2.547, x = 101



At t = 2.89, x = 133

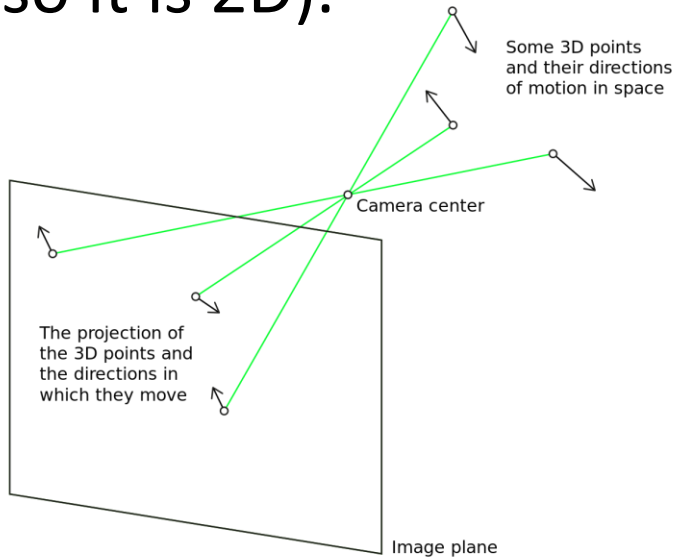


Speed of the edge:

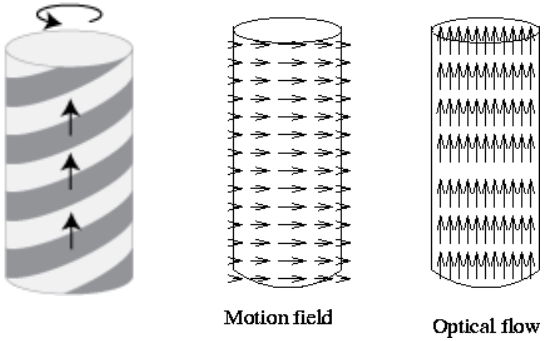
$$v = \frac{(133 - 101)}{(2.89 - 2.547)} = 93.3\ pix/sec$$

What is Optical Flow?

- **Optical Flow** vs. **Motion Field**
- **Motion field** is the physical motion (3D) projected onto the image plane (so it is 2D).



- OF and MF may differ: barber-pole illusion.

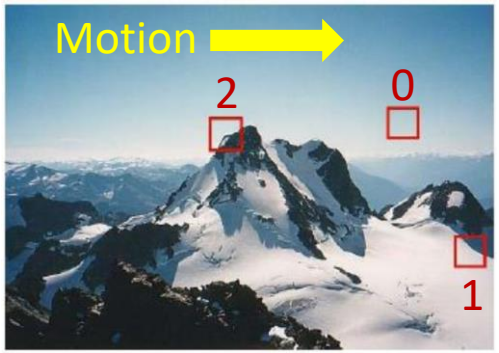


https://en.wikipedia.org/wiki/Motion_field

Image by Robyn Owens

Where is optical flow ≈ motion field?

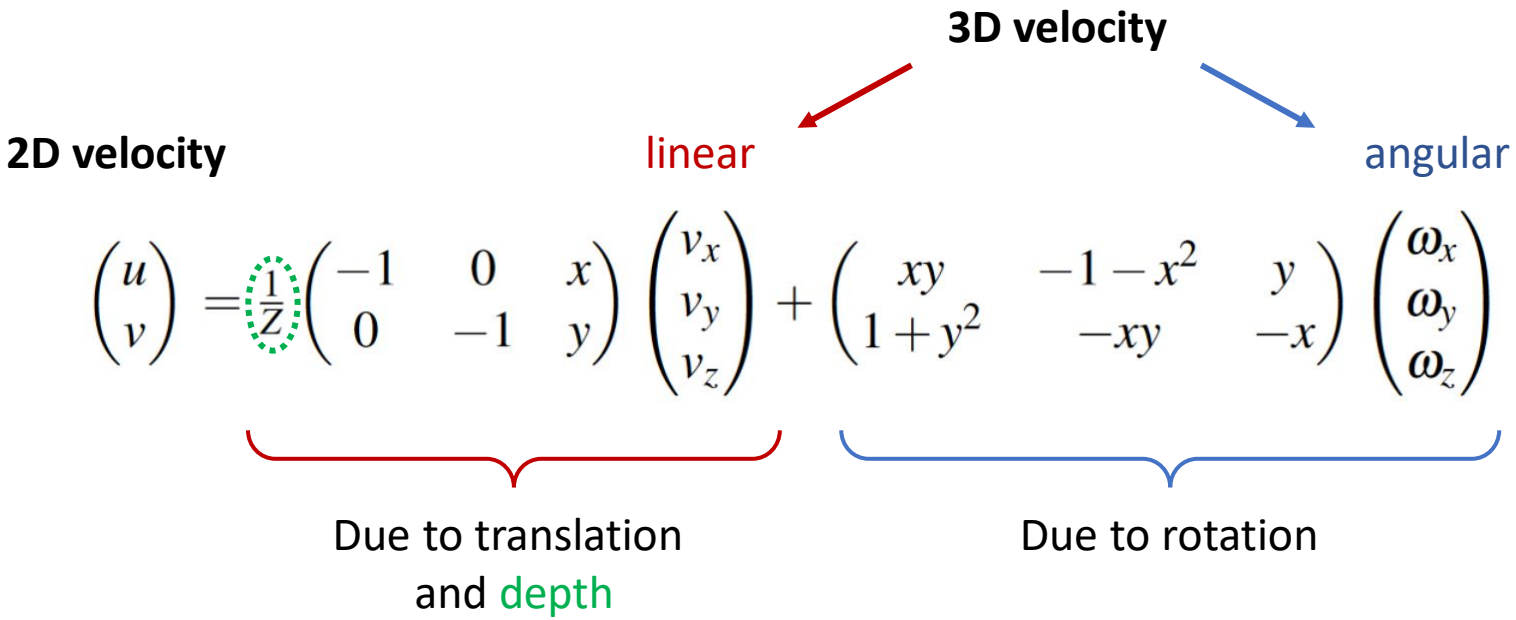
- Three types of neighborhoods of image points:



  - **Region of constant brightness**: "impossible" to determine optical flow. Algorithms provide a solution based on priors (regularizer).
  - **Single edge (1D)**: can only determine the optical flow perpendicular to the edge. (Normal component of motion field, called normal flow)
  - **Corner or complex edge pattern**: brightness gradients in more than one direction. Unambiguous to determine motion
- In all three cases there is a motion field vector, but only in one case the optical flow is a good approximation of it (case 2).
- We are omitting noise, occlusions, etc. in this analysis.

Motion field

- The motion field has two components:
  - **Rotational** component (independent of depth Z)
  - **Translational** component (depends on depth Z). Parallax



x,*y*: pixel coordinates

Brightness Constancy

- It assumes that brightness on the image plane does not change for projections of the same object (3D) point:

$$L(\boldsymbol{x}(t), t) = const$$

  for all points $\boldsymbol{x}(t)$ corresponding to an object point.

- In practice, it holds for short time (differential viewpoint):

$$0 = \frac{dL(\boldsymbol{x}(t), t)}{dt} = \nabla L \cdot \frac{d\boldsymbol{x}}{dt} + \frac{\partial L}{\partial t}$$

  where $\dot{\boldsymbol{x}} = d\boldsymbol{x}/dt$ is the point's velocity $\boldsymbol{v}$, i.e., the **optical flow.**

- If L is known, it gives 1 equation in 2 unknowns → underdetermined
- **Problem**: events cameras do not provide L. How do we get around?

Frame-based vs. event-based Optical Flow

- **Similarities**?
  - Some try to exploit common **principles** (e.g., brightness constancy), even if that produces different methods.
  - Suffer from similar fundamental **problems:**
    - The **aperture problem** does not vanish by having events instead of frames
    - (But frame-based acquisition suffers from motion blur. Events do not)
  - Computing optical flow on every pixel and time (i.e., each point in space-time) is **expensive**.

Frame-based vs. event-based Optical Flow

- **Differences**? Different characteristics of signals:

  - **Geometry**: synchronous and dense vs. asynchronous and sparse
  - **Photometry**: absolute intensity vs. temporal contrast
  - **Noise**: event noise vs. intensity noise

  - **Scenarios**: flow in high-speed & HDR scenarios
  - **Technology**: maturity

Why is event-based optical flow attractive?

- Events allow to obtain flow in **high speed** and **HDR** scenarios, and at **low power**

- Potentially **more efficient** computations by focusing on edges (the informative regions of the image plane)

- Potentially implement in neuromorphic hardware (**biologically plausible** computational model).
  Trying to understand Nature

Opportunities. What's missing?

- **Datasets and benchmarks** (metrics and protocols) for comparison that foster progress in the field, to advance the state of the art.
  - Ground truth optical flow is difficult to obtain.
  - We may use as ground truth the motion field from **simulated** 3D scenes and camera motions. However, event noise is not modeled well yet. Real-world data is needed.
  - Metrics: Quantify **trade-offs** in accuracy, efficiency, latency, etc.
- A **thorough comparison** of multiple methods in a variety of scenes (texture, illumination) and motions (speed, occlusions, parallax, etc.) that will allow us to **identify key ideas** and **best practices**.

(Discussion):

- This is an emerging field of research (> 2008).
  We are still in an **exploratory phase** of methods.
- Presumably, data-driven methods will soon dominate, as it has happened in conventional computer vision.

Literature Review

Questions for each method?

- **Key principle** (main idea of the method):
  - **Assumptions**. What is the **additional knowledge**?
  - What is **being optimized**?
  - Differential method, correlation-based, phase-based, based on event generation model, data-driven (CNN), ...
  - What event **representations** are used (at input or intermediate)?
  - What is the **inference mechanism** to compute flow given events?
- **Characteristics**:
  - Output: full flow vs normal flow?
  - Output: dense vs. sparse?
  - Method: Event-by-event vs. packet of events?
  - Method: Bio-inspired vs not?
  - Method: Does it exploit polarity? yes/no
- **Pros / Cons** (limitations)
  - Supported by **experiments**?

Taxonomy of some event-based OF methods

- Multiple dimensions (criteria) can be considered

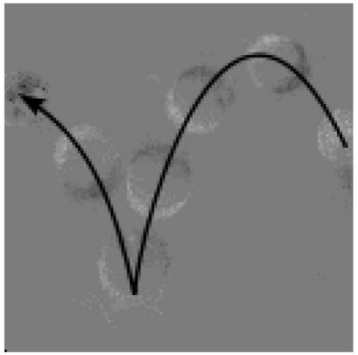
Table 2
Classification of several optical flow methods according to their output and design. Some methods provide full motion flow (F) whereas others only its component normal to the local brightness edge (N). The output may be a dense (D) flow field (i.e., optical flow for every pixel at some time) or sparse (S) (i.e., flow computed at selected pixels). According to their design, methods may be model-based or model-free (Artificial Neural Network - ANN), and neuro-biologically inspired or not.

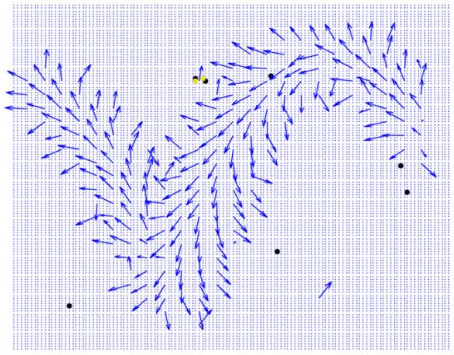
Reference	N/F?	S/D?	Model?	Bio?
Delbruck [80], [190]	Normal	Sparse	Model	Yes
Benosman et al. [190], [191]	Full	Sparse	Model	No
Orchard et al. [157]	Full	Sparse	ANN	Yes
Benosman et al. [21], [190]	Normal	Sparse	Model	No
Barranco et al. [192]	Normal	Sparse	Model	No
Barranco et al. [193]	Normal	Sparse	Model	No
Conradt et al. [194]	Normal	Sparse	Model	No
Brosch et al. [134], [195]	Normal	Sparse	Model	Yes
Bardow et al. [117]	Full	Dense	Model	No
Liu et al. [105]	Full	Sparse	Model	No
Gallego [128], Stoffregen [154]	Full	Sparse	Model	No
Haessig et al. [196]	Normal	Sparse	ANN	Yes
Zhu et al. [22], [119]	Full	Dense	ANN	No
Ye et al. [153]	Full	Dense	ANN	No
Paredes-Vallés [102]	Full	Sparse	ANN	Yes

Gallego et al., *Event-based Vision: A Survey*, TPAMI 2020, Section 4.2

Adaptation of Frame-based methods

- Recall **brightness constancy**: $\nabla L \cdot \frac{dx}{dt} + \frac{\partial L}{\partial t} = 0$
- At each pixel: 1 eq., 2 unknows → Lucas-Kanade (LK) assumed **flow is locally constant** → more equations in the same 2 unknowns.
- Event camera does not provide L …
  - Adapted LK (2012) proposes to **use increment image of events** in place for L.



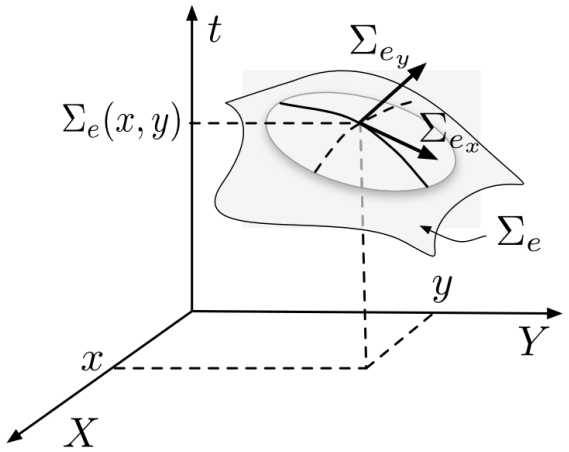


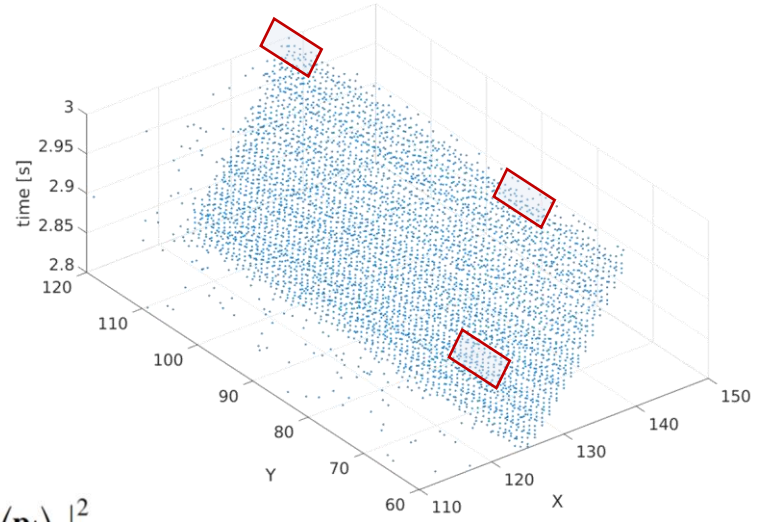
- Critique by Brosch et al (2015):
  - Differential methods do not work so well because computing derivatives with only few events per pixel is not reliable.
  - **Methods working on the event cloud** directly (without derivat.) **are preferred**

Benosman et al., *Asynchronous Frameless Event-Based Optical Flow*, Neural Networks 2012

Optical Flow by Local Plane Fitting

- **Idea**: fit planes locally (in space-time) to the point cloud of events
- Why? A moving edge produces a trail of events that resembles a surface. Fit a plane, get the velocity from the plane coefficients.





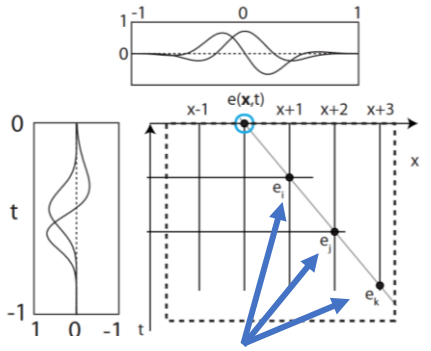
$$\nabla \Sigma_e = \left(\frac{1}{v_x}, \frac{1}{v_y} \right) \qquad \tilde{\Pi}_0 = \underset{\Pi \in \mathbb{R}^4}{\operatorname{argmin}} \sum_i \left| \Pi^T \begin{pmatrix} \mathbf{p}_i \\ t_i \\ 1 \end{pmatrix} \right|^2$$

- Can only determine the **normal component** of the flow
- Sparked the concept of event "lifetime"

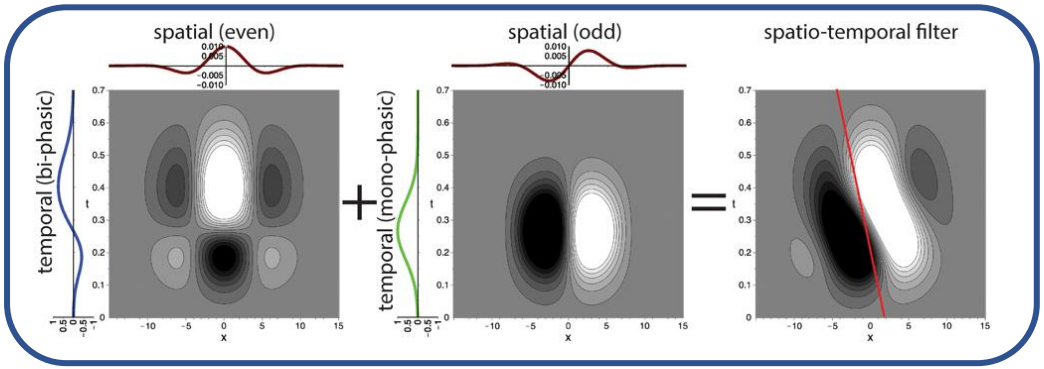
Benosman et al., *Event-Based Visual Flow*, TNNLS 2014. PDF

Space-time Direction-Selective Filter Bank

- **Idea**: Convolve events with **space-time filters** of selected response
  - Biologically / physiologically inspired
- Build **motion-direction sensitive filters** from **separable** components



Trail of events

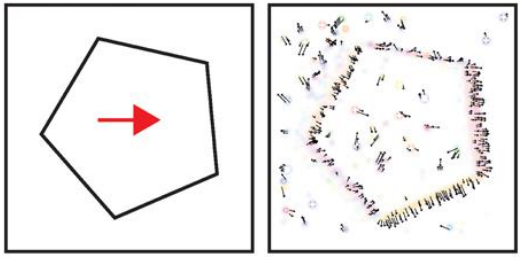


(one of the filters)

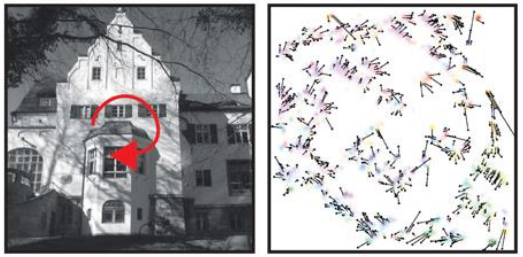
Maximum response

- Results:

Results on translation



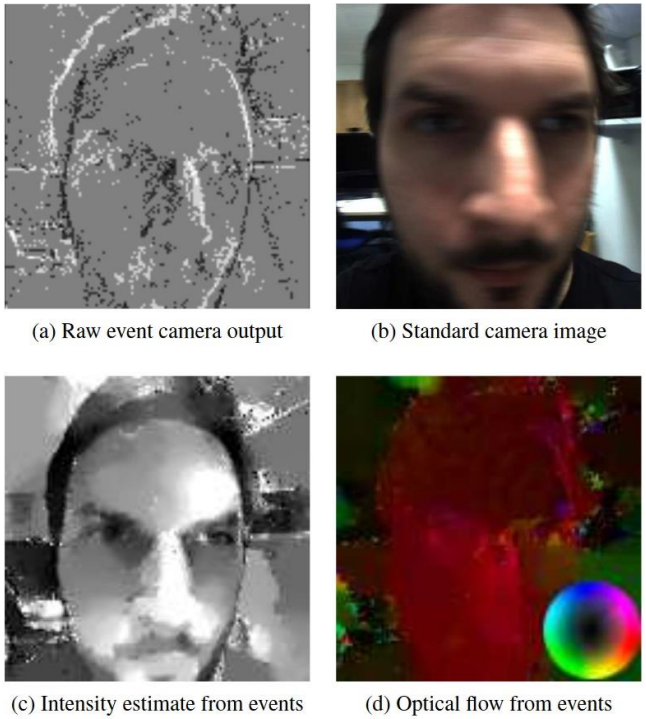
Results on rotation



Brosch et al., *On event-based optical flow detection*, Front. Neurosci. 2015

SOFIE: Simultaneous Optical Flow & IE

- **Joint optimization** over **image brightness** and **optical flow** to explain a volume of events (voxel grid)
- **Idea**: Penalize **deviations** for all equations (brightness constancy and event generation model) and assume **smooth** solution (***u***, ***L***).
- Solve a variational optimization problem:



(a) Raw event camera output (b) Standard camera image

(c) Intensity estimate from events (d) Optical flow from events

$$\min_{\mathbf{u},L} \int_\Omega \int_T \Big(\lambda_1 \|\mathbf{u}_\mathbf{x}\|_1 + \lambda_2 \|\mathbf{u}_t\|_1 + \lambda_3 \|L_\mathbf{x}\|_1 + \lambda_4 \|\langle L_\mathbf{x}, \delta_t \mathbf{u}\rangle + L_t\|_1 + \lambda_5\, h_\theta(L - L(t_p)) \Big) \mathrm{d}t\, \mathrm{d}\mathbf{x} + \int_\Omega \sum_{i=2}^{|P(\mathbf{x})|} \|L(t_i) - L(t_{i-1}) - \theta\rho_i\|_1 \mathrm{d}\mathbf{x},$$

Smoothness terms

Optical flow term (**brightness constancy**)

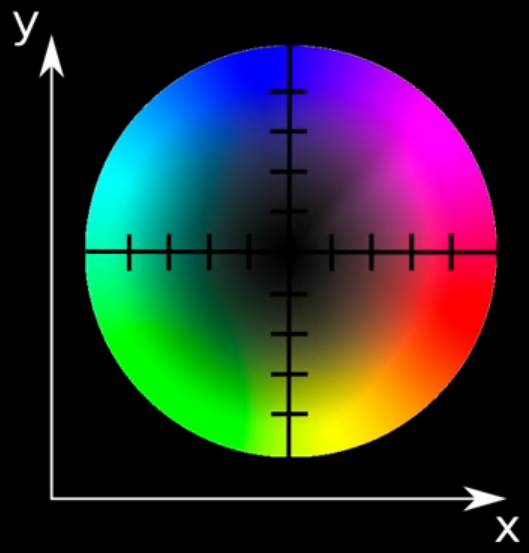
No-event term

Event term

Bardow et al., *Simultaneous Optical Flow and Intensity Estimation from an Event Camera*, CVPR 2016. YouTube

SOFIE: Simultaneous Optical Flow & IE

Conference on Computer Vision and Pattern Recognition (CVPR 2016)

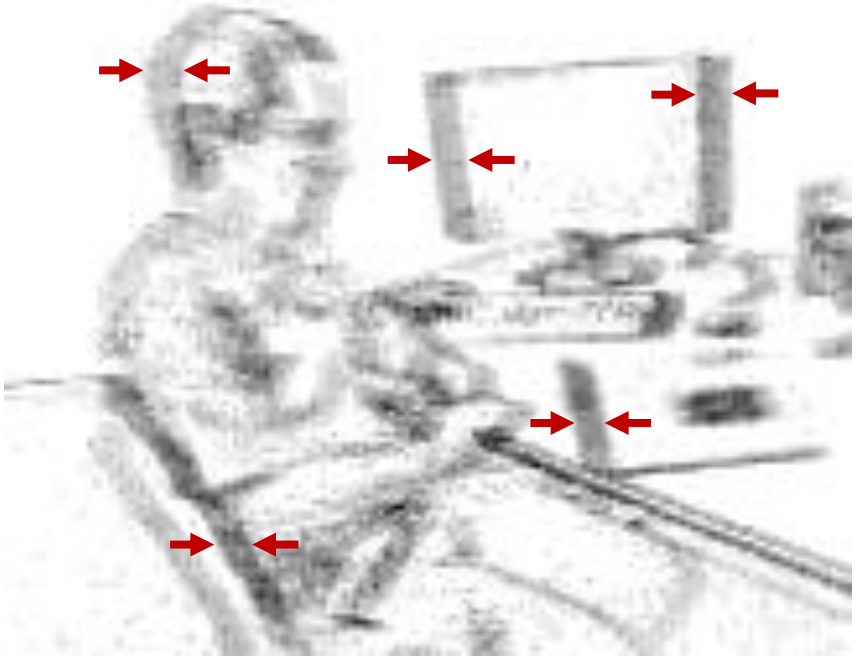


Optical Flow Visualization:
- Each Pixel's
- direction is shown as color and
- velocity as brightness
- (as shown in the color wheel)

Bardow et al., *Simultaneous Optical Flow and Intensity Estimation from an Event Camera*, CVPR 2016. YouTube

Optical Flow from Motion Compensation

- Assume **constant flow** in a small space-time volume (i.e., events in the volume share the same flow vector)
- **Idea**: find the flow $\boldsymbol{v}$ that best ***aligns*** corresponding events.
- Why? as an edge moves, it triggers events at pixels it crosses. Try to "**undo**" (compensate) this motion, to **recover the thin edge**.



Gallego et al., *A Unifying Contrast Maximization Framework for Event Cameras…*, CVPR 2018. PDF, YouTube

Optical Flow from Motion Compensation

- Assume **constant flow** in a small space-time volume (i.e., events in the volume share the same flow vector)
- **Idea**: find the flow $\boldsymbol{v}$ that best ***aligns*** corresponding events.
- Why? as an edge moves, it triggers events at pixels it crosses. Try to "**undo**" (compensate) this motion, to **recover the thin edge**.

Gallego et al., *A Unifying Contrast Maximization Framework for Event Cameras…*, CVPR 2018. PDF, YouTube

Optical Flow from Motion Compensation

Iteration steps:

1. **Transform** events: $e_k = (\boldsymbol{x}_k, t_k, p_k) \mapsto e'_k = (\boldsymbol{x}'_k, t_{\text{ref}}, p_k)$
   - Move events using candidate flow:
     $$\boldsymbol{x}'_k \doteq \boldsymbol{x}_k - (t_k - t_{\text{ref}})\boldsymbol{v}$$
   - t_{ref} : reference time (e.g., first event in the volume)
   - Easy: space = velocity * time

2. Measure how well the warped events **align**. How?
   - Euclidean distance between events e'_k (point sets) or
   - Contrast (focus) of histograms / image of events or
   - Average time per pixel if using time surface-like representation

Gallego et al., *A Unifying Contrast Maximization Framework for Event Cameras…*, CVPR 2018. PDF, YouTube
Stoffregen et al., *Simultaneous Optical Flow and Segmentation (SOFAS) using Dynamic Vision Sensor*, ACRA 2017. PDF, YouTube
Zhu et al., *Event-based Feature Tracking with Probabilistic Data Association*, ICRA 2017. PDF, YouTube

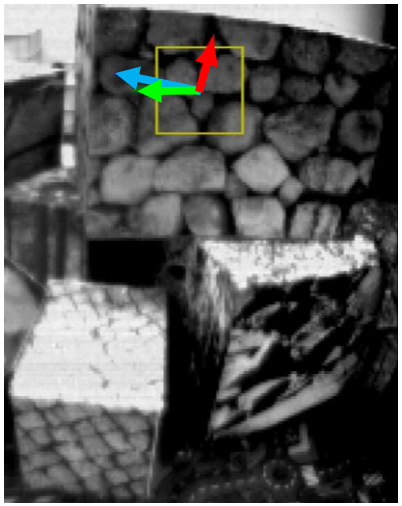
Optical Flow from Motion Compensation

Using **histograms of events**

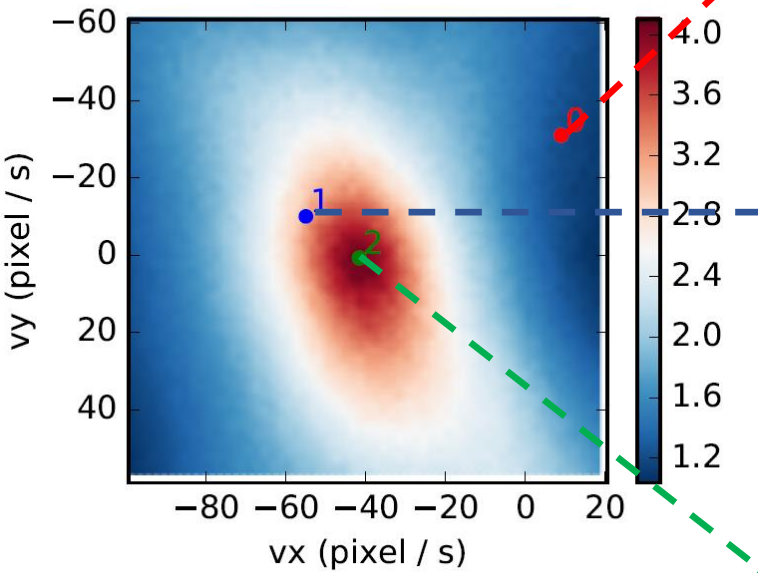
Warped Events

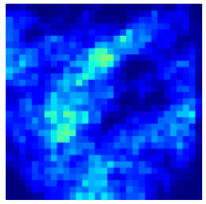
$\boldsymbol{x}'_k = \boldsymbol{x}_k - \boldsymbol{v}\, t_k$ straight trajectories

Events generated by a patch

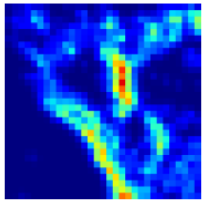


Focus vs. Optical Flow

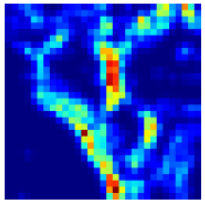




Blurred (without motion correction)



Sharper

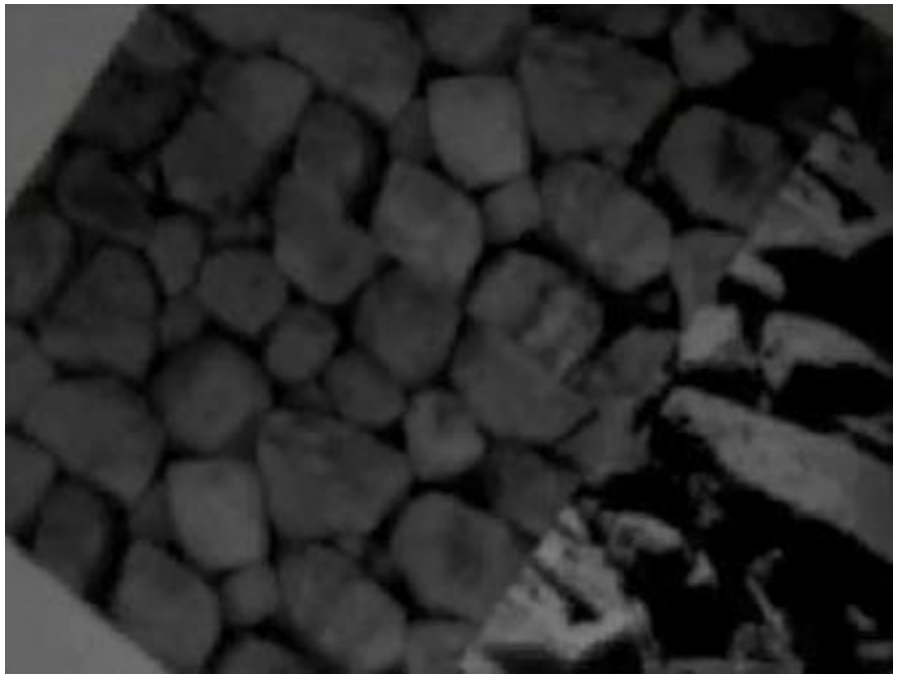


Sharpest (with motion correction)

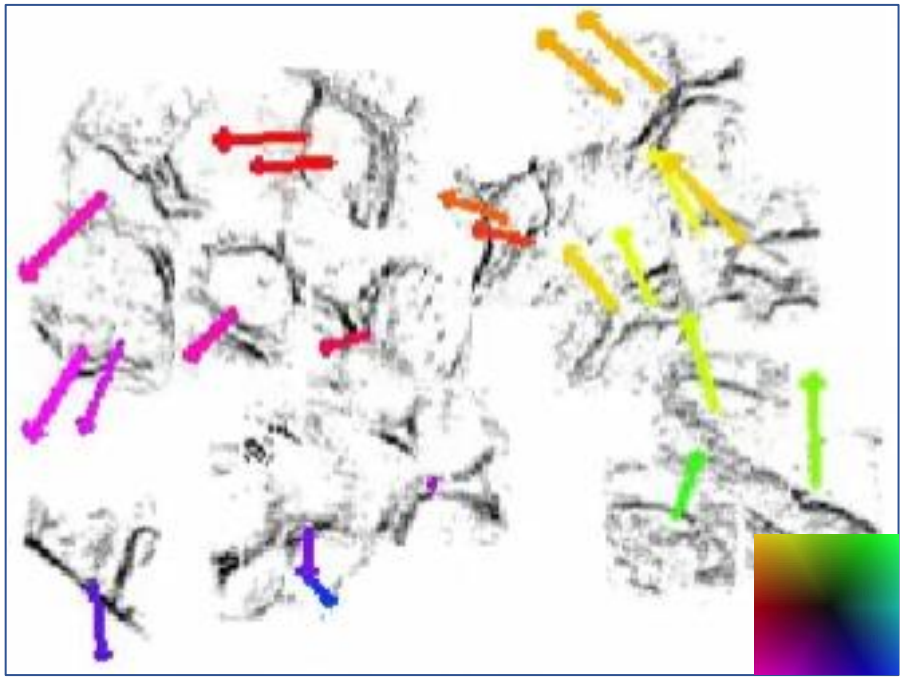
Gallego et al., *A Unifying Contrast Maximization Framework for Event Cameras…*, CVPR 2018. PDF, YouTube

Optical Flow from Motion Compensation

Frames (not used)



Events & Optical Flow



Full flow, at sparse locations (feature-based)

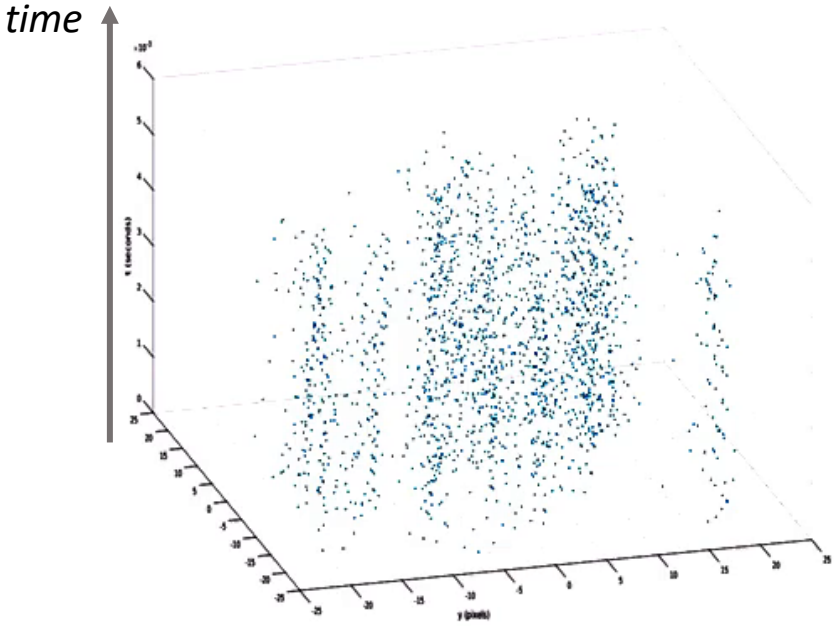
Gallego et al., *A Unifying Contrast Maximization Framework for Event Cameras…*, CVPR 2018. PDF, YouTube

Video by Dan Jia

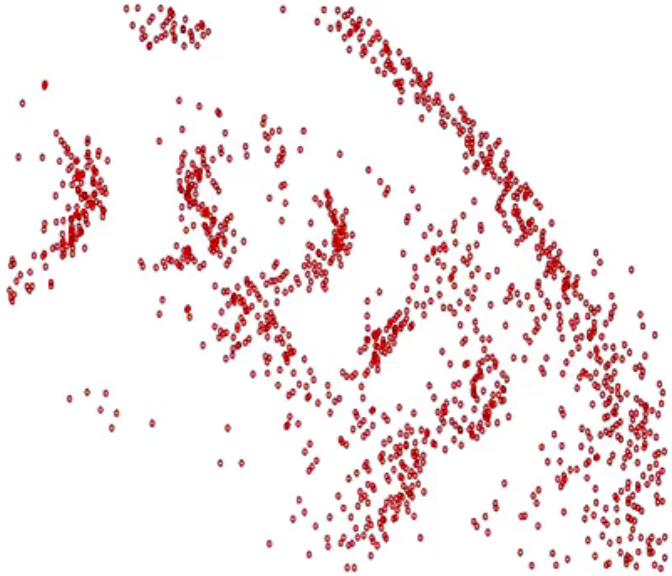
Optical Flow from Motion Compensation

Interpreting events as **point sets**

Space-time (polarity not used)



1. Warped events $\{\boldsymbol{x}'_k\}_{k=1}^{N_e}$ (2D point set)
(like a top-view of space-time)



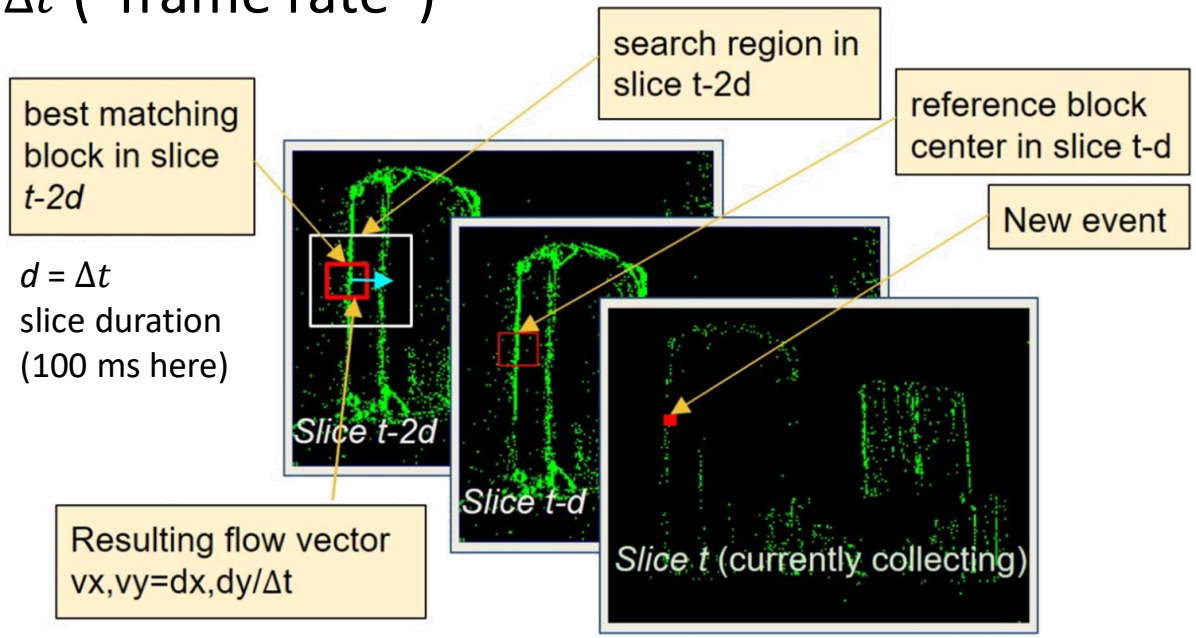
2. Measure the **alignment** of warped events using Euclidean distance between pairs

$$\min_v \sum_{i=1}^{n} \sum_{k=1}^{n} \left[\sum_{j=1}^{m} r_{ij} r_{kj} \right] \|(x_i - t_i v) - (x_k - t_k v)\|^2$$

Zhu et al., *Event-based Feature Tracking with Probabilistic Data Association*, ICRA 2017. PDF, YouTube

Optical Flow from Block Matching

- **Idea**: reuse video processing technique to estimate motion vectors ("flow"): matching patches / "blocks" (e.g., 21 x 21 pix)
- Representation: event frames, i.e., **method compares histograms of events** (uncompensated) and finds the best match.
- **Efficient search** for motion vector
- Adaptive slice duration Δt ("frame rate")
- FPGA implementation



Liu et al., *Adaptive Time-Slice Block-Matching Optical Flow Algorithm for Dynamic Vision Sensors*, BMVC 2018, YouTube

Optical Flow from Block Matching

ETH Zürich

University of Zurich[UZH]

Adaptive Time-Slice Block-Matching Optical Flow Algorithm for Dynamic Vision Sensors

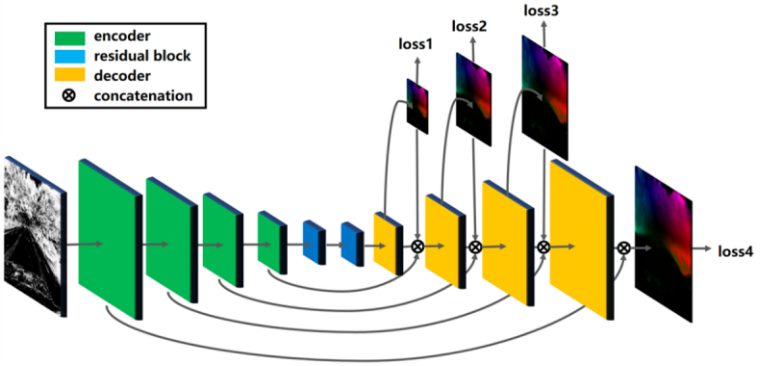
Min Liu and Tobi Delbruck
Institute of Neuroinformatics
University of Zurich & ETH Zurich

robotics Swiss National Centre of Competence in Research

sensors Institute of Neuroinformatics Zurich

Liu et al., *Adaptive Time-Slice Block-Matching Optical Flow Algorithm for Dynamic Vision Sensors*, BMVC 2018, YouTube

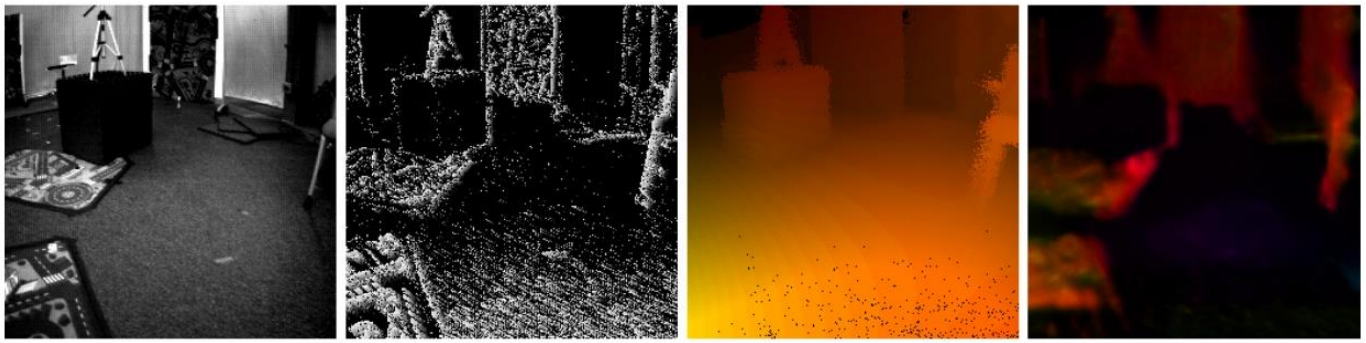
Ev-FlowNet



- **Idea**: learn flow from (lots of) data
- Architecture:
  - **CNN** (conventional computer vision). U-Net with skip connections…
  - Input (**4-channel**): event frames & time surfaces, split by polarity
- **Self-supervised** training using intensity frames
  - **Loss**: Photometric error (using flow to warp frames) + smoothness (regularizer)

$$\ell_{\text{photometric}}(u, v; I_t, I_{t+1}) = \sum_{x,y} \rho(I_t(x, y) - I_{t+1}(x + u(x, y), y + v(x, y)))$$

  - Frames are not needed during testing / inference
- Output: **dense flow** (on all pixels)

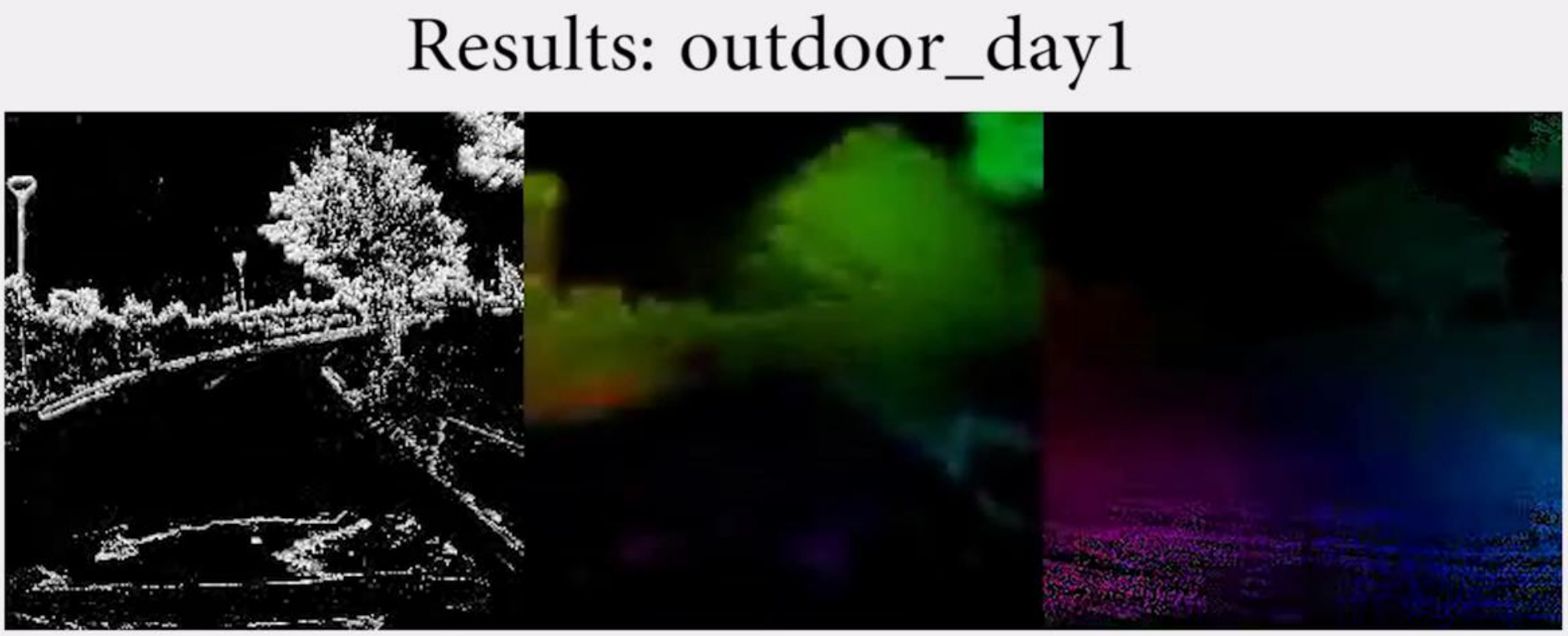


Images | Events | Ground truth | Optical Flow

Zhu et al., *EV-FlowNet: Self-Supervised Optical Flow Estimation for Event-based Cameras*, RSS 2018. PDF, YouTube

Ev-FlowNet

- Trained & tested on MVSEC dataset (on different sequences)
- **Ground truth** from motion field (depth provided by LIDAR)



Events
(4-channel input)

Estimated, dense flow

Ground truth flow

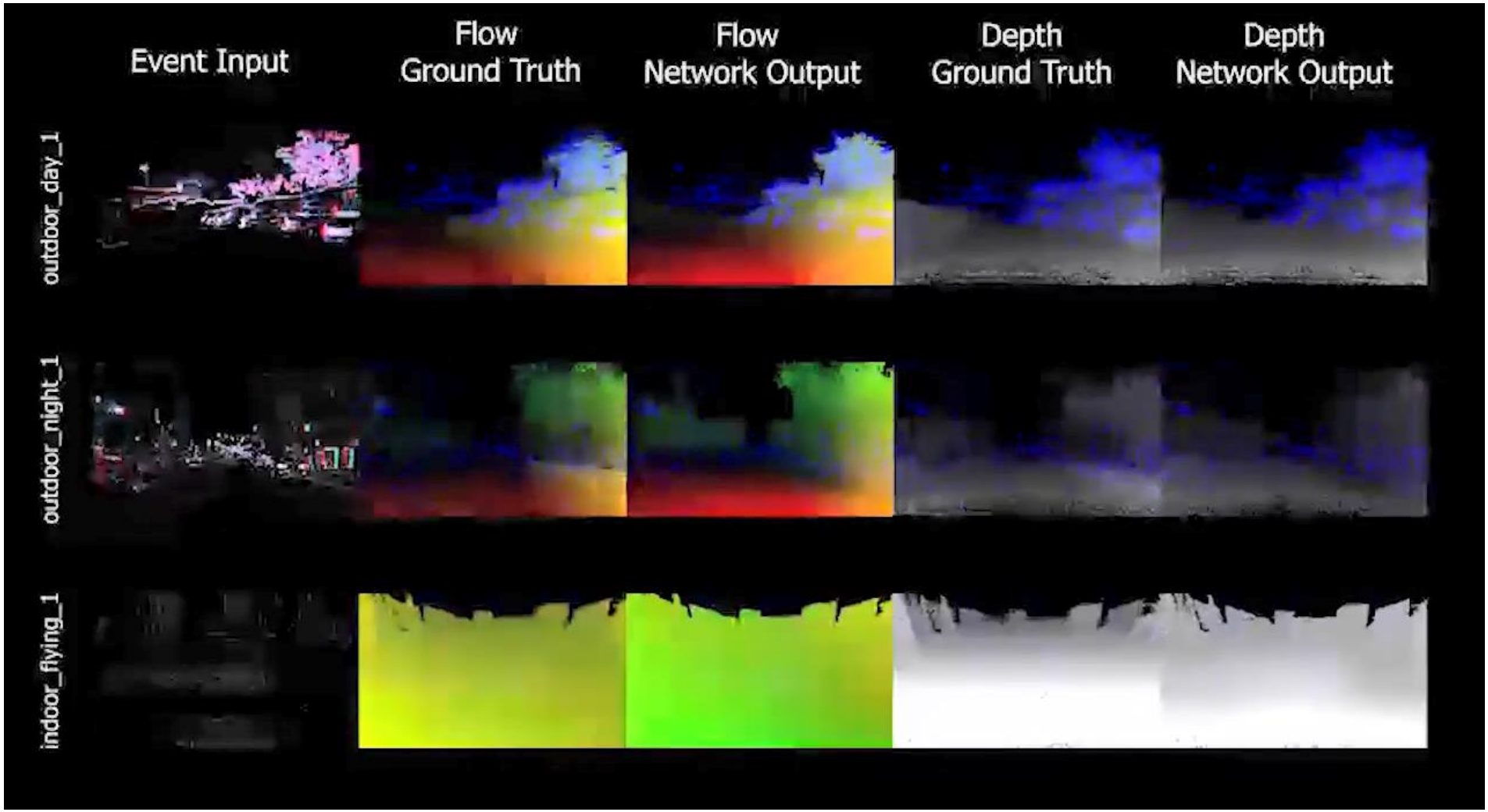
Zhu et al., *EV-FlowNet: Self-Supervised Optical Flow Estimation for Event-based Cameras*, RSS 2018. PDF, YouTube

Learning Structure from Motion

- Input (3-channel): events & average timestamps
- Architecture:
  - Evenly-Cascaded Network (**ECN**) design (small network, lightweight)
  - **Two ECNs**: to predict depth and ego-motion (pose)
- **Unsupervised** training: L1 loss between warped slices using flow
- Output: dense flow (**motion field**) and depth

Ye et al., *Unsupervised Learning of Dense Optical Flow, Depth and Egomotion from Sparse Event Data*, IROS 2020. YouTube

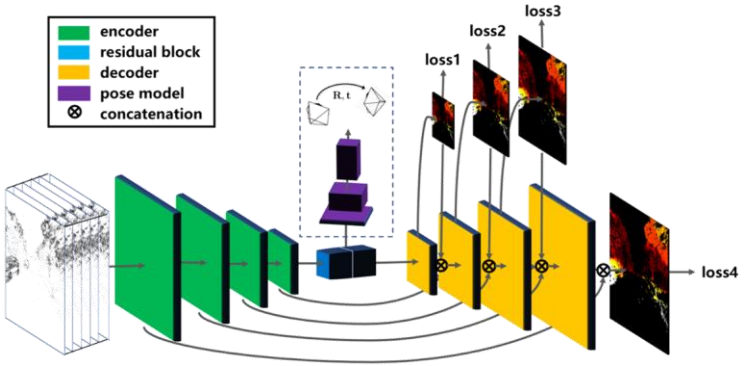
Learning Structure from Motion

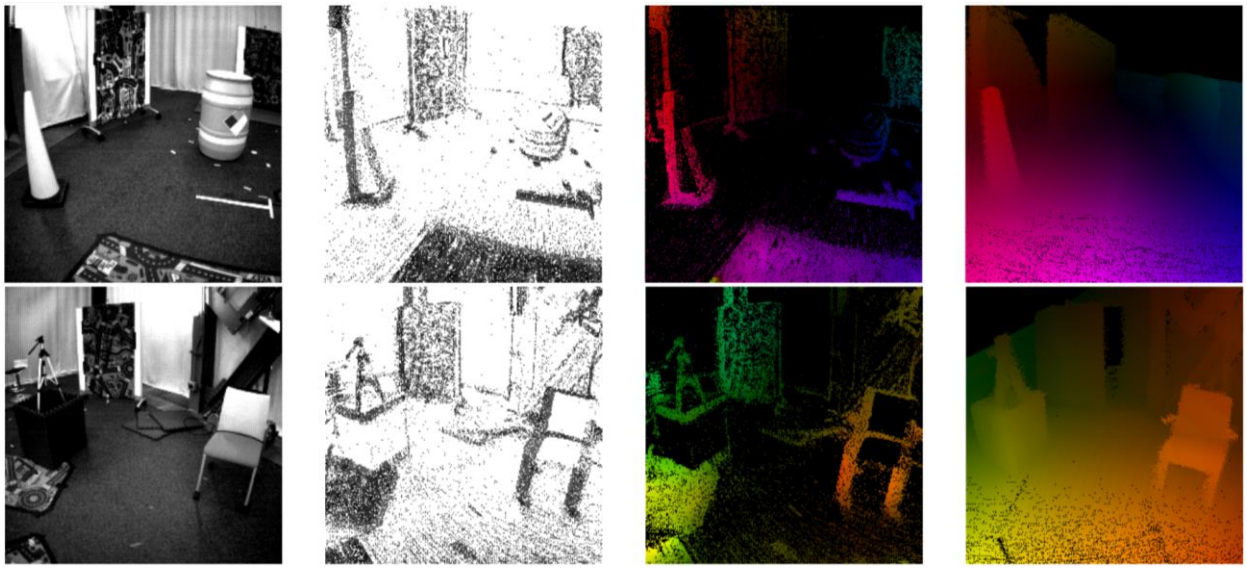


Ye et al., *Unsupervised Learning of Dense Optical Flow, Depth and Egomotion from Sparse Event Data*, IROS 2020. YouTube

Unsupervised CNN-based

- **Optical flow** NN or **Depth + ego-motion** (SFM) NN
- Network architecture: U-Net with skip connections and multi-level loss
- Input: **voxel grid**
- Loss: **motion-compensation**
- Results:





Images | Events | Optical Flow | Ground truth

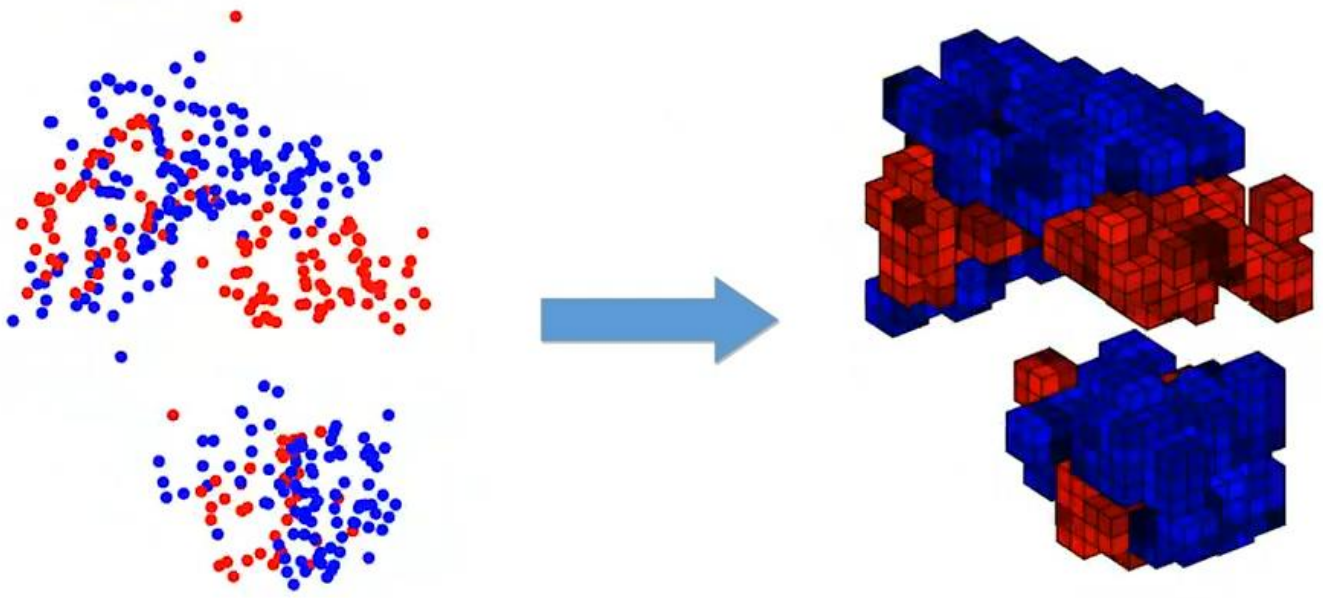
Zhu et al., *Unsupervised Event-Based Learning of Optical Flow, Depth, and Egomotion*, CVPR 2019. PDF, YouTube

Unsupervised CNN-based

Penn Engineering | GRASP Laboratory
General Robotics, Automation, Sensing & Perception Lab

Input Representation

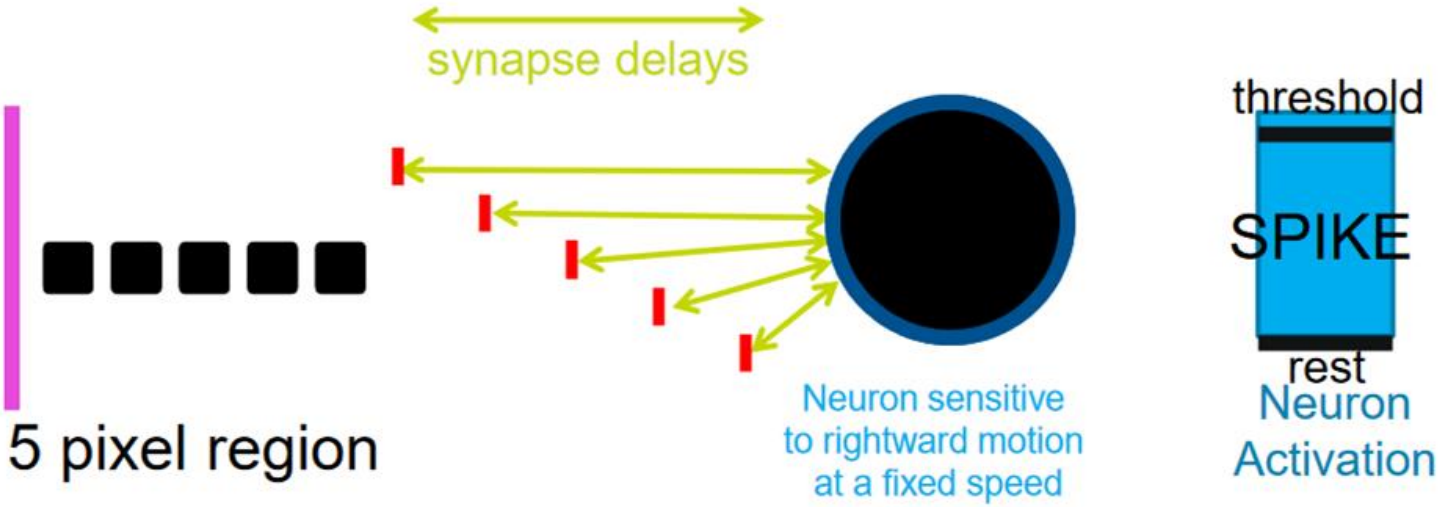
We propose the **discretized event volume.** Events are inserted into the volume with **trilinear interpolation**, resulting in minimal loss in resolution.



Zhu et al., *Unsupervised Event-Based Learning of Optical Flow, Depth, and Egomotion*, CVPR 2019. PDF, YouTube

Optical Flow by SNN-coincidence detection

- What is a Spiking Neural Network (**SNN**)?
- **Idea**: detect a particular flow (direction and speed) by **coincidence detection** of events at a neuron.



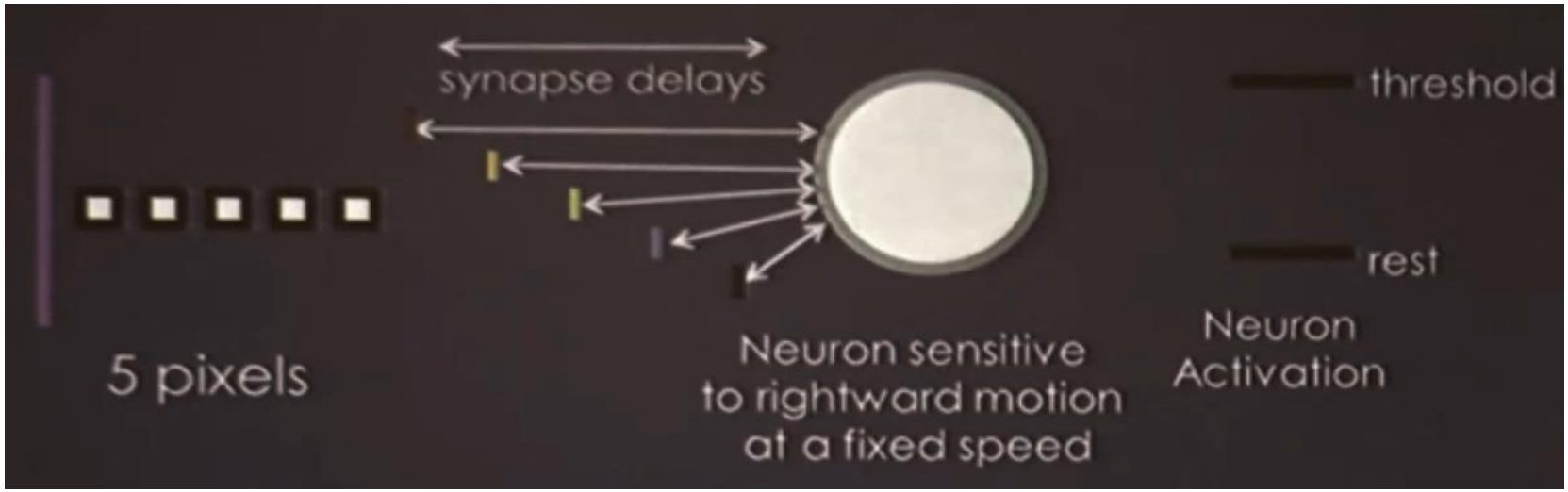
- **Delay** each event by a different amount so that events in the receptive field arrive at the same time to the neuron.

Orchard et al., *A Spiking Neural Network Architecture for Visual Motion Estimation*, BioCAS 2013. PDF

Image by G. Orchard

Optical Flow by SNN-coincidence detection

- What is a Spiking Neural Network (**SNN**)?
- **Idea**: detect a particular flow (direction and speed) by **coincidence detection** of events at a neuron.



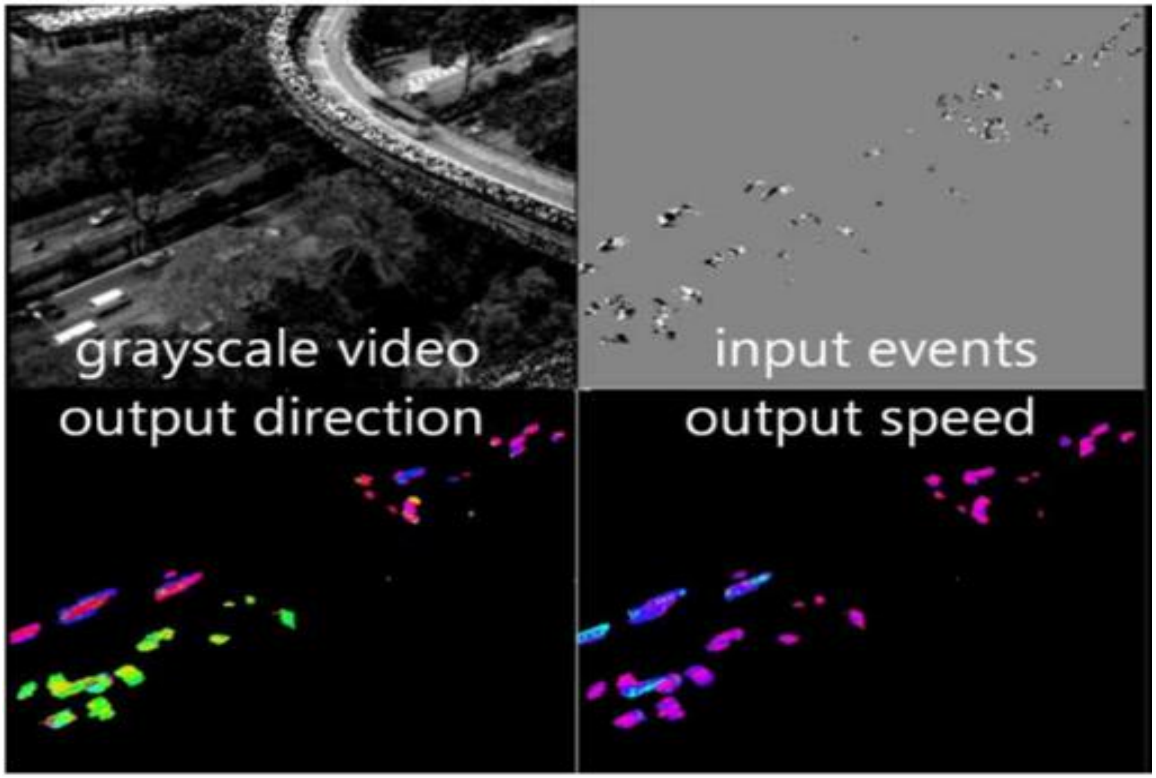
- **Delay** each event by a different amount so that events in the receptive field arrive at the same time to the neuron.

Orchard et al., *A Spiking Neural Network Architecture for Visual Motion Estimation*, BioCAS 2013. PDF https://youtu.be/INqNL1TGjlQ?t=1057

Optical Flow by SNN-coincidence detection

Experiments:

- ATIS camera
- Estimate flow in 8 directions and 8 speeds = 64 possible flow vectors (64 neurons)
- Manually set the SNN synapses



8 directions

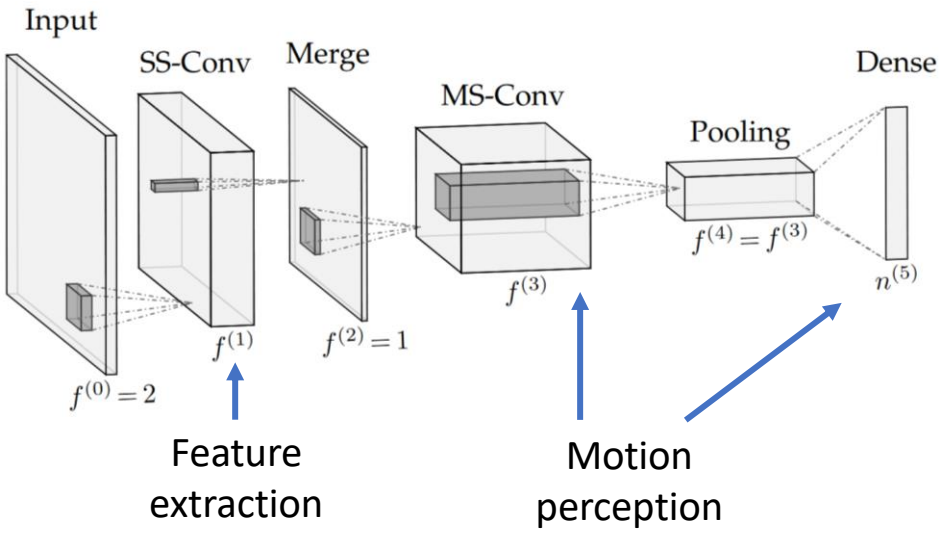
20 30 50 70 90 130 270 290 pix/s

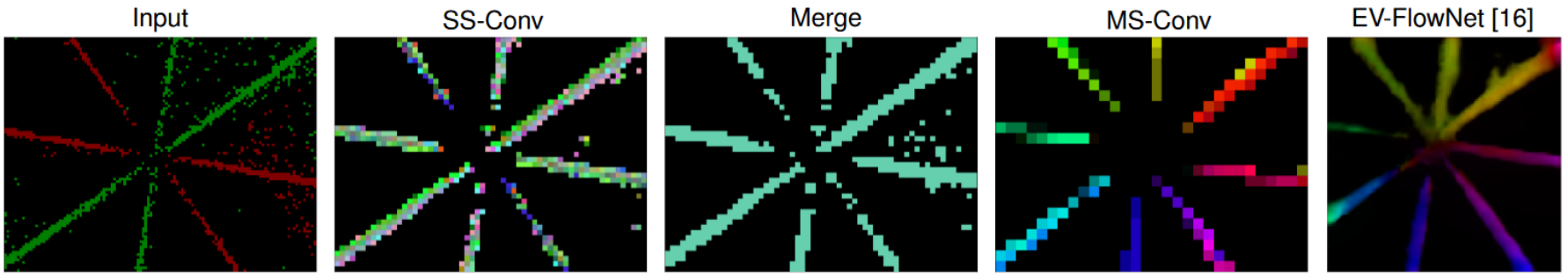
8 speeds

Orchard et al., *A Spiking Neural Network Architecture for Visual Motion Estimation*, BioCAS 2013. PDF https://youtu.be/INqNL1TGjlQ?t=1057

Optical Flow using a Hierarchical SNN

- Architecture: feedforward **SNN**
- **Connections are learned**
  - Synapses with delays
- Biological properties: **feature extraction**, and local and global **motion perception**
- All software, not implemented in hardware.

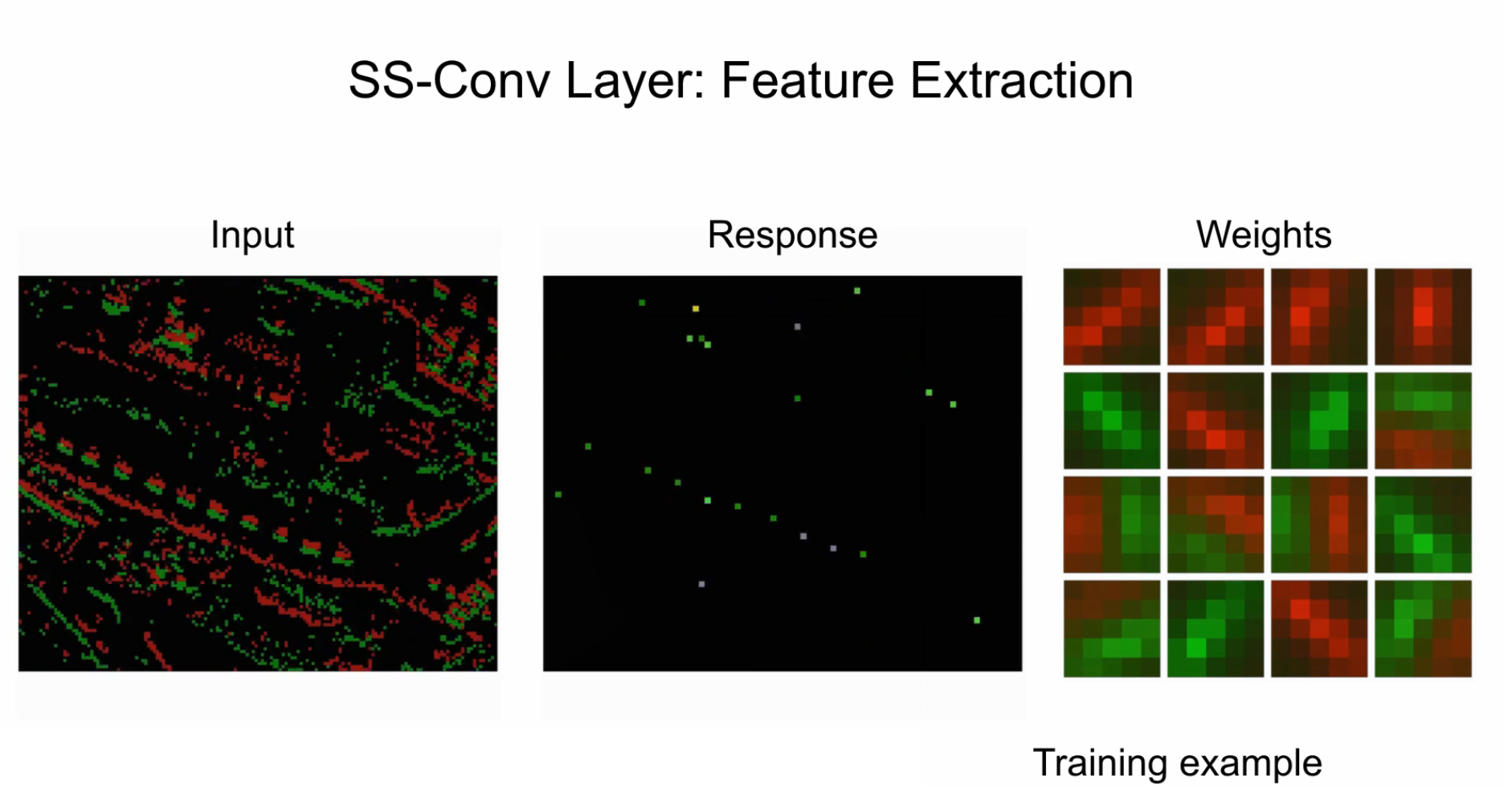




Paredes-Valles et al., *Unsupervised Learning of a Hierarchical Spiking Neural Network for Optical Flow Estimation.*, TPAMI 2019. PDF, YouTube

Optical Flow using a Hierarchical SNN

- Result of the SS-Conv and MS-Conv layers



Paredes-Valles et al., *Unsupervised Learning of a Hierarchical Spiking Neural Network for Optical Flow Estimation.*., TPAMI 2019. PDF, YouTube

How do the methods compare?

- Hard to tell… there is a lack of comparison benchmark
  - 2016 Comparison by Rueckauer et al. Front. Neuroscience
  - But… there have been many developments since 2016
- De facto standard: use MVSEC dataset (Zhu et al., 2018)
  - "Ground truth" flow from motion field obtained by SLAM-based sensor fusion (LIDAR + motion capture) + manual registration.
  - Take the numbers with a grain of salt.

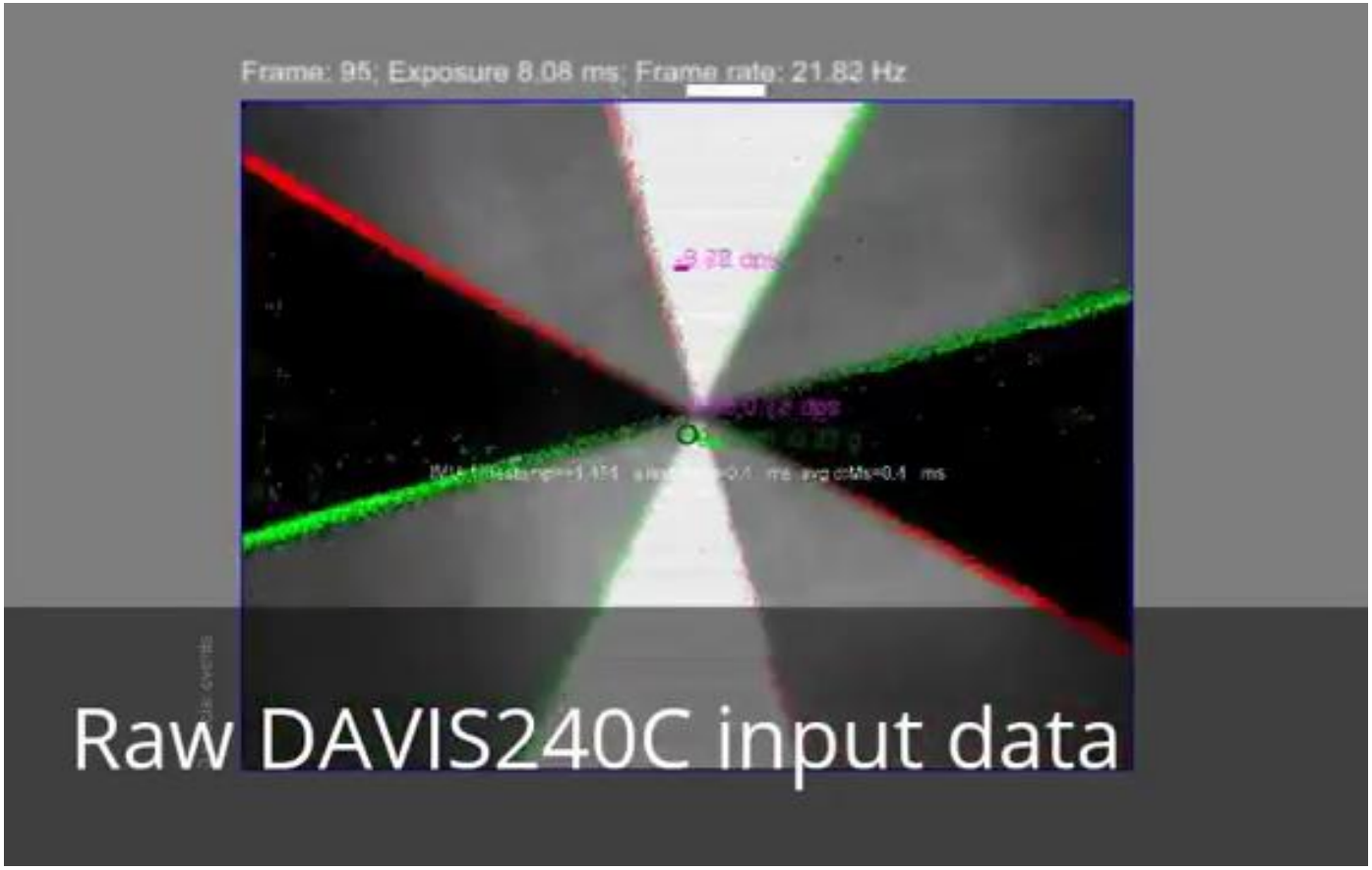
	outdoor day1		indoor flying1		indoor flying2		indoor flying3	
dt=1 frame	AEE	% Outlier	AEE	% Outlier	AEE	% Outlier	AEE	% Outlier
Ours	**0.32**	**0.0**	0.58	**0.0**	1.02	4.0	0.87	3.0
EV-FlowNet	0.49	0.2	1.03	2.2	1.72	15.1	1.53	11.9
UnFlow	0.97	1.6	0.50	0.1	0.70	1.0	0.55	**0.0**
ECN_{masked}	0.36	0.2	**0.20***	**0.0***	**0.24***	**0.0***	**0.21***	**0.0***

	outdoor day1		indoor flying1		indoor flying2		indoor flying3	
dt=4 frames	AEE	% Outlier	AEE	% Outlier	AEE	% Outlier	AEE	% Outlier
Ours	1.30	9.7	**2.18**	**24.2**	**3.85**	46.8	3.18	47.8
EV-FlowNet	**1.23**	**7.3**	2.25	24.7	4.05	**45.3**	3.45	39.7
UnFlow	2.95	40.0	3.81	56.1	6.22	79.5	**1.96**	**18.2**
ECN_{masked}	-	-	-	-	-	-	-	-

Zhu et al, CVPR 2019

2016 comparison of previous methods

- Nicely compare 4 + 4 + 1 algorithms vs ground truth (from IMU)
  - 4 variations of LK-method, 4 variations of plane-fitting, 1 direction selective
- But it lacks parallax



Rueckauer et al., *Evaluation of Event-Based Algorithms for Optical Flow with Ground-Truth from Inertial Measurement Sensor*, FNINS 2016. YouTube

References

- Section 4.2 of Event-based Vision: A Survey, TPAMI 2020
- Papers referenced at the bottom of each slide.
- Classical computer vision books for the topic of feature detection and tracking.
  - Optic Flow – Scholarpedia: http://www.scholarpedia.org/article/Optic_flow
  - D. J. Fleet & Y. Weiss, *Optical Flow Estimation*, Handbook of Mathematical Models in Computer Vision, 2005 http://www.cs.toronto.edu/~fleet/research/Papers/flowChapter05.pdf
  - "What Is Optical Flow for?": Workshop Results and Summary, ECCVW 2018.