Recent advances in neuromorphic vision sensors: A survey

Jianing Li
lijianing@pku.edu.cn
https://jianing-li.github.io/

Multimedia Learning Group
School of EE & CS, Peking University

Dec.18th, 2019
Overview

- **Introduction**
  - Conventional camera
  - Biological visual sensing system
  - Neuromorphic vision milestones
  - Neuromorphic vision Sensors

- **Event-based signal processing**
  - Spatial-temporal filter
  - Spike metric
  - Spike coding

- **Feature representations (Applications)**
  - Rate-based image
  - Hand-crafted feature
  - End-to-end deep network
  - Spiking neural network

- **Discussion**
  - Better input representations for event data
  - Event-based vision in the future
Overview

- Introduction
  - Conventional camera
  - Biological visual sensing system
  - Neuromorphic vision milestones
  - Neuromorphic vision Sensors

- Event-based signal processing
  - Spatial-temporal filter
  - Spike metric
  - Spike coding

- Feature representations (Applications)
  - Rate-based image
  - Hand-crafted feature
  - End-to-end deep network
  - Spiking neural network

- Discussion
  - Better input representations for event data
  - Event-based vision in the future
Conventional camera

**Disadvantage of conventional camera**

- Over-sampling, data redundancy
- Under-sampling, motion blur
- Low dynamic range
- High power

**High frame camera**

- **Too expensive**
  - 16000FPS
  - 118,000 $
- **Short storage**
  - 720p frame only
  - Storing 10s
- **Light requirement**
  - Exposure time < 60μs
  - High illuminance

**High-frame camera**

Phantom V612
Biological visual sensing system

- Retina sampling
  - Three layers structures
  - Fovea – visual texture
  - Peripheral – high time resolution
Neuromorphic vision milestones

Mcculloch Pitts

1943

Spike firing model

Hodgkin Huxley

1952

Carver Mead

1980s

silicon retina

Mahowald

1991

Octopus retina

Culurciello

1993

ATIS

Christoph Posh

2003

DAVIS

Shoushun Chen

2005

DVS

IVS

Mahowald & Boahen

2008

AER

Tobias Delbruck

2013

Color-DVS

Carver Mead

2015

Christoph Posh

2017

Mahowald & Boahen

2018

Tobias Delbruck

2018

Carver Mead
Neuromorphic sampling

- Differential sampling model
  - Address event representation (AER)
  - Illumination change

\[ \Delta L = \ln L(u, t) - \ln L(u, t - \Delta t) = p \theta \]

- DVS, DAVIS, ATIS, CeleX

Neuromorphic vision sensors

- Dynamic vision sensor (DVS) – Differential sampling

Neuromorphic vision sensors

- Integral sampling model
  - Leaky integrate and fire (LIF) model
    \[ A(t) = \int_0^t I(t)dt \geq \varphi \]
  - Octopus retina, FSM

Diagram:
- Photoreceptor cell
- Bipolar cell
- Ganglion cell
- FSM
- Texture image
- Integral firing
- Reset unit
- Readout unit
- SPIKES
- Voltage pulse
- Integration
- X, Y axes
- Spike train
- Texture pattern
Neuromorphic vision sensors

- Fovea-like sampling model (FSM) – Integral sampling
Neuromorphic vision sensors

- Asynchronous time-based image sensor (ATIS)
  - **DVS**: change detector
  - Time-based: greyscale events

![DVS Diagram]

\[ I_{ph} \approx \frac{1}{t_2 - t_1} \]

Neuromorphic vision sensors

- Dynamic and active pixel vision sensor (DAVIS)
  - DVS
  - APS: 50 Hz texture image
  - Two stream: asynchronous
Neuromorphic vision sensors

- CeleX - NTU
  - DVS
  - Voltage -> grayscale
  - Optical flow
# Neuromorphic vision sensors

## Comparison between different commercialized neuromorphic vision sensors

<table>
<thead>
<tr>
<th>Sensors</th>
<th>DAVIS128</th>
<th>ATIS</th>
<th>DAVIS346</th>
<th>DVS-G2</th>
<th>CeleX-V</th>
<th>FSM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supplier</td>
<td>iniVation</td>
<td>Prophesee</td>
<td>iniVation</td>
<td>Samsung</td>
<td>CelePixel</td>
<td>PKU</td>
</tr>
<tr>
<td>Year</td>
<td>2008</td>
<td>2011</td>
<td>2017</td>
<td>2017</td>
<td>2018</td>
<td>2018</td>
</tr>
<tr>
<td>Resolution</td>
<td>128×128</td>
<td>304 ×240</td>
<td>346×260</td>
<td>640×480</td>
<td>1280×800</td>
<td>400×250</td>
</tr>
<tr>
<td>Sampling (Hz)</td>
<td>1×10⁶</td>
<td>1×10⁶</td>
<td>1.2 ×10⁷</td>
<td>3×10⁹</td>
<td>1.6 ×10⁸</td>
<td>4×10⁴</td>
</tr>
<tr>
<td>DR (dB)</td>
<td>120</td>
<td>143</td>
<td>120</td>
<td>90</td>
<td>120</td>
<td>70</td>
</tr>
<tr>
<td>Power (mW)</td>
<td>23</td>
<td>50-175</td>
<td>10-170</td>
<td>27-50</td>
<td>390-470</td>
<td>370</td>
</tr>
<tr>
<td>Chip Size (mm²)</td>
<td>6.3×6</td>
<td>9.9×8.2</td>
<td>8×6</td>
<td>8×5.8</td>
<td>14.3×11.6</td>
<td>10×6</td>
</tr>
<tr>
<td>Pixel Size (μm²)</td>
<td>40×40</td>
<td>30×30</td>
<td>18.5×18.5</td>
<td>9×9</td>
<td>9.8×9.8</td>
<td>20×20</td>
</tr>
<tr>
<td>Fill factor</td>
<td>8.1%</td>
<td>20%</td>
<td>22%</td>
<td>100%</td>
<td>9%</td>
<td>13.75%</td>
</tr>
<tr>
<td>Latency (μs)</td>
<td>12</td>
<td>3</td>
<td>20</td>
<td>65-410</td>
<td>1</td>
<td>25</td>
</tr>
<tr>
<td>Voltage (V)</td>
<td>3.3</td>
<td>1.8&amp;3.3</td>
<td>1.8&amp;3.3</td>
<td>1.2&amp;3.3</td>
<td>1.2&amp;3.3</td>
<td>1.5&amp;3.3</td>
</tr>
<tr>
<td>Texture image</td>
<td>No</td>
<td>gray</td>
<td>color</td>
<td>No</td>
<td>gray</td>
<td>gray</td>
</tr>
</tbody>
</table>
Neuromorphic vision sensors

- **Machine vision**
  - High speed
  - Challenging illumination

- Drone
- iCub robot
- Autonomous driving
Neuromorphic vision sensors

- Research institutes on neuromorphic
Neuromorphic vision sensors

- Related works
  - Paper numbers
  - Papers in CVPR
  - Research fields

Papers

Papers in CVPR

Research fields

- stereo vision
- SLAM
- tracking/detection
- classification/recognition
- databases
- signal processing
- reconstruction
- Review
Overview

- **Introduction**
  - Conventional camera
  - Biological visual sensing system
  - Neuromorphic vision milestones
  - Neuromorphic vision Sensors

- **Event-based signal processing**
  - Spatial-temporal filter
  - Spike metric
  - Spike coding

- **Feature representations (Applications)**
  - Rate-based image
  - Hand-crafted feature
  - End-to-end deep network
  - Spiking neural network

- **Discussion**
  - Better input representations for event data
  - Event-based vision in the future
Event-based signal processing

- Spatial-temporal filter
  - Denoising
  - Intensity estimation


Event-based signal processing

- **Spike metric --- the basis of event-based signal processing**
  - **Spike coding**
    - Distortion measurement
    - Motion estimation
  - **Neuroscience**
    - Retinal prostheses
    - Multi-neuron synchrony
  - **Event-based vision**
    - Retrieval and tracking
    - Matching in 3D stereo
    - Loss function for ANN or SNN

1. No taking *polarity* into account
2. No *spatial structure* information

---

Event-based signal processing

- Spike coding
  - Prediction framework
    - Intra-cube
    - Inter-cube

Overview

- Introduction
  - Conventional camera
  - Biological visual sensing system
  - Neuromorphic vision milestones
  - Neuromorphic vision Sensors

- Event-based signal processing
  - Spatial-temporal filter
  - Spike metric
  - Spike coding

- Feature representations (Applications)
  - Rate-based image
  - Hand-crafted feature
  - End-to-end deep network
  - Spiking neural network

- Discussion
  - Better input representations for event data
  - Event-based vision in the future
Feature representation for event streams

- **Rate-based image**

<table>
<thead>
<tr>
<th>Framework</th>
<th>Rate-based image + Deep Network</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asynchronous Events $e(x,y,t,p)$</td>
<td>Integration time $T$</td>
</tr>
<tr>
<td></td>
<td>$h^+(x,y)$</td>
</tr>
<tr>
<td></td>
<td>Synchronous Event Frame</td>
</tr>
<tr>
<td></td>
<td>EV-Gait: Event-based robust gait recognition using dynamic vision sensors, CVPR 2019</td>
</tr>
<tr>
<td></td>
<td>EKLT: Asynchronous photometric feature tracking using events and frames, IJCV 2019</td>
</tr>
<tr>
<td></td>
<td>1. Event-based vision meets deep learning on steering prediction for self-driving cars, CVPR 2018</td>
</tr>
</tbody>
</table>

**Representation works**
1. Directly applied in image-based method;
2. Failed to exploit spatial-temporal attribute.

**Reviews**

---

Feature representation for event streams

- Event volume

<table>
<thead>
<tr>
<th>Framework</th>
<th>Event volume + Deep Network</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_i^* = (B - 1)(t_i - t_0)/(t_N - t_1)$</td>
<td></td>
</tr>
<tr>
<td>$V(x, y, t) = \sum_i p_i k_b(x - x_i) k_b(y - y_i) k_b(t - t_i^*)$</td>
<td></td>
</tr>
<tr>
<td>$k_b(a) = \max(0, 1 -</td>
<td>a</td>
</tr>
</tbody>
</table>

| Representation works | 1. Unsupervised event-based learning of optical flow, depth, and ego-motion, CVPR 2019  
2. Events-to-video: Bring modern computer vision to event cameras, CVPR 2019 |
|----------------------|-------------------------------------------------------------------|
| Reviews | 1. Transform image by sampling function;  
2. Failed to exploit spatial-temporal attribute. |
# Feature representation for event streams

- **Hand-crafted feature**

<table>
<thead>
<tr>
<th>Framework</th>
<th>Representation works</th>
<th>Reviews</th>
</tr>
</thead>
</table>
| **Layer 1** | 1. HOTS: A hierarchy of event-based time-surfaces for **pattern recognition**, *PAMI 2017*  
2. HATS: Histograms of averaged time surfaces for robust event-based **object classification**, *CVPR 2018* | 1. **Strong prior performance**;  
2. **Time-consuming by hand-crafted debug.** |

**Time surface**

\[
T_i(u, p) = \max_{j \leq i} \{ t_j | x_j = (x_i + u), p_j = p \}
\]

\[
S_i(u, p) = e^{-\frac{t_i - T_i(u, p)}{\tau}}
\]
Feature representation for event streams

- Learning feature

<table>
<thead>
<tr>
<th>Framework</th>
<th>Learning feature via deep network</th>
</tr>
</thead>
<tbody>
<tr>
<td>Representation works</td>
<td>1. <strong>End-to-end learning of representations</strong> for asynchronous event-based data, <em>ICCV 2019</em></td>
</tr>
<tr>
<td>Reviews</td>
<td>1. <strong>Obtaining better performance than rate-based image</strong>; 2. <strong>Multi-stage design</strong>,</td>
</tr>
</tbody>
</table>
Feature representation for event streams

- 3D convolution

<table>
<thead>
<tr>
<th>Framework</th>
<th>3D convolution (deep network)</th>
</tr>
</thead>
</table>

| Representation works | 1. Constant velocity 3D convolution, *3D Vision 2018*  
2. Constant velocity 3D convolution, *IEEE Access 2018* |

| Reviews | 1. **Obtaining better performance than rate-based image**;  
2. **High computing complexity**. |
Feature representation for event streams

- **Event clouds**

<table>
<thead>
<tr>
<th>Framework</th>
<th>Event clouds + PointNet</th>
</tr>
</thead>
</table>

| Representation works | 1. Space-time event clouds for **gesture recognition**: from RGB cameras to event cameras, *WACV 2019*  
2. **EventNet**: Asynchronous recursive event processing, *CVPR 2019*  
3. Modeling **point clouds** with self-attention and Gumbel subset sampling, *CVPR 2019* |

| Reviews | 1. **Research hotspot**  
2. The state-of-the-art method. |
Feature representation for event streams

- **Graph-based**

<table>
<thead>
<tr>
<th>Framework</th>
<th>Event as graph + GNN</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><img src="image" alt="Graph-based Neural Network Diagram" /></td>
</tr>
</tbody>
</table>

| Representation works | 1. Graph-based **object classification** for neuromorphic vision sensing, *ICCV 2019*  
2. Graph-based **spatial-temporal feature learning** for neuromorphic vision sensing, *arXiv 2019*  
3. **Graph based event processing**, *Imaging and Applied Optics, 2019* |
| Reviews | 1. **First work using graph-based neural network**;  
2. The state-of-the-art methods. |
# Feature representation for event streams

- **Spiking neural network**

## Framework

| Representation works | 1. SLAYER: Spike layer **error reassignment** in time, *NIPS 2019*  
|----------------------|---------------------------------------------------------------|
|                      | 2. **Direct training** for spiking neural networks: Faster, larger and better, *AAAI 2019*  
|                      | 3. A spiking neural network model of **3D perception** for event-based neuromorphic stereo vision, *Scientific Reports, 2017*  
|                      | 4. A spiking neural network model of **depth from defocus** for event-based neuromorphic vision, *Scientific Reports, 2019*  

| Reviews | 1. **Focus on learning theory (such as supervised learning) and computing ability (GPU)**  
|---------|------------------------------------------------------------------------------------------|
|         | 2. **Low performance, especially in complex tasks**.  

---

![Spiking Neural Networks](image)
# Feature representation for event streams

## Comparison between different recognition methods on publication datasets

<table>
<thead>
<tr>
<th>Representations</th>
<th>Methods</th>
<th>N-MNIST</th>
<th>MNIST-DVS</th>
<th>N-Caltech</th>
<th>CIFAR10-DVS</th>
<th>N-CARS</th>
<th>DVS-Gesture</th>
<th>ASL-DVS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Rate-based image</strong></td>
<td>VGG_19</td>
<td>0.972</td>
<td>0.983</td>
<td>0.549</td>
<td>0.334</td>
<td>0.728</td>
<td>-</td>
<td>0.806</td>
</tr>
<tr>
<td></td>
<td>ResNet_50</td>
<td>0.984</td>
<td>0.982</td>
<td>0.637</td>
<td>0.558</td>
<td>0.903</td>
<td>-</td>
<td>0.886</td>
</tr>
<tr>
<td></td>
<td>LSTM</td>
<td>0.837</td>
<td>0.824</td>
<td>0.196</td>
<td>0.245</td>
<td>0.789</td>
<td>0.769</td>
<td>0.621</td>
</tr>
<tr>
<td><strong>Hand-crafted feature</strong></td>
<td>HOTS</td>
<td>0.808</td>
<td>0.803</td>
<td>0.210</td>
<td>0.271</td>
<td>0.624</td>
<td>0.785</td>
<td>0.656</td>
</tr>
<tr>
<td></td>
<td>HATS</td>
<td>0.991</td>
<td>0.984</td>
<td>0.642</td>
<td>0.524</td>
<td>0.902</td>
<td>0.933</td>
<td>0.871</td>
</tr>
<tr>
<td><strong>End-to-end deep network</strong></td>
<td>PATs</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.974</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>RG-CNNs</td>
<td>0.990</td>
<td><strong>0.986</strong></td>
<td><strong>0.657</strong></td>
<td><strong>0.540</strong></td>
<td><strong>0.914</strong></td>
<td>0.938</td>
<td><strong>0.901</strong></td>
</tr>
<tr>
<td><strong>Spiking neural network</strong></td>
<td>H-Frist</td>
<td>0.712</td>
<td>0.595</td>
<td>0.054</td>
<td>0.077</td>
<td>0.561</td>
<td>0.529</td>
<td>0.479</td>
</tr>
<tr>
<td></td>
<td>Direct-SNN</td>
<td><strong>0.995</strong></td>
<td>-</td>
<td>-</td>
<td><strong>0.605</strong></td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>SLAYER</td>
<td>0.992</td>
<td>0.956</td>
<td>0.598</td>
<td>0.532</td>
<td>0.907</td>
<td>0.936</td>
<td>0.869</td>
</tr>
</tbody>
</table>
Overview

- Introduction
  - Conventional camera
  - Biological visual sensing system
  - Neuromorphic vision milestones
  - Neuromorphic vision Sensors

- Event-based signal processing
  - Spatial-temporal filter
  - Spike metric
  - Spike coding

- Feature representations (Applications)
  - Rate-based image
  - Hand-crafted feature
  - End-to-end deep network
  - Spiking neural network

- Discussion
  - Better input representations for event data
  - Event-based vision in the future
## Discussion

- **How to find better representation for event data?**
  - **Learning feature**
  - **Event clouds**
  - **Graph-based**

<table>
<thead>
<tr>
<th>Representations</th>
<th>Works</th>
<th>Vision tasks</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Learning feature</strong></td>
<td><em>End-to-end learning of representations</em> for asynchronous event-based data, <em>ICCV 2019</em></td>
<td>Recognition &amp; optical flow</td>
</tr>
<tr>
<td></td>
<td>DART: <em>Distortion aware retinal transform</em> event-based cameras, <em>TPAMI 2019</em></td>
<td>Recognition &amp; tracking</td>
</tr>
<tr>
<td><strong>Event clouds</strong></td>
<td>Space-time event clouds for <em>gesture recognition</em>: from RGB cameras to event cameras, <em>WACV 2019</em></td>
<td>recognition</td>
</tr>
<tr>
<td></td>
<td><em>EventNet</em>: Asynchronous recursive event processing, <em>CVPR 2019</em></td>
<td>segmentation</td>
</tr>
<tr>
<td></td>
<td>Modeling <em>point clouds</em> with self-attention and Gumbel subset sampling, <em>CVPR 2019</em></td>
<td>recognition</td>
</tr>
<tr>
<td><strong>Graph-based</strong></td>
<td>Graph-based <em>object classification</em> for neuromorphic vision sensing, <em>ICCV 2019</em></td>
<td>recognition</td>
</tr>
<tr>
<td></td>
<td><em>Graph based event processing</em>, <em>Imaging and Applied Optics, 2019</em></td>
<td>Signal processing</td>
</tr>
</tbody>
</table>
Discussion

How many challenges are there in neuromorphic vision?

- Large datasets for supervised learning
- Asynchronous spatial-temporal spike metric
- Better representations for spatial-temporal streams
- Spiking neural networks for complex vision
- High computing using neuromorphic chip

Loihi, Intel
TureNorth, IBM
SpiNNaker, UM
Discussion

What is the future of neuromorphic vision sensors?
- Memory mechanism using neuromorphic vision sampling
- Multi-spectral sampling using neuromorphic concept
- Multi-sensors fusion, such as DVS, DAS, neuromorphic torch
- Neuroscience & Neuromorphic Engineering

Asynchronous spatial-temporal point processes

“Listen to the technology, find out what it’s telling you”

Carver Mead
THANKS

Q&A?
1. Can we lay a practical mathematical foundation that allows deriving efficient event-driven signal processing algorithms, analogous to the Z-transform of DSP?

2. Can we find better noise reduction (NR) algorithms than existing ones?

3. Can we find general methods for adaptively controlling sensor parameters like threshold, bandwidth, and refractory period?

4. Can we find better input representations for event cameras data for CNN?

5. What can we do to combine DVS events with color vision?
Organizers:

Davide Scaramuzza
UZH

Guillermo Gallego
UZH

Kostas Daniilidis
UPenn
Call for papers and demos

- Event-based / neuromorphic vision.
- Algorithm: Visual odometry, SLAM, 3D reconstruction, Optical flow estimation, image intensity reconstruction, recognition, stereo depth reconstruction, feature/object detection and tracking, calibration, sensor fusion.
- Model based, embedded or learning approaches.
- Event-based signal processing, control, bandwidth control.
- Event-based active vision.
- Event-based camera datasets and/or simulators.
- Applications in: robotics(navigation, manipulation, drones...), automotive, IoT, AR/VR, space, inspection, surveillance, crowd counting, physics.
- Biologically-inspired vision and smart cameras
- Novel hardware (cameras, neuromorphic processors, etc.) and/or software platforms.
- New trends and challenges in event-based and/or biologically-inspired vision.
CVPR 2019 Workshops

Invited speakers

- Tobi Delbruck
  ETH
- Garrick Orchard
  NUS
- Jorg Conradt
  KTH
- Giacomo Indiveri
  ETH
- Piotr Dudek
  Univ. Manchester
- Andrew Davision
  ICL
- Cornelia Fermuller
  Univ. Maryland
- Yulia Sandamirskaya
  ETH
- Chiara Bartolozzi
  Italiano di Tecnologia
- Margarita Chli
  ETH
- Robert Mahony
  ANU

Invited companies

- ATIS, France
- DVS(640*480), SK
- Loihi, USA
- Insightness, DVS, Switzerland
- DVS, Switzerland
- DVS, China
Appendix

- Paper list (1+8) in CVPR 2019
  - Bring a blurry frame alive at high frame-rate with an event camera, Liyuan Pan et. al, ANU. (oral)
  - Unsupervised event-based learning of optical flow, depth and ego-motion, Alex Z. Zhu et al, University of Penn.
  - Events-to-video: bringing modern computer vision to event cameras, Henri Rebecq et al, UZH & ETH.
  - EventNet: Asynchronous recursive event processing, Yusuque Sekikawa et al, Denso IT Laboratory.
  - Event-based high dynamic range image and very high frame rate video generation using conditional generation adversarial networks, S. M. Mostafavi et al, GIST.
  - Speed invariant time surface for learning to detect corner points with event-based cameras, J. Manderscheid et al, PROPHESSEE.
  - Focus loss functions for event-based vision, Guilleromo Gallego et at, UZH & ETH.
  - Event cameras, contrast maximization and reward functions: an analysis, T. N. Stoffregen et al, Monash University.
Appendix

- Paper list (1+3) in *ICCV 2019*
  - Learning an event sequence embedding for dense event-based deep stereo, Stepan Tulyakov et. al, *EPFL*. (oral)
  - Events-based motion segmentation by motion compensation, Timo stoffrengen et al, *UZH & ETH*.
  - Graph-based object classification for neuromorphic vision sensing, Yin Bin et al, *University of College London*.
Appendix

Paper list (1+3) in CCF A journal 2019
