Silicon retina technology

Tobi Delbruck

Inst. of Neuroinformatics, University of Zurich and ETH Zurich

Sponsors: Swiss National Science Foundation NCCR Robotics project, EU projects SEE BETTER and VISUALISE, Samsung, DARPA

Sensors Group sensors.ini.uzh.ch
Sponsors: Swiss National Science Foundation, NCCR Robotics, EU projects CAVIAR, SEEBETTER, VISUALISE, Samsung, DARPA, University of Zurich and ETH Zurich
Conventional cameras (**Static vision sensors**) output a stroboscopic sequence of frames.

**Good**

Compatible with 50+ years of machine vision

Allows small pixels (1μm for consumer, 3-5μm for machine vision)

**Bad**

Redundant output

Temporal aliasing

Limited dynamic range (60dB)

Fundamental “latency vs. power” trade-off
The Human Eye as a digital camera

- 100M photoreceptors
- 1M output fibers carrying max 100Hz spike rates
- 180dB ($10^9$) operating range
- >20 different “eyes”
- Many GOPs computing
- 3mW power consumption

Output is sparse, asynchronous stream of digital spike events
This talk has 4 parts

- Dynamic Vision Sensor Silicon Retinas
- Simple object tracking by algorithmic processing of events
- Using probabilistic methods for state estimation
- “Data-driven” deep inference with CNNs
DVS (Dynamic Vision Sensor) Pixel

Photoreceptor

$log/\Delta I$

Change amplifier (bipolar cells)

Event reset

Comparators (ganglion cells)

Threshold

Brightness change events

$Lichtsteiner$ et al., ISSCC 2007, JSSC 2009
DVS pixel has wide dynamic range

780 lux ÷ 5.8 lux

Edmund 0.1 density chart
Illumination ratio=135:1

ISSCC 2007
Using DVS for high speed (low data rate) imaging

Data rate <1MBps

“Frame rate” equivalent to 10 kHz but 100x less data

(10 kHz image sensor x 16k pixels = 160 MBps)
DAVIS (Dynamic and Active Pixel Vision Sensor) Pixel

- Photoreceptor
- Intensity reset
- Intensity value

\[ \pm \Delta \log I \]

- Change amplifier (bipolar cells)
- Comparator (ganglion cells)
- Event reset
- ON threshold
- OFF

From Rodieck 1998

Brandli et al., Symp VLSI, JSSC 2014
DVS/DAVIS +IMU demo

Start DAVIS Demo

Brandli, Berner, Delbruck et al., Symp. VLSI 2013, JSSC 2014, ISCAS 2015
DAVIS (Dynamic and Active Pixel Vision Sensor) Pixel

From Rodieck 1998
DAVIS346

AER DVS asynch. event readout

180nm CIS
346x260 18.5um pixel DAVIS

APS col-parallel ADCs and scanner

Bias generator

8mm
Important layout considerations
1. Post layout simulations to minimize parasitic coupling
2. Shielding parasitic photodiodes
Event threshold matching measurement

Experiment: Apply slow triangle wave LED stimulus to entire array, measure number of events that pixels generate

Conclusion: Pixels generate $11 \pm 3$ events per factor 3.3 contrast. Since $\ln(3.3) = 1.19$ and $1.19/11 = 0.11$, contrast threshold = 11% ± 4%
Measuring DVS pixel latency

Experiment: Stimulate small area of sensor with flashing LED spot, measure response latencies from recorded event stream

Conclusion: Pixels can have minimum latency of about 12us under bright illumination. But “real world” latencies are more like 100us-1ms.
DVS pixel has built-in temperature compensation

Since photoreceptor gain and threshold voltage both scale with absolute temperature $T$, it cancels out.

Photoreceptor

$V_p \propto T \ln(I_p)$

Threshold

$\Theta_{on} \propto T \ln(I_{on}/I_d)$
Integrated bias generator and circuit design enables operation over extended temperature range.

DAVIS240C

0.7°C  25°C  81°C

Nozaki, Delbruck 2017 (submitted)
Event camera silicon retina developments

Commercial entities
Inilabs (Zurich) – R&D prototypes
Insightness (Zurich) – Drones and Augmented Reality
Samsung (S Korea) – Consumer electronics
Pixium Vision (Paris) – Retinal implants
Inivation (Zurich) – Industrial applications, Automotive
Chronocam (Paris) - Automotive
Hillhouse (Singapore) - Automotive
Neuromorphic sensor R&D prototypes

Open source software, user guides, app notes, sample data

Shipped devices based on multiproject wafer silicon to 100+ organizations
• Dynamic Vision Sensor
  Silicon Retinas
• Simple object tracking
  by algorithmic
  processing of events
• Using probabilistic
  methods for state
  estimation
• “Data-driven” deep
  inference with CNNs
Tracking objects from DVS events using spatio-temporal coherence

1. For each event, find nearest cluster
   - If event within a cluster, move cluster
   - If event not within cluster, seed new cluster

2. Periodically prune starved clusters, merge clusters, etc (lifetime mgmt)

Advantages
1. Low computational cost (e.g. <5% CPU)
2. No frame memory (~100 bytes/object).
3. No frame correspondence problem

Litzenberger 2007
Using DVS allows 2 ms reaction time at 4% processor load with USB bus connections
This talk has 4 parts

• Dynamic Vision Sensor Silicon Retinas

• Simple object tracking by algorithmic processing of events

• Using probabilistic methods for state estimation

• “Data-driven” deep inference with CNNs
Simultaneous Mosaicing and Tracking with DVS

Simultaneous Mosaicing and Tracking with DVS

Event Camera & Scene

Visualisation of Events

Hanme Kim, A. Handa, … Andy J. Davison, BMVC 2014.
Goal: To do event-based, semi-dense visual odometry

• We want to estimate State vector $s$ (camera pose, visual scene spatial brightness gradients and sensor event thresholds) using Bayesian filtering from the events $e$: $p(s|e)$
• Sensor likelihood $p(e|s)$ is modeled as mixture of inlier Gaussian distribution and outlier uniform distribution
• A tractable posterior $q(s|e) \approx p(s|e)$ is approximated by Kullback-Leibler (KL) divergence
• Leads to closed-form update equations in the form of a classical Kalman filter, thus computationally efficient (unlike particle filtering)
Towards event-based, semi-dense SLAM: 6-DOF pose estimation

G. Gallego et al., PAMI (submitted 2016).
This talk has 4 parts

• Dynamic Vision Sensor Silicon Retinas
• Simple object tracking by algorithmic processing of events
• Using probabilistic methods for state estimation
• “Data-driven” deep inference with CNNs
RoShamBo CNN architecture

Conventional 5-layer LeNet with ReLU/MaxPool and 1 FC layer before output.

Total 18MOp (~9M MAC)

Compute times:
On 150W Core i7 PC in Caffe: 2ms
On 1W CNN accelerator on FPGA: 8ms

Conclusions

1. The DVS was developed by following a neuromorphic approach of emulating key properties of biological retinas
2. The wide dynamic range and sparse, quick output make these sensors useful in real time uncontrolled conditions
3. Applications could include vision prosthetics, surveillance, robotics and consumer electronics
4. The precise timing could improve learning and inference
5. The main challenges are to reduce pixel size and to develop effective algorithms. Only industry can do the first but academia has plenty of room to play for the second.
6. Event sensors can nicely drive deep inference. There is a lot of room for improvement of deep inference power efficiency at the system level!