Spiking silicon retina for digital vision

Tobi Delbruck
Inst. of Neuroinformatics
UZH-ETH Zurich
Switzerland

Patrick Lichtsteiner PhD project

Funding: UZH-ETH Zurich, EU Project CAVIAR, ARCs research
Silicon design: K. Boahen (Stanford) G. Indiveri & S. Mitra (UZH-ETH) C. Posch, (ARCs)

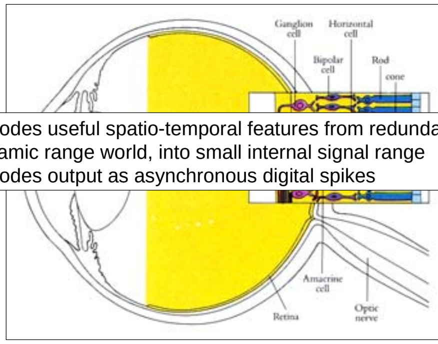
Inst. of Neuroinformatics, UNI-ETH Zurich

Telluride Neuromorphic Engineering Workshop

The problem with frames

Frame-based image sensors
- Have dominated machine vision for 40+ years
  - + Are compatible with displays
  - + Everyone understands them
  - + Allow lots of small (cheap) pixels
  - – impose a uniform limited sample rate
  - – make very redundant output
  - – generally have poor dynamic range (<60dB)

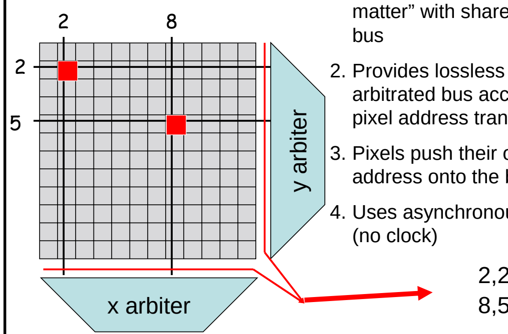
Function of the retina



1. Encodes useful spatio-temporal features from redundant, wide dynamic range world, into small internal signal range
2. Encodes output as asynchronous digital spikes

Rodieck, 1998

Address-Event Representation (AER) Transmitter



1. Replaces biology's “white matter” with shared digital bus
2. Provides lossless arbitrated bus access for pixel address transmission
3. Pixels push their own address onto the bus
4. Uses asynchronous logic (no clock)

Spike-based vision sensors

1990 1995 2000 2005

Mahowald & Mead outer retina (SciAm 91)

UPenn Magno-Parvo silicon retina (TBME 02)

CSEM VISe steerable-filter contrast vision sensor (JSSC 03)

Yale Univ. “Octopus” imager (JSSC 04)

JHU Temporal Change Threshold Detection Imager (JSSC 07)

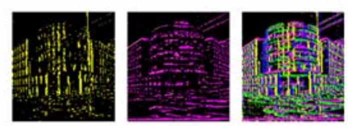
ETH Dynamic Vision Sensor (JSSC 08)

www.devise.ch VISe steerable-filter contrast vision sensor

CONTRAST

Ordered Evolution of Image Feature Information in Time

ORIENTATION



Rüedi et al. JSSC 2003

www.devise.ch VISe vision sensor output and use

Spike-based vision sensors

1990 1995 2000 2005

Mahowald & Mead outer retina

UPenn Magno-Parvo silicon retina

Devise steerable-filter contrast vision sensor

Yale Univ. “Octopus” imager

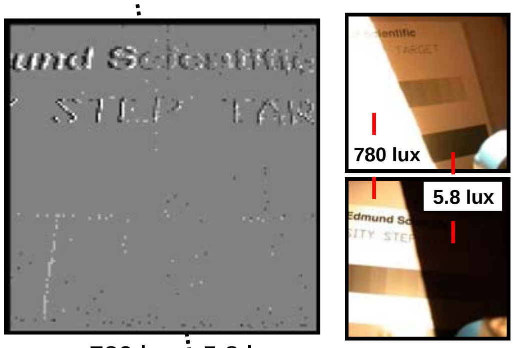
JHU Temporal Change Threshold Detection Imager

ETH Dynamic Vision Sensor (DVS)

2. Basic characteristics of dynamic vision sensor

- This asynchronous vision sensor responds to *temporal contrast.*
- It emits digital *address-events* that encode the *identities* of changing pixels.
- Each event means that the log intensity has changed by a quantized amount
- These events signify scene reflectance change

Uniform event threshold and wide dynamic range



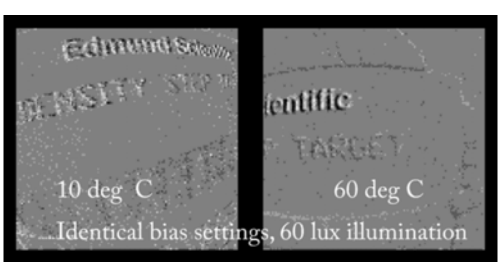
780 lux ⋮ 5.8 lux

Edmund 0.1 density chart
Illumination ratio=135:1

Low light performance

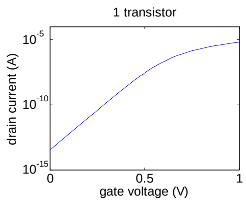
Shot under moonlight (<0.1 lux) with high contrast text
Photocurrent is <20% of dark current!
Keys to this ability
1) Low threshold mismatch
2) Pixels remember all change since last event

Integrated biases enable unadjusted operation over a wide temperature range



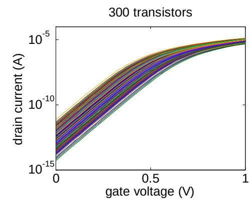
3. Chip and pixel architecture

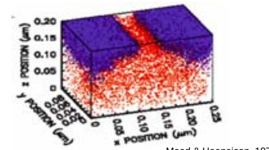
The hard problem – transistor mismatch



1. Transistor gain and exponential I-V relationship are very useful

The hard problem – transistor mismatch





Mead & Hoeneison, 1972

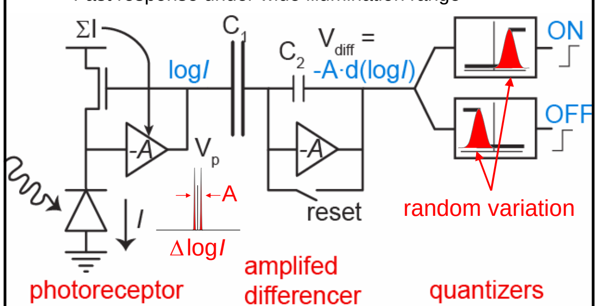
1. Transistor gain and exponential I-V relationship are very useful
2. *But* - transistor current matching is terrible
3. *But* – capacitors depend on tightly-controlled oxide and can match 100x better

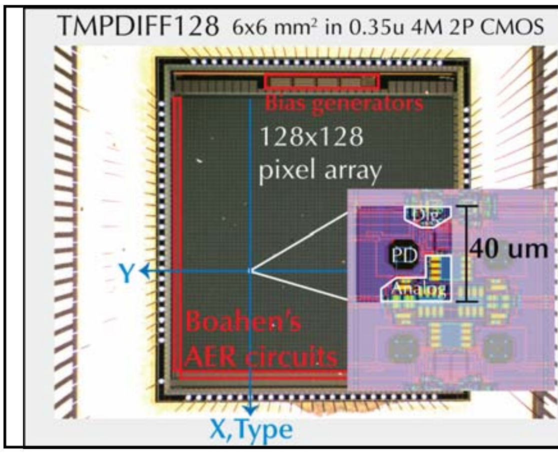
How do the pixels work?

Objectives:

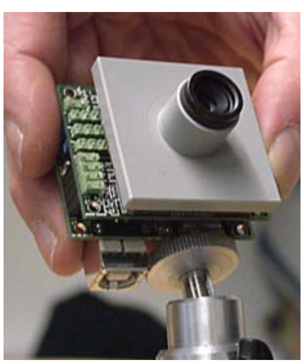
Good event-threshold uniformity

Fast response under wide illumination range





Present implementation



USB2 interface
(or direct AER interface)

Delivers stream of time-stamped addresses

Components:

- Tmpdiff128 retina
- Cypress USB chip
- 16 bit timestamp counter

Temperature & process insensitive

4. Application examples

CAVIAR spike-based vision system

High speed imaging

Low level vision (feature extraction)

High level vision (object tracking)

4. Application examples

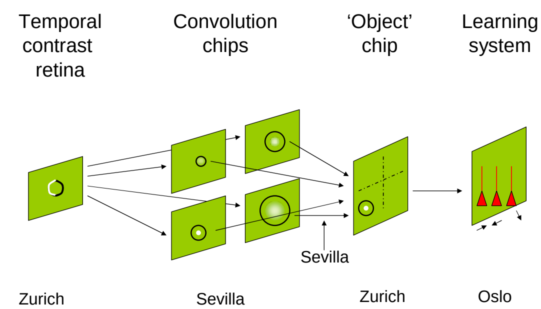
CAVIAR spike-based vision system

High speed imaging

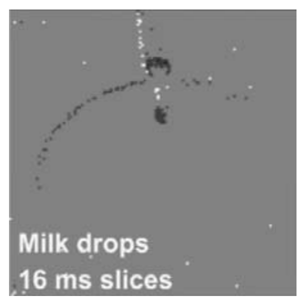
Low level vision (feature extraction)

High level vision (object tracking)

CAVIAR spike-based vision system



High speed (low data rate) imaging

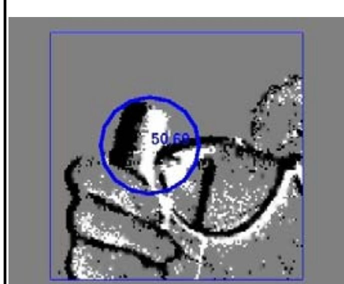


Data rate <1MBps

"Frame rate" equivalent to 10 kHz but 100x less data

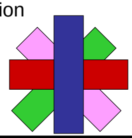
(10 kHz image sensor x 16k pixels = 160 MBps)

Low level vision: using spatio-temporal coincidence to label events with orientation and velocity

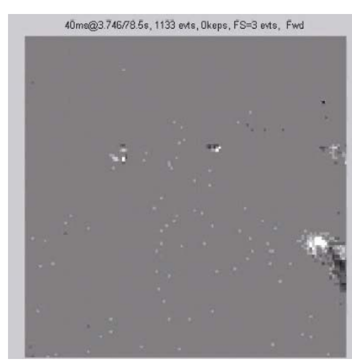


For each event:

1. Record event time in spatial map
2. Find most coincident orientation
3. Output an *orientation-event* encoding this orientation
4. Use these to compute local motion



High level vision: Tracking

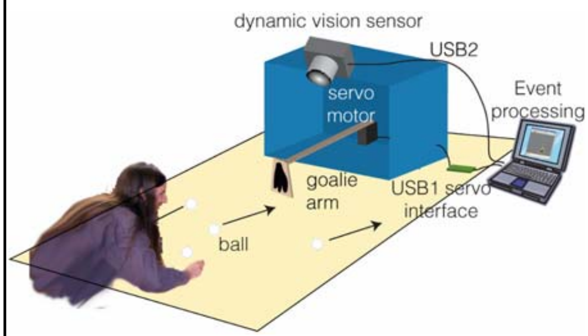


For each packet

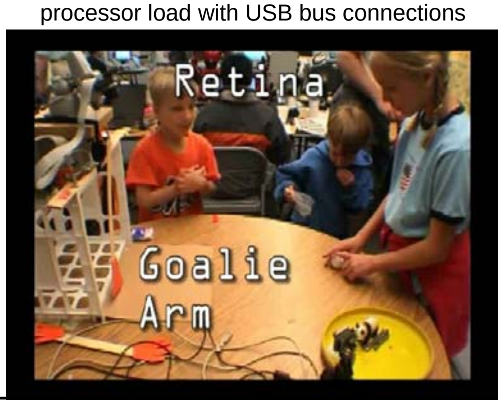
1. For each event
   1. Find nearest cluster
      a) If not within cluster, seed new cluster
      b) If within cluster, move cluster
2. Prune starved clusters
3. Merge clusters (iteratively)

Austria Research Centers Siebersdorf
“SmartEye” traffic camera

Robot Goalie



Achieves 550 “FPS” and 3 ms reaction time at 4% processor load with USB bus connections



Other applications

- Highway surveillance (SmartEye, ARCS, Vienna)
- Assembly line part identification (ARCS, Vienna)
- Tracking grasping for spinal cord recovery (Rogister, INI)
- Eye tracking (Ersboell, DTU Lyngby, EU NoE COGAIN)
- Sleep – humans, mice, worms (Tobler/Winsky, UZH Zurich)
- Hydrodynamics (Hafliger, Oslo)
- Tracking fruit fly wing beats (Fry, UZH-ETH Zurich)
- Tracking walking flies (Dickenson lab, Caltech)
- Human movement analysis (Perona lab, Caltech)
- Locust antennal movements (Huston, Caltech)
- Microscopic organisms and Brownian motion (Wu, Caltech)
- Tracking satellites (Assad, JPL)
- Fluorescence / Phosphorescence imaging (Arian, JPL)
- Calcium imaging of neural activity (Kanold, Maryland)
- Driving with spikes (Delbruck, UZH-ETH Zurich)
- Reinforcement learning for slot car racing (Riedmiller, Germany)

siliconretina.ini.uzh.ch

Telluride Neuromorphic Engineering Workshop