Building high performance neuromorphic and bioinspired systems:

Spike-based digital vision

Tobi Delbruck

Inst. of Neuroinformatics UZH-ETH Zurich Switzerland

Patick Lichtsteiner, PhD/postdoc project

Raphael Berner PhD, Christoph Posch, (ARC)

Funding: UZH-ETH Zurich, EU FET-Open Project CAVIAR, ARC
Silicon design: Kwabena Boahen (Stanford)
Giacomo Indiveri & Srinjoy Mitra (UZH-ETH)

Computer vs. Brain (What is neuromorphic engineering?)

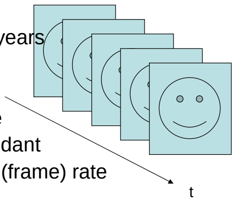
At the system level, brains are at least 1 million times more power efficient than computers. Why?

Cost of elementary operation (turning on transistor or activating synapse) is about the same. It's not some magic about physics.

Computer	Brain
Fast global clock	Self-timed
Bit-perfect deterministic logical state	Synapses are stochastic! Computation dances: digital→analog→digital
Memory distant to computation	Memory at computation
Fast, high resolution, constant sample rate, power hungry analog-to-digital converters	Low resolution adaptive data-driven quantizers (spiking neurons)

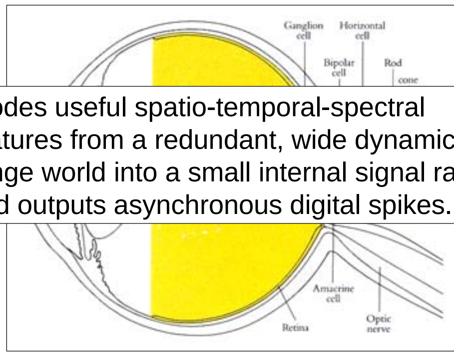
An event-driven approach to vision

Frame-based imagers
- dominated machine vision for 40+ years
- + everyone understands them
- + allow small pixels
- – usually have poor dynamic range
- – make images that are very redundant
- – impose a uniform limited sample (frame) rate



Biology
- + uses data-driven digital events with analog time
- + uses local gain control to
- + massively reduce spatial and temporal redundancy
- + is good at doing vision

Function of the retina



Encodes useful spatio-temporal-spectral features from a redundant, wide dynamic range world into a small internal signal range and outputs asynchronous digital spikes.

Rodieck, 1998

Spike-based vision sensors

1990 1995 2000 2005

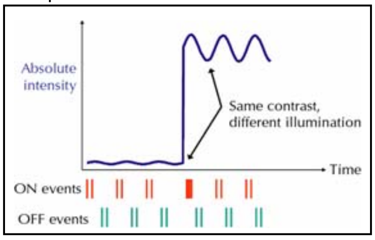
- Mahowald & Mead outer retina (SciAm 91)
- UPenn Magno-Parvo silicon retina (TBME 02)
- **CSEM VISe steerable-filter contrast vision sensor (JSSC 03)**
- Yale Univ. "Octopus" imager (JSSC 04)
- JHU Temporal Change Threshold Detection Imager (JSSC 07)
- **ETH Dynamic Vision Sensor (JSSC 08)**
- Sevilla Spatial Contrast Silicon Retina (TCAS 2008)

*See **siliconretina.ini.uzh.ch** for links*

ETH Dynamic Vision Sensor (DVS)

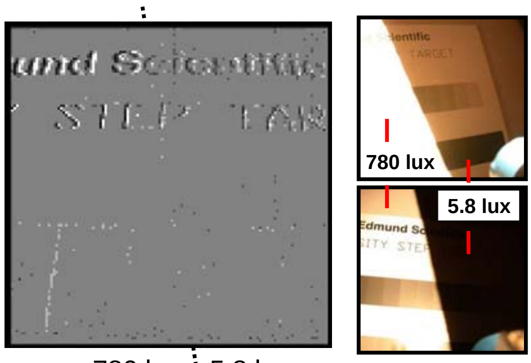
Principle of temporal-contrast operation

- This *asynchronous* vision sensor responds to *relative intensity change.*
- 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



This operation efficiently encodes local changes in scene reflectance with good temporal resolution and wide dynamic range

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

DVS camera

USB interface

Delivers stream of time-stamped addresses

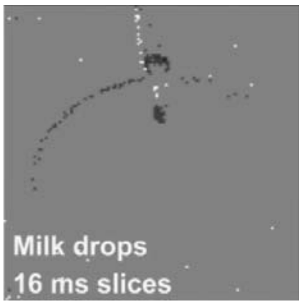
Components:
Tmpdiff128 retina, USB chip, CPLD

Temperature & process insensitive

DVS application areas

- Low level features (Delbruck, Zurich)
- Car and person counting (ARCS, Vienna)
- Fast robotic vision (Delbruck, Zurich)
- Neuromorphic spike-based hardware systems: CAVIAR
- Assembly line part identification (ARCS, Vienna)
- Tracking grasping for spinal cord recovery (Rogister, Zurich)
- Eye tracking (Ersboell, DTU Lyngby, EU NoE COGAIN)
- Sleep – humans, mice, worms (Tobler/Winsky, UZH Zurich)
- Hydrodynamics (Hafliger and Jensen, 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 (Besselmann & Delbruck, Zurich)
- Reinforcement learning for slot car racing (Riedmiller, Germany)

High speed (low data rate) imaging



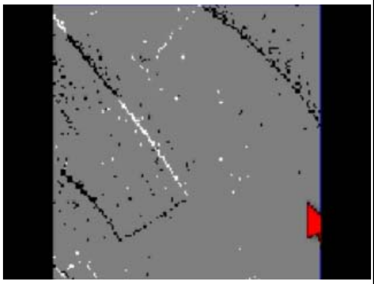
Data rate <1MBps

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

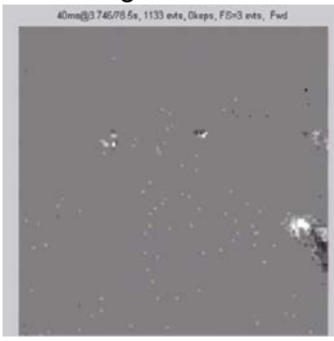
(10 kHz image sensor x 16k pixels = 160 MBps)

Low level vision: Measuring local and global motion

1. Start with raw events that represent temporal contrast
2. Use **spatio-temporal event coincidence** to label events with local edge **orientation**
3. Label these ***OrientationEvents*** with local direction and speed
4. Integrate these ***DirectionSelectiveEvents*** to compute global translational, rotational, and radial 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)

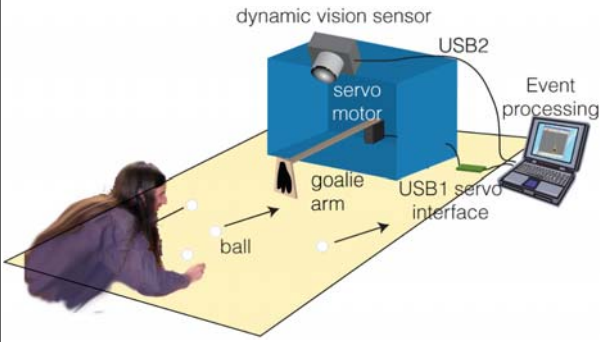
No frame memory (100 bytes/object).
No frame correspondence problem

Commercial applications by ARC (Austria Research Corporation)

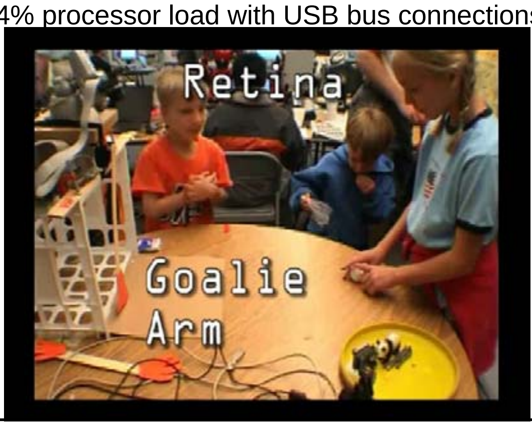
TDS "SmartEye" - traffic data sensing. DVS+2.5W Blackfin DSP.

UCOS - people counting using 3D

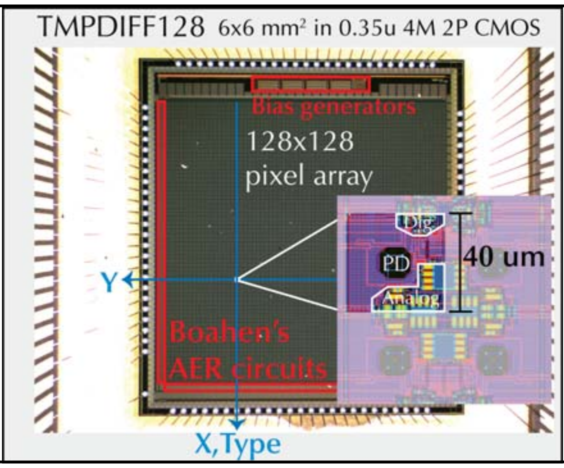
Robot Goalie



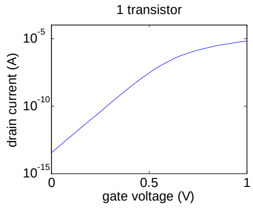
Achieves 550 "FPS" and 3 ms reaction time at 4% processor load with USB bus connections



DVS chip and pixel architecture

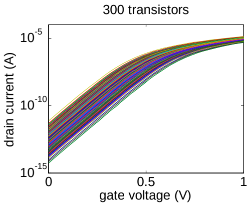


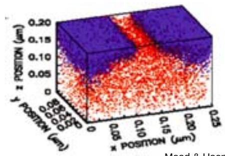
The hard problem – transistor mismatch



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

The hard problem – transistor mismatch



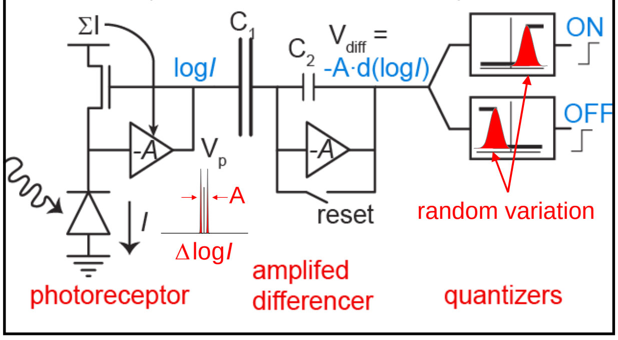


Mead & Hoeneison, 1972

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



siliconretina.ini.uzh.ch

The DVS (**dynamic vision sensor)** provides

- Meaningful fast asynchronous events
- Logarithmic, self-timed, quantizing pixel
- Rapid and precise timing of scene reflectance changes with wide intra-scene illumination range
- Unprecedented specifications: **2% mismatch, 120dB dynamic range, 23mW power consumption, 15us minimum latency**
- **A new way to think about doing vision**

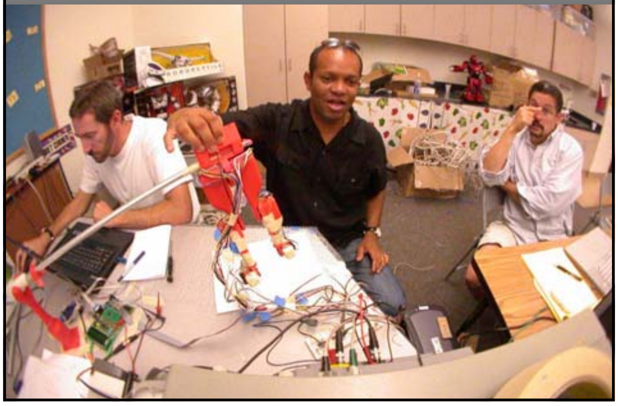
Additional accomplishments

- Integrated digital on-chip biases (open sourced)
- A standard high-speed USB computer interface (open)
- 350+ classes for event-driven digital vision (*jAER*, *open*)
- Winner of 7 IEEE awards including 2006 ISSCC Jan Van Vessem Outstanding European Paper

Telluride Neuromorphic Engineering Workshop
3 week annual July workshop
Telluride, Colorado, USA

www.ini.uzh.ch/~tobi

Hands-on projects, new collaborations



Telluride "Grand Challenge" 2006
Robot race

Housing and part of travel subsidized
Application deadline March 23

Inst. of Neuroinformatics, UZH-ETH Zurich, Switzerland
www.ini.uzh.ch

Specifications of TMPDIFF128

Spatial resolution	128 x 128
Pixel size Fill factor	40 x 40 um^2 (200 x 200 lambda^2) 9.4% (photodiode area 151 um^2)
Events/sec	max ~1 M events/sec
Power consumption	23 mW @ 3.3V (core 1.5mA, AER 0.5mA)
Illumination operating range	120 dB: <1 lux to > 100 klux scene illumination with f/1.4 lens
Response latency	Down to 15 us; ~1ms indoor illum.
Uniformity (1-σ contrast threshold variation)	2.1% contrast