Building an asychronous silicon retina and using it for digital vision

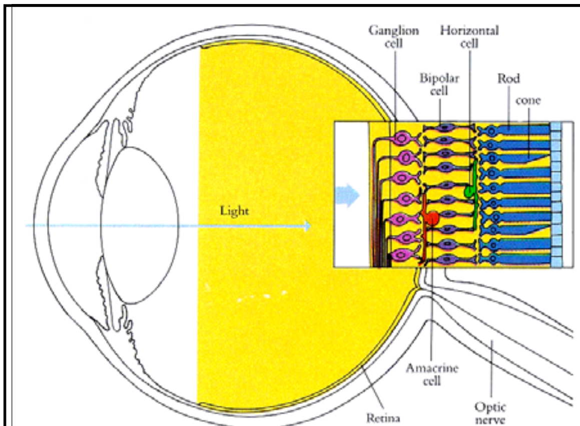
Tobi Delbruck
Inst. of Neuroinformatics
UZH-ETH Zurich

Patrick Lichtsteiner PhD project

Also C. Posch (ARCS)

siliconretina.ini.uzh.ch

Funding: UNI-ETH Zurich, EU Project CAVIAR, ARCs research
Silicon IP: K. Boahen (UPenn/Stanford) G. Indiveri & S. Mitra (UZH-ETH)



Rodieck, 1998

Related work on spike-based vision sensors

- 1992: Mahowald & Mead outer retina
- 2002: UPenn Magno-Parvo silicon retina
- **2003: Devise steerable-filter contrast vision sensor**
- 2004: JHU Temporal difference detection imager
- 2004: Yale Univ. “Octopus” imager

See Delbruck Wiki for references

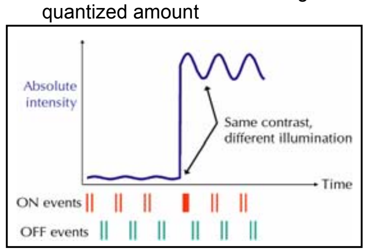
Outline

2. Basic characteristics of the temporal contrast silicon retina
3. Motivation for this design
4. Pixel and chip design
5. Application examples in digital vision

2. Basic characteristics

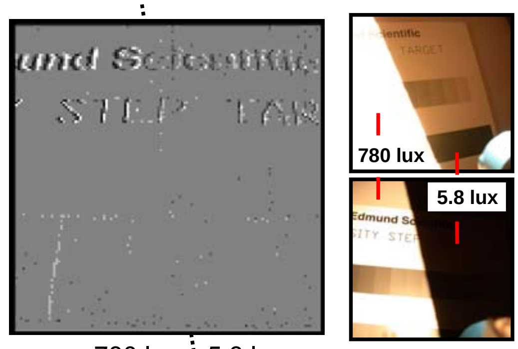
Principle of temporal-contrast operation

- This *asynchronous* vision sensor responds to *relative intensity change.*
- It emits digital *address-events* that encode the *identities* of pixels that see these changes.
- 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

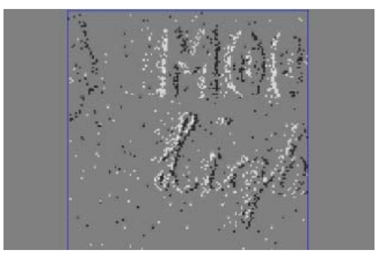
Uniform event threshold and wide dynamic range



780 lux ⋮ 5.8 lux

Edmund 0.1 density chart
Illumination ratio=135:1

Low light performance



Shot under ¾ moon (<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

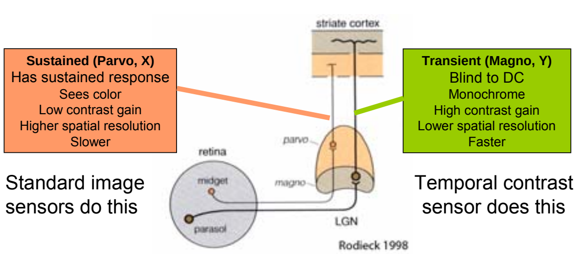
2a. Engineering motivation

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

Biological retinas
- + use data-driven digital events with analog time (spikes)
- + use local gain control
- + massively reduce spatial and temporal redundancy
- + are good at doing vision

2b. Neuromorphic motivation for temporal contrast response

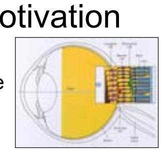
A partitioning into sustained and transient visual pathways is universal in biology from insects to us



Standard image sensors do this

Temporal contrast sensor does this

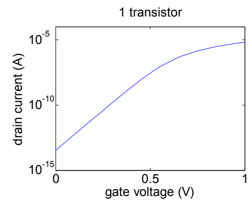
2a. Engineering motivation

Camera  Eye 

Camera	Eye
Clocked	Asynchronous
Image frames	Data driven events (spikes)
All pixels have identical integration time and gain → low dynamic range	Local gain is controlled by local spatio-temporal context
Uses one or a few fast, high precision, power hungry ADCs	Low resolution, adaptive, data-driven quantizers (spiking ganglion cells)
Highly redundant output, both within and between frames	Every spike event probably *means* something

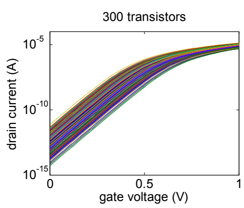
3. Chip and pixel architecture

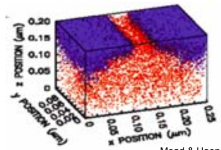
The hard problem – transistor mismatch



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

The hard problem – transistor mismatch





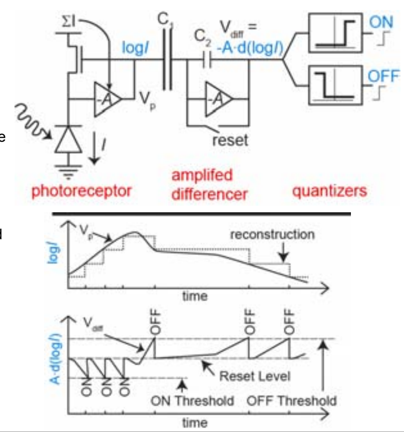
Mead & Hoeneison, 1972

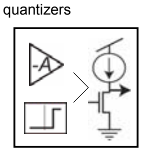
1. Transistor gain and exponential I-V relationship are very useful
2. *But* - transistor current matching is terrible
3. *But* – capacitors depend on tightly-controlled oxide and can match 100x better

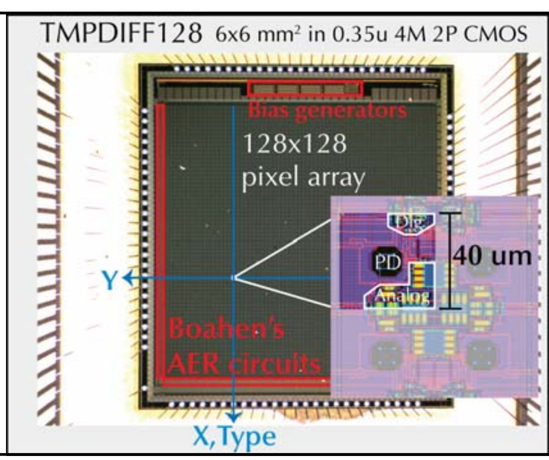
Key challenge:
Transistor mismatch

Objectives:
1. Good threshold uniformity
2. Fast response under wide illumination range

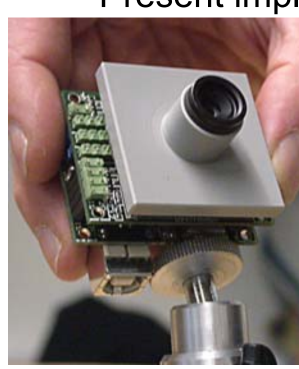
Key techniques:
1. Active logarithmic transimpedance
2. DC offset removal and amplifying precisely *before* imprecise quantizers







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

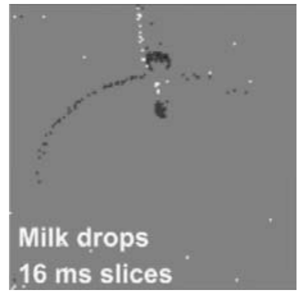
High speed imaging

Spike-based vision systems

Low level vision

High level vision

High speed (low data rate) imaging

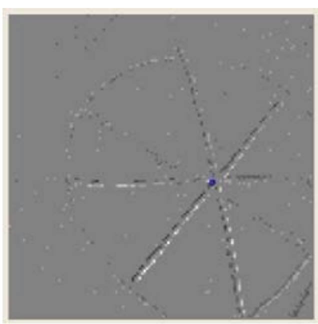


Data rate <1MBps

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

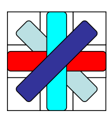
(10 kHz image sensor x 16k pixels = 160 MBps)

Low level vision: Orientation from temporal coincidence



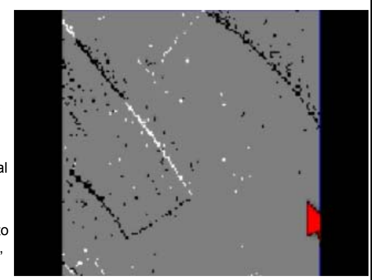
For each event:

1. Record event time at pixel
2. Find most coincident orientation, it has Δt
3. If Δt **<threshold**, output an event encoding this orientation

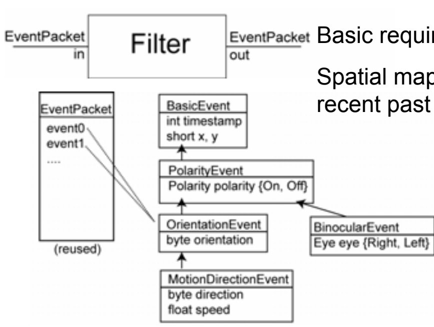


Low level vision continued: Measuring local and global motion

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



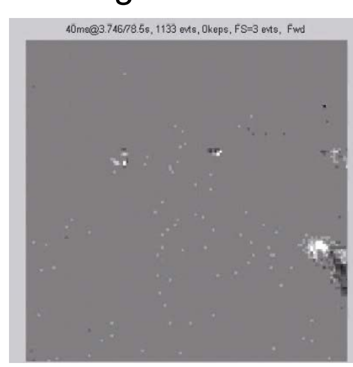
Events gain extended interpretation as they are processed



Basic requirement:

Spatial map of most recent past events

High level vision: Tracking



For each packet

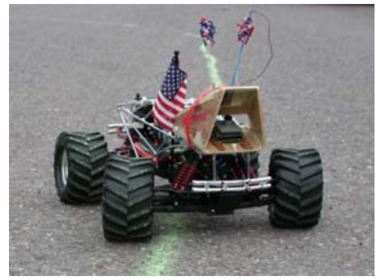
1. For each event
   1. Find nearest cluster
      a) If not within cluster, seed new cluster
      b) If within cluster, move cluster
2. Prune starved clusters
3. Merge clusters (iteratively)

Using tracking for robotics: Fast Goalie

Servo Controller

Retina tracking view

Driving with spikes (ongoing)



Traxxus E-Maxx: **50 kph**; Retina; Sony Vaio WinXP; USB hub; USB Servo Controller
Goal: Drive fast on drawn racetrack

Austria Research Centers Siebersdorf "SmartEye"

Other applications being explored

- Highway surviellance (SmartEye, ARCS, Vienna)
- Assembly line part indentification (ARCS, Vienna)
- Street Surveillance video (Delbruck)
- Tracking rat grasping for spinal cord recovery (Rogister, INI)
- Sleep – humans, mice, worms (Winsky, UZH Zurich)
- Tracking fruit fly wing beats (Fry, UZH-ETH Zurich)
- Locust antennal movements (Huston, Caltech)
- Hydrodynamics (Hafliger, Oslo)
- Measuring motion of microscopic cilia (Wu, Caltech)
- Sports motion analysis, e.g. baseball (Arian, Caltech)
- Tracking satellites (Assad, JPL)
- Flourescence/Phospheresence imaging (Arian, JPL)
- Calcium imaging of neural activity (Kanold, Maryland)

Summary

This temporal contrast vision sensor provides
- Meaningful asynchronous events
- Precise timing of scene reflectance changes
  - Wide intra-scene illumination range
  - Low power consumption
- Unprecedented specifications: **2% mismatch, 120dB dynamic range, 23mW power consumption, 15us minimum latency**
- A novel logarithmic, self-timed quantizing pixel
- **A new way to think about doing vision**

Additional accomplishments
- Integrated digital on-chip biases
- A standard high-speed USB computer interface
- 200+ classes for event-driven digital vision (*jAER*)
- Winner of 5 IEEE awards including 2006 ISSCC Jan Van Vessem Outstanding European Paper