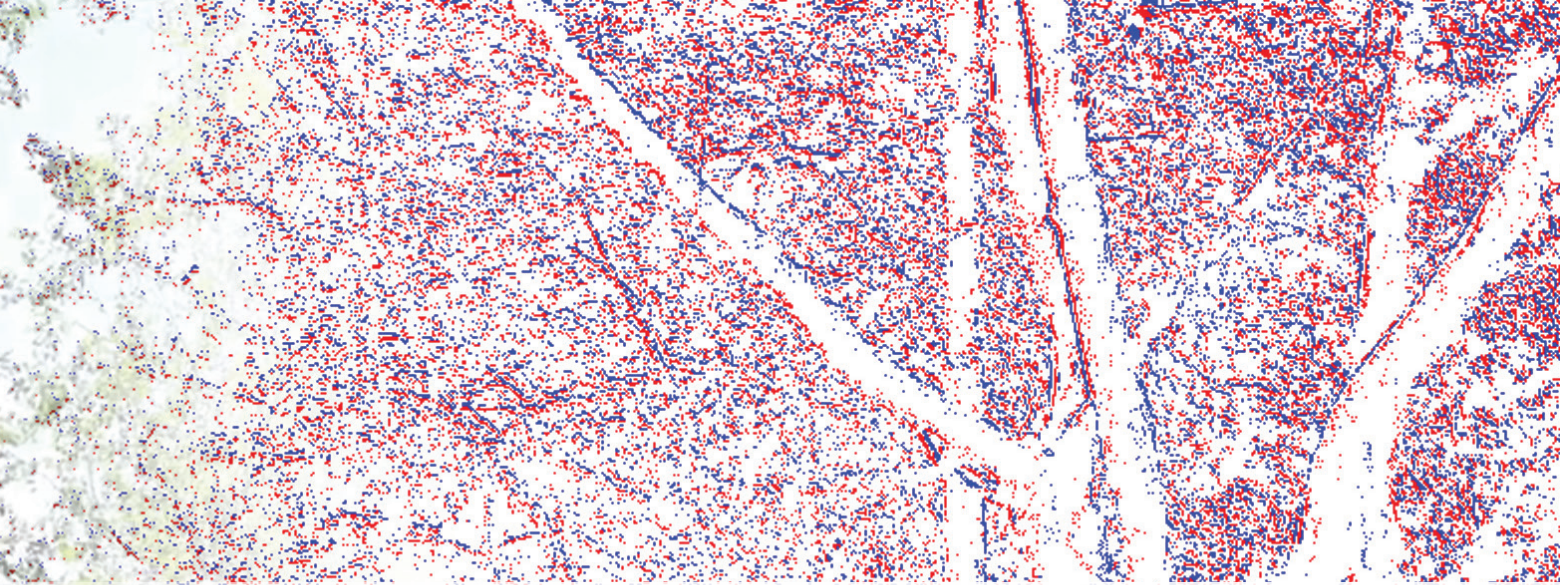
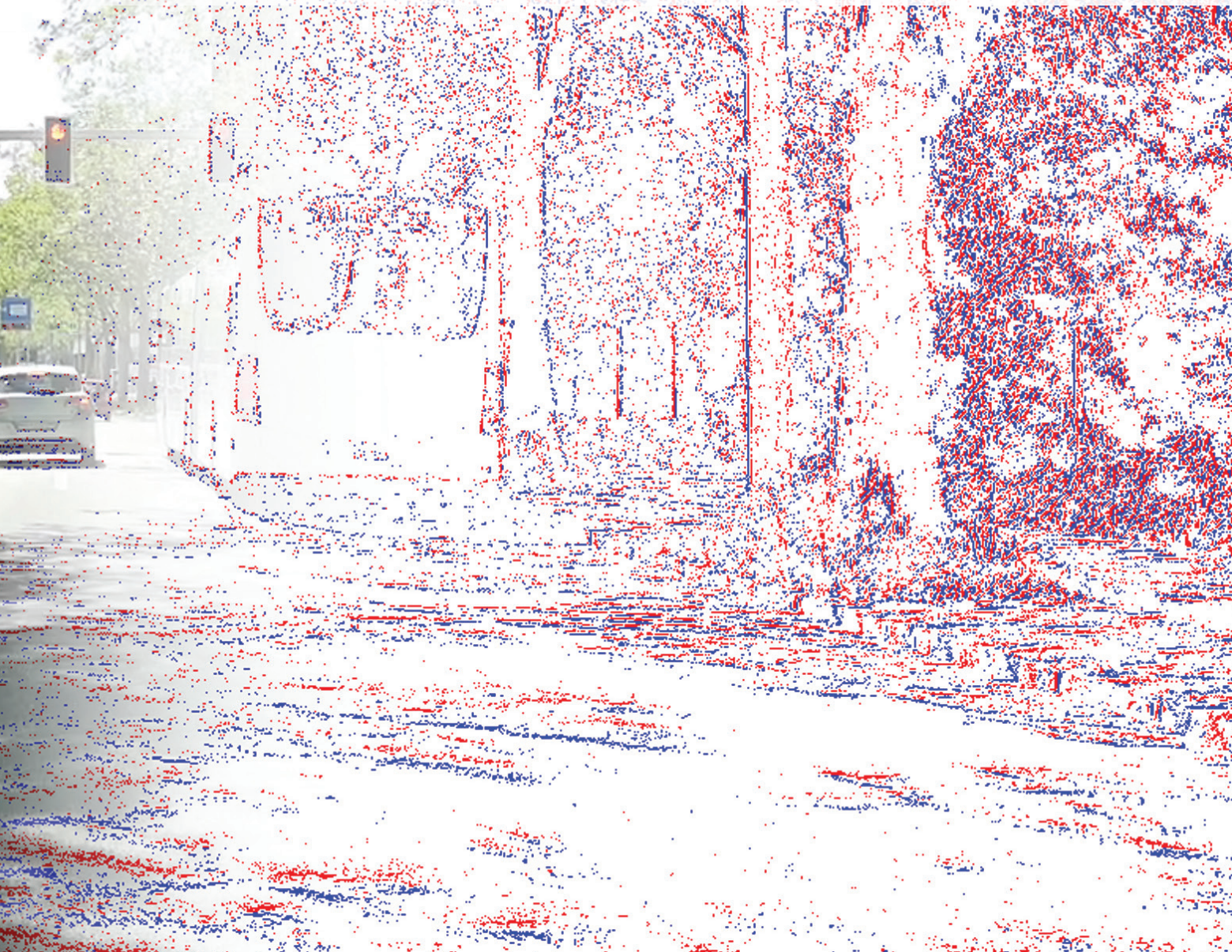


Susan Curtis

Event Cameras: A New Imaging Paradigm



Unlike conventional cameras, neuromorphic cameras mimic how human eyes work by detecting and recording only changes in a scene, opening doors to new imaging possibilities.



Conventional camera

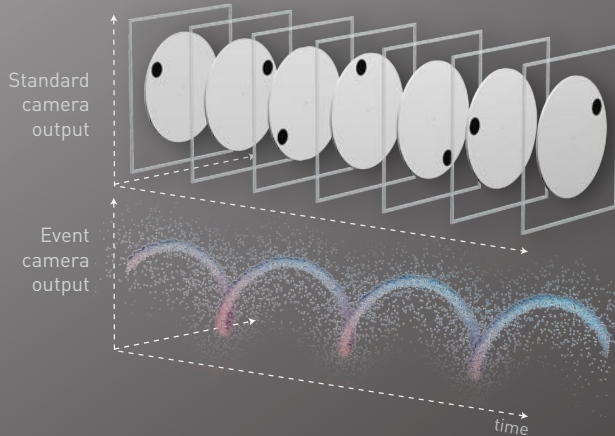
Intensity changes

Event camera

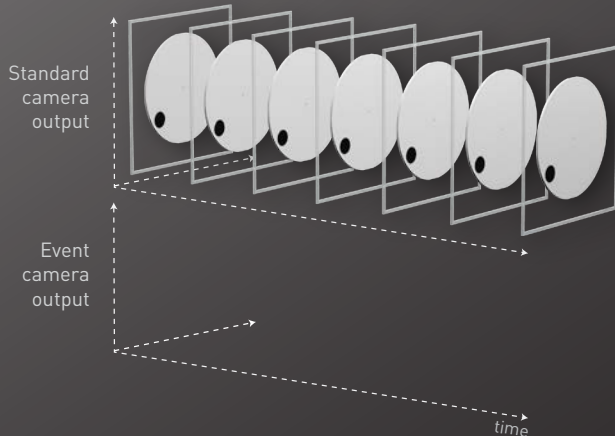
How event cameras work

Unlike ordinary cameras, which capture an entire scene, event cameras have independent pixels that respond to changes in intensity—making it possible to capture fast-moving events that would appear blurred in a conventional camera

For a rotating disk marked with a dot, a standard camera records a series of images of the entire disk; an event camera continuously captures the position of the dot alone, the only changing element



If the disk stops rotating, the standard camera still records the separate images; the event camera records nothing, as nothing is changing



If the disk rotates very fast, the standard camera output is blurred, but the event camera captures the detail

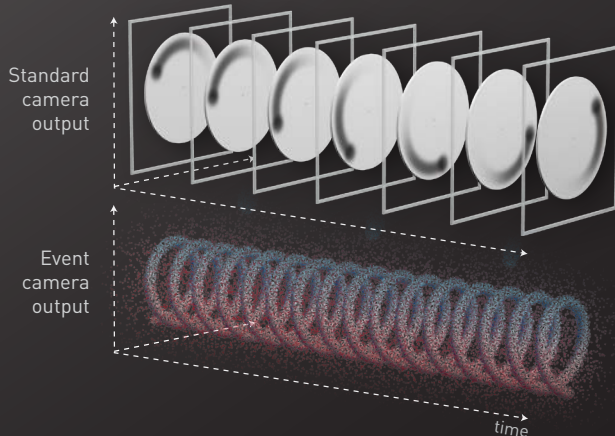


Illustration by Phil Saunders

Imagine the human eye as an optical sensor. It captures a constant stream of visual data, converting incoming light into electrochemical signals that allow the brain to create a panoramic view of the surroundings. But the retina that lines the back of the eye achieves much more than a simple photodiode: a succession of specialized cells decodes the raw optical data, extracting the most important features, discarding redundant information, and sending to the brain only what's useful for decision making and producing a dynamic image. This crucial pre-processing step allows the brain to make sense of vast amounts of raw optical data more quickly, reconstructing a 3D view of the world in real time and allowing humans to react almost instantaneously to fast-moving events.

It's hardly surprising, then, that this natural imaging solution's elegance and efficiency have inspired scientists and engineers to create artificial vision systems that mimic biology. In the late 1980s, at the California Institute of Technology (Caltech), USA, Ph.D. student Misha Mahowald worked with microelectronics pioneer Carver Mead to create the first silicon chips to emulate the biological function of the retina, generating a similar response in real time to the signals observed in the human visual system.

"We have taken the first step in simulating the computations done by the brain to process a visual image," wrote Mahowald and Mead in "The Silicon Retina," a landmark 1991 article in *Scientific American*. "Our success persuades us that this approach not only clarifies the nature of biological computation but also demonstrates that the principles of neural information processing offer a powerful new engineering paradigm."

Only the changes

The breakthrough demonstrations by the two Caltech researchers ignited the field of neuromorphic engineering—the idea that building electronic systems that replicate the neural architecture of the brain might enable more efficient analysis and computation by taking a differential approach instead of an integrated one, while also offering crucial insights into the way the brain works. Neuromorphic engineering now extends to all areas of sensory perception and processing, with initiatives such as the EU's Human Brain Project even attempting to build artificial brains to explore the complexity of human cognition and to devise more powerful and more efficient computer architectures.

It's hardly surprising that a natural imaging solution's elegance and efficiency have inspired scientists and engineers to create artificial vision systems that mimic biology.

Meanwhile, the early silicon retinas built by Mahowold and Mead have evolved into a new paradigm of neuromorphic vision systems that can respond much more quickly to high-speed events—while also consuming far less power than standard cameras.

These bio-inspired image sensors exploit the same optical design and layout as a standard CMOS camera, and indeed have benefited from the manufacturing improvements that over the last decade have reduced the pixel size and enhanced the spatial resolution of all silicon-based devices. The difference is that each pixel in a neuromorphic camera operates independently, triggering a response only when the intensity of light exceeds a predefined threshold. In contrast to a conventional camera, which records a complete picture of the scene at regular time intervals, these image sensors capture an event only when one of these smart pixels detects some sort of change.

"An event camera only measures motion in the scene," explains Davide Scaramuzza, a professor of robotics and perception at the Institute of Neuroinformatics, a joint research center of the University of Zurich and ETH Zurich in Switzerland that Mahowold helped to establish. "Whenever a pixel detects a change in intensity, possibly caused by motion or by blinking patterns, that pixel triggers an event. If nothing moves, there is no information." In practice, that means the images recorded by an event camera show only the contours of moving objects, rather than the full-color pictures captured by a normal video camera.

This event-driven paradigm has several important consequences for image sensors. For a start, recording the time and location of each discrete event makes it possible to detect any movement in the scene with microsecond resolution. "Most digital cameras record a new frame once every 30 milliseconds, but an event-based camera produces a constant stream of asynchronous events," explains Scaramuzza. "The output is continuous in both space and time."

That makes event cameras ideal for applications that demand fast response times, such as robotics and self-driving cars. As an example, Scaramuzza and his colleagues have tested the performance of an event camera when a ball is thrown at an autonomous



Event cameras can capture the super-fast dynamics of a scene more effectively than a standard image sensor, and can enable an autonomous drone to detect and dodge a flying ball around 10 times more quickly than one equipped with a frame-based camera.

Robotics and Perception Group, University of Zurich, Switzerland

drone. "The ability to dodge a moving obstacle is particularly difficult in robotics since in this case the speed between the drone and the ball can reach around 10 m/s, or 36 km/h," he says. "With a standard camera, you would need to wait at least 30 milliseconds to capture two successive frames, which is too long for the drone to safely avoid the ball."

In contrast, Scaramuzza's tests showed that an event camera can detect the ball and instruct the drone to make an evasive maneuver within just 3.5 ms—almost 10 times faster than with a standard camera. In addition to allowing drones to react more quickly to unpredictable events, this type of system could improve the responsiveness and safety of autonomous driving systems, potentially reducing reaction times to fractions of a millisecond.

Such low latency also offers important advantages for industrial automation and machine vision, as well as for augmented- and virtual-reality systems. For the holograms produced by these systems to appear realistic, explains Scaramuzza, any gesture or movement of the head must be rendered on the virtual scene



A photo of Astrosite, a mobile observatory built by Greg Cohen and colleagues that fits inside a shipping container. The high dynamic range of an event camera allows the observatory to operate day or night.
International Centre for Neuromorphic Systems, Western Sydney University

within about 10 ms. “Your brain perceives that something is wrong if it takes any longer,” he says. “You might get motion sickness, or it just doesn’t seem true to life. Anything you can do to shorten the interval is an advantage.”

Sparse data, high dynamic range

Event-based sensors offer other benefits, too. Each discrete event contains only a tiny amount of information, which can massively reduce the amount of data that must be processed and stored—particularly in applications where the data are inherently sparse. Such low-data-rate operation can drive down the device’s power consumption to the milliwatt regime, an order of magnitude lower than for frame-based image sensors.

Event cameras also boast a very high dynamic range—typically up to 140 dB, compared with 60 dB for a smartphone camera. That means they can capture high-quality data both in bright sunlight and at night, and thereby cope with the changing lighting conditions typical of automotive applications and industrial environments.

Such a unique combination of properties has proved particularly useful for tracking satellites and other objects in space, as Greg Cohen at Western Sydney University in Australia has discovered. “For space applications, all we really want is high dynamic range and a camera that captures the movement of stars and satellites,” he says. “With a normal telescope, you spend a lot of energy and bandwidth taking pictures of empty space, but a camera that only senses movement strips out all that irrelevant information.”

Once Cohen and his team had fitted an event camera to a telescope, they realized that it offered a whole new approach to space observation. “An event-based device cannot compete with a sensitivity of a camera that integrates over time, but it can capture small changes that might otherwise get missed,” he says. “That’s really important for certain tasks, such as looking at satellites, because you can detect small movements that indicate whether it’s tumbling or drifting off course.”

Since event cameras do not capture data over a fixed time interval, images can even be recorded when the telescope is being moved. That has inspired Cohen and his colleagues to build a mobile observatory that fits inside a shipping container, allowing it to be transported anywhere in the world. “You can just put it down, plug it in and start doing some observing wherever you might be,” he says. “You can even stop near a highway because it’s much less sensitive to vibrations and other types of motion. Sometimes I even like to tap the telescope to make it easier to see things.” The low power and low data rate of event cameras have also enabled Cohen and his team to design two space-based instruments that are currently in orbit on the International Space Station.

Commercial potential

The novel applications of event cameras now being demonstrated in both academia and industry have been enabled in large part by ongoing advances in sensor technology. In the 1990s, Mahowold collaborated with Tobi Delbrück at Zurich’s Institute of Neuroinformatics to improve the design of the early silicon retinas, with Delbrück then working with Patrick Lichtsteiner to unveil the first practical system for event-based sensing in 2005. This demonstration device, with its 64×64-pixel array, kickstarted more than a decade of innovations in design and engineering, yielding event-based sensors that pack in more than a million pixels to improve both their resolution and dynamic range.

Commercial devices are also now emerging that leverage the same manufacturing technologies as more

The novel applications of event cameras now being demonstrated in both academia and industry have been enabled in large part by ongoing advances in sensor technology.

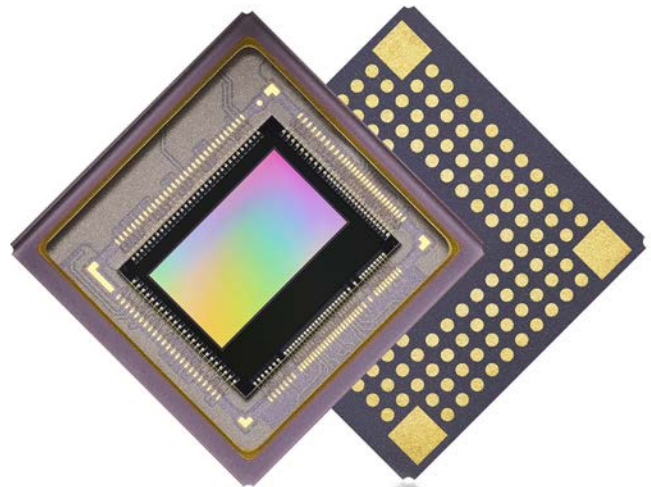
established types of image sensor. The French startup Prophesee has collaborated with Sony to release a new event-based vision sensor with a center-to-center distance between pixels, or a pixel pitch, of just $4.86\ \mu\text{m}$ —compared with $15\ \mu\text{m}$ for Prophesee's previous product, Metavision. In the new device, the photodiode is fabricated using a dedicated process and then stacked on top of the transistor layer, improving the electro-optic performance and using the silicon area more efficiently. "In our previous sensor, the photodiode occupied only a quarter of the pixel's surface, which meant we were losing more than half of the photons," explains Luca Verre, Prophesee's CEO. "The fill factor in the new design is more than 80%, which improves the sensitivity of the sensor as well as its low-light performance."

Two versions of the sensor are now available—one with a 1280×720 -pixel array and a smaller 640×512 device—with an evaluation kit to support rapid prototyping. Initially, Prophesee and Sony are targeting applications in industrial automation, with Prophesee having already worked with companies specializing in machine vision such as Imago, CenturyArks and Lucid to demonstrate practical event-based solutions that can be used for real-time monitoring and control of production processes. "The partnership with Sony provides extra credibility for our efforts so far," says Verre. "We will also be working together to develop customer opportunities and to boost the adoption of the technology across different applications."

Need for software

One key market in the two companies' sights is the Internet of Things (IoT), where intelligent vision systems operating at the edge of the network could capture and process information locally, rather than transmitting all the data to a central computer. Such always-on systems could be used for autonomous monitoring of dynamic processes, such as the flow of traffic or people in busy areas, as well as for smart-home devices and human-machine interfaces.

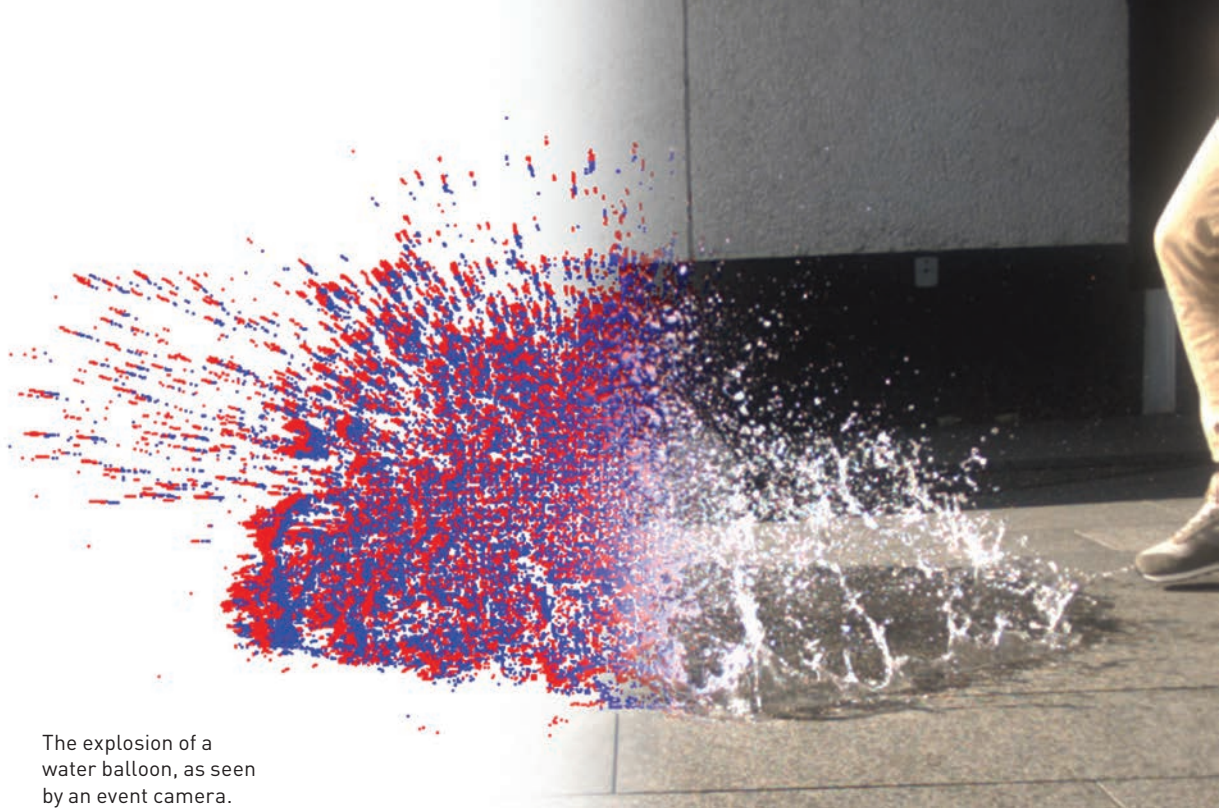
However, an intelligent vision system also needs software to make sense of the raw optical data, and most conventional algorithms for computer vision



The IMX636 event-based vision sensor was produced in a collaboration between Sony and Prophesee.
Prophesee

and deep learning process frame-based information. One simple solution is to impose an artificial frame rate on the event data, essentially summing all of the discrete events that have been triggered within a short time interval. This offers flexible frame rates while still using readily available software, but clearly sacrifices some of the temporal resolution that event cameras can achieve. "It's not great to grab this interesting data out of the camera—which is providing both the time and location of the change—and then wrangling it back into a frame," says Cohen. "To realize the benefit, you really need to rethink the way you process the information."

Researchers have therefore devised alternative algorithms that return an output every time a new event is triggered. As with standard computer vision, most of these approaches rely on neural networks consisting of thousands or even millions of artificial neurons, or nodes. In conventional systems, all the nodes in the network are connected together, so the whole network is updated when a new frame of data is processed. In an event-by-event framework, however, each artificial neuron only produces a response—or a spike—once it reaches a specific threshold.



The explosion of a water balloon, as seen by an event camera.

Robotics and Perception Group, University of Zurich, Switzerland

These so-called spiking neural networks more closely mimic the way the brain works, with each neuron transmitting a signal to other nodes in the network only when it is triggered by an event. “We try to build systems that work iteratively, and that also use the time between events as a source of information,” explains Cohen. “Processing the data as it arrives means that you don’t have to store it, which makes it possible to handle large amounts of information with very little bandwidth.”

Spiking algorithms have already proved their ability to track moving objects with very low power consumption, but in most cases, they still run on computer processors that use around 10 W in standby mode. “We need to co-design the hardware and software for processing this event-based data,” says Scaramuzza. “That will make it possible to demonstrate complete vision solutions that leverage the microsecond temporal resolution of event cameras while also operating on very little power.”

Such a complete imaging solution would combine an event camera with a so-called neuromorphic processor—an alternative computing platform rooted in the ideas of Carver Mead that exploits transistors to emulate the way the brain works. Instead of the sequential approach taken by conventional digital processing technologies, such bio-inspired processors exploit massively parallel circuits that act as artificial neurons, ready to fire whenever a new event is detected. Some of these neuromorphic processors, such as the Loihi neuromorphic research chip developed

by Intel, have already been shown to operate much more quickly and on much less power than a conventional computer chip.

However, the best performance can be achieved by developing application-specific solutions that combine neuromorphic architectures for sensing and processing. “If you can specialize the camera and specialize the processing, you can achieve the power efficiency, robustness and reliability that biology offers,” says Cohen. “That’s what we’re all aiming and striving for, but we’re not quite there yet.”

Technologies such as field-programmable gate arrays (FPGAs) offer an accessible way for researchers to experiment with these neuromorphic approaches. “Once we have developed an algorithm to solve a particular problem, we simulate its performance using a conventional computer,” Cohen explains. “If the results look promising we can start to build the algorithm on an FPGA, and if that looks good we can start to build the circuitry in silicon. With each progression, we can improve the power efficiency by orders of magnitude.”

Integrated solutions

Building a solution directly in silicon presents a whole new challenge, but fully integrated neuromorphic solutions for vision applications are now starting to emerge. In 2019, for example, the Chinese/Swiss startup SynSense released a neuromorphic processor that has been optimized to work with event-based image sensors. SynSense’s DYNAP-CNN chip incorporates more

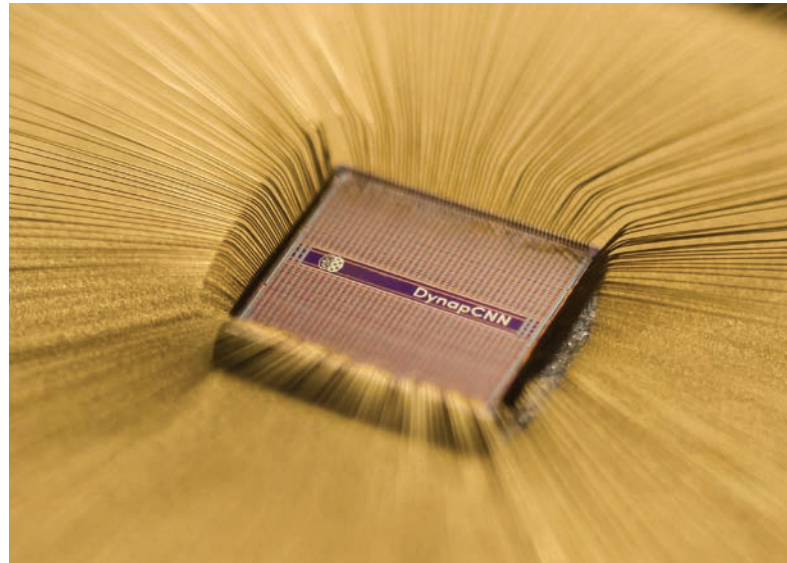
“If you can specialize the camera and specialize the processing, you can achieve the power efficiency, robustness and reliability that biology offers,” says Cohen.

than a million spiking neurons and four million programmable parameters, allowing implementation of different algorithms for processing event-based data, and offers a latency below 5 ms and a power efficiency that SynSense claims is 10 to 100 times greater than standard processors.

SynSense has now partnered with Prophesee to build a single-chip solution that combines its DYNAP-CNN processor with the Metavision event sensor. The initial objective will be to create a small device with a 128×128-pixel array, which should be sufficient for short-range applications such as smart-home devices, facial recognition and simple human-machine interfaces. “By integrating everything on the same chip, we will be able to process data on the fly, allowing us to drive down both the power consumption and the latency,” says Verre. “We also believe we can manufacture the chip at a low enough cost to address always-on IoT applications.”

The first devices are expected to be ready to ship to customers by the end of the year. Verre points out that Prophesee has already shown that the data recorded with its sensor can be processed by Intel’s Loihi chip, while the DYNAP-CNN processor has been designed to accept a direct feed of data from an event camera. “One of the challenges will be programming the chip for a specific task,” says Verre. “At least in the beginning, we need to work with our customers to develop their applications because it will require some dedicated software tools as well as specialized data collection for training the neural network.”

Despite such impressive advances in technology, event cameras have yet to enter the mainstream. One issue has been the cost, with state-of-the-art event cameras typically priced at around US\$5,000. But Verre believes that Prophesee’s manufacturing breakthrough with Sony should make event cameras more competitive with more established imaging technologies. “We are not using any exotic technology to manufacture the sensor or the camera module,” he says. “Our pixel size is now in the same ballpark as for a time-of-flight or global-shutter sensor, which should enable the more aggressive price positioning



The DYNAP-CNN chip from SynSense has been optimized for processing event-based data.

SynSense

that’s needed to open up high-volume applications in the consumer space.”

Indeed, France-based market analysis firm Yole Développement predicts that the global market for neuromorphic technologies could reach US\$5 billion by 2030, driven largely by novel imaging solutions for mobile phones. Other consumer applications are likely to emerge in wearable technologies and smart-home devices, along with automotive applications, autonomous drones and industrial robotics. “It’s no longer a technology or manufacturing challenge, it’s more of a business challenge to get the technology adopted in some of these consumer applications,” says Verre. “That will enable us to reach the scale we need to access more advanced technology nodes and further reduce the size and cost of these neuromorphic solutions.” [OPN](#)

Susan Curtis (sxcurtis@gmail.com) is a freelance science and technology writer based in Bristol, UK.

For references and resources, go online:
optica-opn.org/link/neuromorphic-cameras.

Article References and Resources

- ▶ M.A. Mahowald and C. Mead. "The Silicon Retina," *Sci. Am.* **264**(5), 76 (1991).
- ▶ P. Lichtsteiner and T. Delbruck. "64×64 Event-Driven Logarithmic Temporal Derivative Silicon Retina," [PDF] *IEEE Workshop on Charge-Coupled Devices and Advanced Image Sensors* (2005).
- ▶ M. Davies et al. "Advancing Neuromorphic Computing With Loihi: A Survey of Results and Outlook," *Proc. IEEE* **109**, 911 (2021).
- ▶ "Neuromorphic Computing and Sensing 2021 report," *Yole Développement Market and Technology Report* (2021).