Stereo vision using computing architecture inspired by the brain
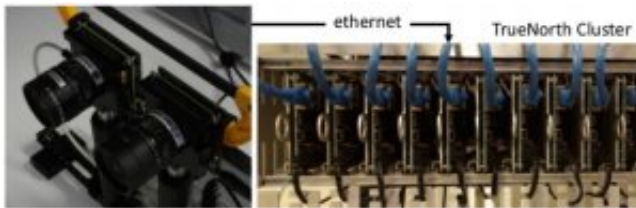
June 20 2018, by Alexander Andreopoulos

FIGURE 1: A fully event-based stereo vision system comprised of a pair of Dynamic Vision Sensors (left) which sends their output to a cluster of TrueNorth processors (right). Credit: IBM

The Brain-Inspired Computing group at IBM Research-Almaden will be presenting at the 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2018) our most recent paper titled "A Low Power, High Throughput, Fully Event-Based Stereo System." The paper describes an end-to-end stereo vision system that uses exclusively spiking neural network computation and can run on neuromorphic hardware with a live streaming spiking input. Inspired by the human vision system, it uses a cluster of IBM TrueNorth chips and a pair of digital retina sensors (also known as Dynamic Vision Sensors, DVS) to extract the depth of rapidly moving objects in a scene. Our system captures scenes in 3-D with low power, low latency and high throughput, which has the potential to advance the design of intelligent systems.

What is stereo vision?

Stereo vision is the perception of depth and 3-D structure. When you look at an object, for example, your eyes produce two disparate images of it because their positions are slightly different. The disparities between the two images are processed in the brain to generate information about the object's location and distance. Our system replicates this ability for computers. The relative positions of an object in images from the two sensors are compared, and the object's location in 3-D space is computed via triangulation of those data.

Stereo vision systems are used in intelligent systems for industrial automation (completing tasks such as bin picking, 3-D object localization, volume and automotive part measurement), autonomous driving, mobile robotics navigation, surveillance, augmented reality, and other purposes.

Neuromorphic technology

Our stereo vision system is unique because it is implemented fully on event-based digital hardware (TrueNorth neurosynaptic processors), using a fully graph-based non von-Neumann computation model, with no frames, arrays, or any other such common data structures. This is the first time that an end-to-end real-time stereo pipeline is implemented fully on event-based hardware connected to a vision sensor. Our work demonstrates how a diverse set of common sub-routines necessary for stereo vison (rectification, multi-scale spatio-temporal stereo correspondence, winner-take-all, and disparity regularization) can be implemented efficiently on a spiking neural network. This architecture uses much less power than conventional systems, which could benefit the design of autonomous mobile systems.

Furthermore, instead of conventional video cameras, which capture a scene as a series of frames, we use a pair of DVS cameras, which

respond only to changes in the scene. This results in less data, lower energy consumption, high speed, low latency, and good dynamic range, all of which are also key to the design of real-time systems.

Both the processors and the sensors mimic human neural activity by representing data as asynchronous events, much like neuron spikes in the brain. Our system builds upon the early influential work of Misha Mahowald in the design of neuromorphic systems. The Brain-Inspired Computing group previously designed an event-based gesture-recognition system using similar technology.

Our end-to-end stereo system connects a pair of DVS event cameras (iniLabs DAVIS240C models) via USB to a laptop, which distributes the computation via ethernet to a cluster of nine TrueNorth processors. Each TrueNorth processor is responsible for the stereo disparity calculations on a subset of the input. In other words, this is a scale-out approach to the computation of stereo, since the system enables, in principle, the addition of many more TrueNorth processors in order to process larger inputs.
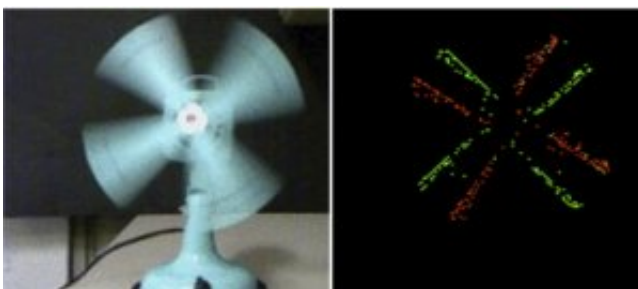

FIGURE 2: Output from a conventional frames-based camera (left) versus Dynamic Vision Sensors (right) for a rotating fan. The Dynamic Vision Sensors produce crisper edges for rapidly moving objects. Credit: IBM Blog Research

The DAVIS cameras provide two 3.5-mm audio jacks, enabling the events produced by the two sensors to be synchronized. This is critical to the system design. The disparity outputs of the TrueNorth chips are then sent back to the laptop, which converts the disparity values to actual 3-D coordinates. An openGL-based visualizer running on the laptop enables the user to visualize the reconstructed scene from any viewpoint. The live-feed version of the system running on nine TrueNorth chips is shown to calculate 400 disparity maps per second with up to 11-ms latency and a ~200X improvement in terms of power per pixel per disparity map compared to the closest state-of-the-art. Furthermore, the ability to increase this up to 2,000 disparities per second (subject to certain trade-offs) is discussed in the paper.
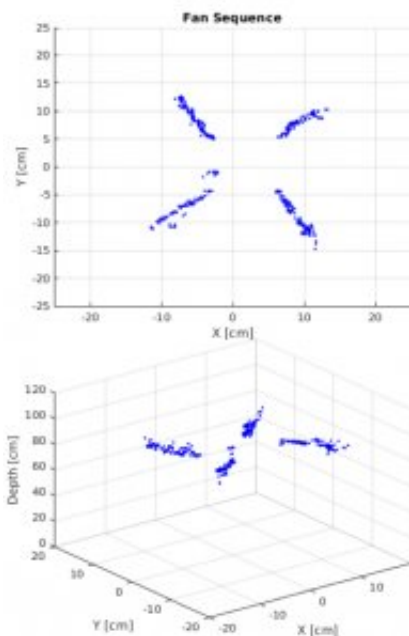

FIGURE 3: Depth reconstruction for a rotating fan from the camera view (top) and from an oblique view (bottom). Credit: IBM

More information: A Low Power, High Throughput, Fully Event-Based Stereo System: [researcher.watson.ibm.com/rese …
aandreo/cvpr2018.pdf](researcher.watson.ibm.com/rese)

Provided by IBM

Citation: Stereo vision using computing architecture inspired by the brain (2018, June 20) retrieved 18 April 2024 from
https://phys.org/news/2018-06-stereo-vision-architecture-brain.html