Accelerating Brain Circuit Simulation for the Real World

Felch, A. C., Chandrashekar, A., Moorkanikara Nageswaran, J., Furlong, J.,
Miller, C., Dutt, N., Nicolau, A., Veidenbaum, A., Granger, R. H.
Neukom Institute at Dartmouth College & University of California, Irvine

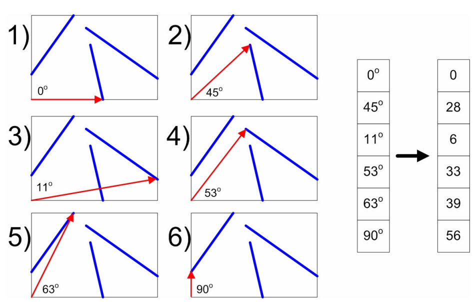
Background and Motivation

Humans outperform computers on many natural tasks including vision. Given the human ability to recognize objects rapidly and almost effortlessly, it is pragmatically sensible to study and attempt to imitate algorithms used by the brain. Analysis of the anatomical structure and physiological operation of brain circuits has led to derivation of novel algorithms that, in initial study, successfully address issues of known difficulty in visual processing. These algorithms are slow on conventional uni-processor based systems, but as might be expected of algorithms designed for highly parallel brain architectures, they are intrinsically parallel and lend themselves to efficient implementation across multiple processors. Adaptation to a cluster of three CELL processors yielded a performance improvement of 140x . Extending the algorithms to new modalities such as speech recognition and language is also under study.

The Building Blocks

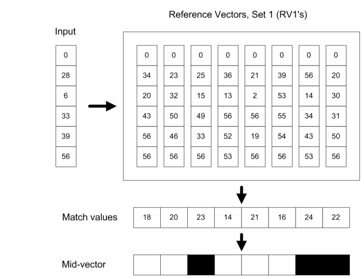
(1) Line segments are extracted and encoded into bit-vectors

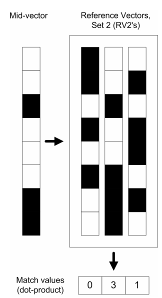
The vision system follows a process of first extracting line segments from an image and then encoding the image features into bit-vectors similar to neuron firing patterns. Local sets of three line segments are used, termed line-triplets. The encoding process creates representations that are translation and scale invariant, by measuring relative angles between line segment endpoints. The process is inherently lossy and redundant similar to the brain, but results in line-triplets having higher bit-vector dot-products with other similar shapes (dissimilar shapes have lower dot-product). Each angle is represented as an 8-hot (contiguous) of 64-bit vector.



Accelerating Comparisons of Shape

(2) Input shapes are compared to a reference set.

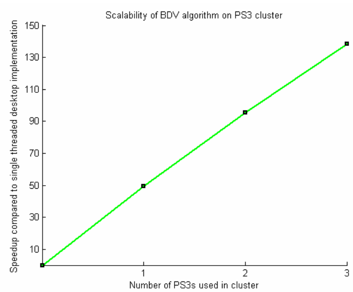




1) Input shapes are compared to a reference set of shapes
-- 8-bit angle representations allow 16 angles to be compared per SIMD instruction.

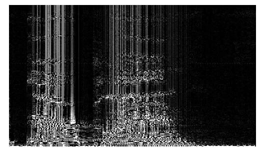
2) The results are converted to a second bit-vector that is compared to a second reference set of shapes
-- 128 bits can be dot-product in 3 instructions (AND, POP_CNT, SUM)

One PlayStation 3 delivers ~50x speedup, the algorithm scales linearly, and three PS3 have achieved 140x.

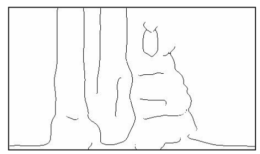


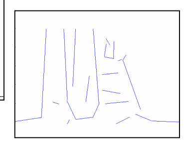
Extensions to Speech and Language

(3) Recognizing Sound Structures

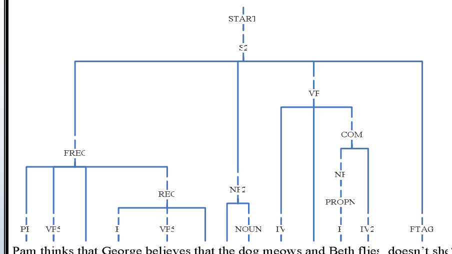


Spoken words can be visualized as spectrograms. Methods for converting sound features into bit-vectors are currently in development. One method is to extract line segments from the audio spectrogram and perform conversion using line-triplets, similar to vision.



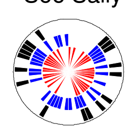


(4) Recognizing Language Structure

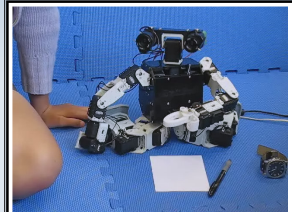


Algorithms derived from brain circuitry have been previously found to perform clustering, classification, and recognition. Combining these lower level networks has yielded algorithms that derive grammar structure from sentences. Sequences of words can be encoded into bit-vectors by assigning dense vectors to each word and performing a SHIFT-AND operation for each word in the sequence.

"See Sally"



Real-time Performance for Interactive Robotics



Just as CELL processors, FPGA's, GPU's, and clusters are each suited to somewhat different tasks, each piece of brain architecture has different functions and different characteristics. As newly developed models of brain architecture continue to extend functionality, integration into real-time robotics can be attempted. An inexpensive robot platform has been developed to allow the study of visual shape hierarchies that comprise the additional pieces of brain architecture responsible for recognizing new objects in previously learned categories. Speech recognition, language semantics, speech generation, real-world navigation, and motor plan execution are also under study. Wireless communication of sensor data and motor movement allows brain architectures to be developed and prototyped on high performance computing systems.

Conclusion and Future Work

By achieving real-time performance, new hypotheses will be tested in real-world environments that will hopefully lead to more robust robots that, through learning, will be able to provide more flexible services without hard-coded programs. A crucially enabling research track will be the development of computer architectures able to support brain-derived algorithms with human-level performance in real-time. Current issues include sparse matrix operations, multicast routing, memory capacity, and memory latency. It may be the case that the characteristics of brain architectures such as high latency communication and low-precision communication will allow novel computer architectures that are highly suited to brain-like computation but are quite poorly suited for general purpose computation.

References

Granger R. (2004) Brain circuit implementation: High-precision computation from low-precision components. In: Toward Replacement Parts for the Brain (T.Berger, D.Glanzman, Eds.) Mass: MIT Press., pp.277-294.