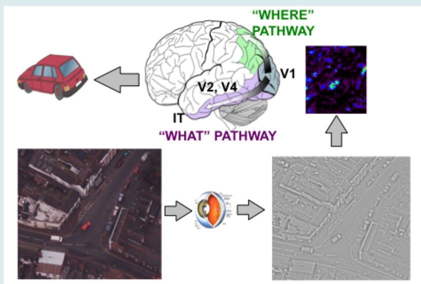
Visual Cortex on a Chip

Large-scale, real-time functional models of mammalian visual cortex on a GPGPU

Steven P. Brumby, Amy E. Galbraith, Michael Ham, Garrett Kenyon, and John S. George
Los Alamos National Laboratory, Los Alamos, New Mexico, USA

Goal

Los Alamos National Laboratory's Petascale Synthetic Visual Cognition project is exploring full-scale, real-time functional models of human visual cortex using the Roadrunner petaflop (1000 teraflop) supercomputer and future GPGPU-based exaflop (1000 petaflop) computers. The project's goal is to understand how human vision achieves its accuracy, robustness and speed. Commercial-off-the-shelf hardware for this modeling is rapidly improving, e.g., a teraflop GPGPU card for a workstation now costs ~$500 and is ~size of mouse cortex. We now present initial results demonstrating whole image classification using standard computer vision image data-sets, and object extraction from UAV video using a model of primary visual cortex running on a GPGPU (240-core NVIDIA GeForce GTX285).
As this technology continues to improve, cortical modeling on GPGPU devices has the potential to revolutionize computer vision.



Functional Models of Visual Cortex. Processing in the human visual system starts in the retina of the eye, continues in the lateral geniculate nucleus (LGN), and then reaches the cortex at region V1 (primary visual cortex), where the input is processed by layers of cortical neurons operating in a massively-parallel assembly of columns of feature-specific cells. Hubel and Wiesel's model of V1 consists of layered "simple" S-cells and "complex" C-cells [1]. Fukushima [2] and Poggio, et al., [3-5] have proposed hierarchical models of the ventral visual pathway ("what" pathway) comprised of visual cortical regions V1, V2, V4, supporting a model of whole object detection in inferotemporal cortex (IT).

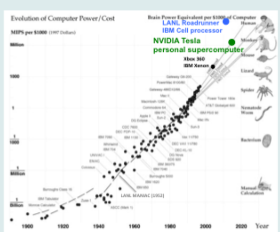


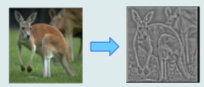
Figure based on Hans Moravec, "When will computer hardware match the human brain?", J. Evolution &Technology, 1998.

Necessary computing hardware is available now. LANL's Roadrunner supercomputer reached a petaflop in 2009 [6], marking the arrival of computing platforms large enough and fast enough for full-scale, real-time functional modeling of human cortex [7]. However, small mammals are capable of excellent visual acuity and object recognition with brains orders of magnitude smaller than humans (11G cortical neurons in human vs 0.3G in cat v 0.004G in mouse). GPGPU technology could enable widespread use of large-scale, real-time models of mammalian visual cortex for a wide range of computer vision tasks.

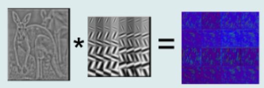
Model

LANL's Petascale Artificial Neural Network (PANN) [8] is a high performance C++, C, and Python implementation of a Neocognitron-type hierarchical model [2-5] of human visual cortex regions V1 (primary visual cortex), V2, V4, and inferotemporal cortex (IT), the ventral pathway of visual processing ("what" pathway). PANN exploits conventional clusters of multi-core CPUs, or hybrid machines such as IBM Cell-accelerated clusters (Roadrunner architecture [6]), or GPGPU-accelerated clusters that are currently in development. Pinto, Cox & DiCarlo [9,10] have recently shown high-throughput image and video processing with V1-like models using a multi-GPGPU machine.

The key scientific question is how does visual cortex organize itself in response to large amounts of visual stimulus? PANN is designed to process large amounts of still and video imagery to match the ~10^15 pixels/year taken in by the 6M cones of the human retina. PANN processes this imagery using unsupervised learning algorithms to build a hierarchical representation of natural scenes, combined with a relatively small amount of supervised learning required to train a back-end classifier (e.g., linear kernel support vector machine (SVM) [11]). The hierarchical representation of natural scenes is believed to be sparse, drawing from a redundant dictionary of scene elements (Olshausen & Field, [12]), which is driving development of a rich set of new ideas about pattern recognition. The speed achievable by running on clusters of GPGPU-accelerated compute nodes will enable testing of the properties of some of these learning rules at the full scale of human visual experience (~months of video).



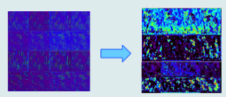
Processing in Retina
Local contrast equalization, for each patch set mean to zero and local Euclidean norm to 1.



Processing in S-cells
Radial Basis Functions with standard oriented Gabor neuron weight vectors [3-5].

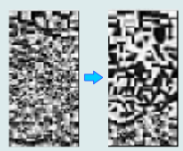
$$s_{RBF}(\mathbf{r},\mathbf{w}_G) = \exp\left(-\frac{1}{2}\beta(\mathbf{r}-\mathbf{w}_G)^2\right)$$

$$\mathbf{w}_G = \mathbf{w}(\theta,\gamma,\sigma,\lambda,\phi) = \exp\left(-\frac{1}{2\sigma}(x_\theta^2+\gamma y_\theta^2)\right)\cos\left(\frac{2\pi x_\theta}{\lambda}+\phi\right)$$



Processing in C-cells
MAX Function over S-cell receptive fields.

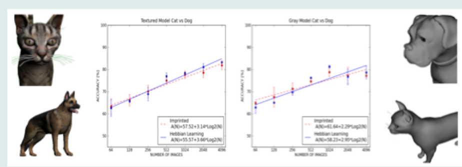
$$c_{MAX}(s) = \max_{i \in I_j}(\{s_i\})$$



Learning
PANN uses Hebbian learning to set the features detected by neurons from large amounts of natural imagery.

Hebb–Oha learning rule [13] :

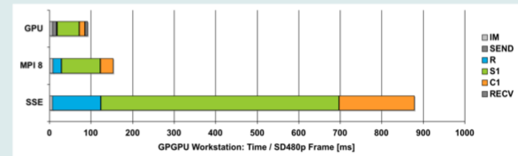
$$\Delta\mathbf{w} = \lambda\, y(\mathbf{r},\mathbf{w})(\mathbf{r}-\mathbf{w})$$



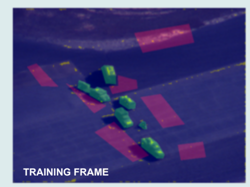
Scaling of image classification accuracy with size of dataset
Neocognitron-type models have been demonstrated to achieve whole-scene classification accuracy comparable to human subjects under conditions of speed-of-sight psychophysics experiments [4]. Here we show results with simulated data sets of rendered models of cats versus dogs (using 3D models with the 3DMax rendering engine) [14]. Model accuracy improves linearly with the logarithm of the size of the training set. Larger training sets require more computing resources to explore scaling properties. Hybrid systems with only one or a few GPGPUs enable large-scale, fast simulations of visual cortex, which has the potential to revolutionize research in this field.

Results

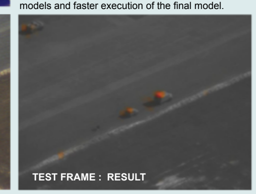
We ported our PANN C++ code to NVIDIA GPGPU using CUDA to develop C++ host code and C device code. The porting process was straight-forward and fast to complete. The PANN code is optimized for ease of algorithm exploration, e.g., we use global device memory to store neuron columns of arbitrary complexity, and the port to GPU was designed to allow ease of validation of host versus GPU code. We report significant speed-up without loss of accuracy on GPGPU, and have been able to develop our code on a range of machines, from simple 16-core GPUs in laptops (NVIDIA GeForce 9600M GT), to 240-core GPUs in workstations (NVIDIA GeForce GTX285), to multi-GPU servers.



GPGPU significantly accelerates visual cortex models
We compare speed of a vectorized PANN model (using SSE intrinsics) running on a 2.6 GHz Intel core (Apple MacPro under OSX 10.5) versus model performance on a NVIDIA GTX285 240-core GPGPU card. We compared outputs at each stage to verify agreement of calculations. We compared average performance over many frames of a 640x480 pixel/frame video stream. As shown above, the retinal component of the model showed **x100** speed-up. The S-cell module, which is the most computationally intensive part of the model, showed **x10** speed-up. The C-cell module showed **x10** speed-up. Host file i/o and round-trip communication time between host and device was a small fraction of frame processing time, and can be parallelized with frame processing. Hence we conclude that GPGPU with CUDA enables significant visual cortex model speed-up without significant constraint on algorithm exploration.



Vehicle Tracking in UAV-like Video
We have preliminary results using a learned model of primary visual cortex on public-release UAV-like imagery of vehicle detection and tracking scenarios provided by the DARPA/DSO Neovision2 program. We are currently working to explore and optimize the object-recognition capabilities of visual cortex models under different conditions of learning rule, initial network topology, and training set size. Initial results suggest that visual cortex models such as PANN are a promising approach to this hard visual pattern recognition problem. GPGPU acceleration may be the key enabling technology for this type of application, as exploiting the GPGPU enables a better than order-of-magnitude speed-up of execution of the model on workstations, enabling learning of better models and faster execution of the final model.



Acknowledgements
Work supported by Department of Energy (LDRD-DR-2009006), National Science Foundation (Award No. NSF-OCI-0749348), and DARPA/DSO Neovision2 program. LANL publication approved for unlimited release [LA-UR-10-05430].

References

[1] Hubel, D.H., Wiesel, T.N.: Receptive fields and functional architecture of monkey striate cortex. J. Physiol. (Lond.) 195, 215–243 (1968).
[2] K. Fukushima: Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position, Biological Cybernetics, 36(4), pp. 193-202 (April 1980).
[3] Riesenhuber, M. & Poggio, T. (1999). Hierarchical Models of Object Recognition in Cortex, Nature Neuroscience 2: 1019-1025.
[4] T. Serre, A. Oliva and T. Poggio. A feedforward architecture accounts for rapid categorization. Proceedings of the National Academy of Science, 104(15), pp. 6424-6429, April 2007.
[5] T. Serre, L. Wolf, S. Bileschi, M. Riesenhuber and T. Poggio. Object recognition with cortex-like mechanisms. In: IEEE Transactions on Pattern Analysis and Machine Intelligence, 29 (3), pp. 411-426 , 2007.
[6] Paul Henning and Andrew White, Trailblazing with Roadrunner, Computing in Science & Engineering, July/August 2009, pp. 91-95.
[7] Hans Moravec, When will computer hardware match the human brain?, J. Evolution &Technology, 1998.
[8] S.P. Brumby, G. Kenyon, W. Landecker, C. Rasmussen, S. Swaminarayan, and L.M.A. Bettencourt, "Large-scale functional models of visual cortex for remote sensing", Proc. 38th IEEE Applied Imagery Pattern Recognition, Vision: Humans, Animals, and Machines, 2009.
[9] Pinto N, Cox DD, and DiCarlo JJ. Why is Real-World Visual Object Recognition Hard? PLoS Comp. Biology, 4(1):e27 (2008).
[10] Pinto N, Cox DD, and DiCarlo JJ. A High-Throughput Screening Approach to Discovering Good Forms of Biologically-Inspired Visual Representation, GTC 2009.
[11] Chang, C.-C. and C.-J. Lin (2001). LIBSVM: a library for support vector machines. http://www.csie.ntu.edu.tw/~cjlin/libsvm
[12] Olshausen, B. A. and Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature, 381:607–609.
[13] Oja, Erkki, Simplified neuron model as a principal component analyzer, Journal of Mathematical Biology 15 (3): 267–273, Nov 1982.
[14] S. Brumby, L. Bettencourt, M. Ham, R. Bennett, and G. Kenyon, "Quantifying the difficulty of object recognition tasks via scaling of accuracy versus training set size", Computational and Systems Neuroscience (COSYNE) 2010.