THE NEUROMORPHIC ENGINEER

VOLUME 2 NUMBER 2 DECEMBER 2005

Neuromorphic robot vision with mixed analog-digital architecture

Robotic vision is the most fascinating and feasible application for neuromorphic engineering, since processing images in real time with low power consumption is the field's most critical requirement. Conventional machine vision systems have been implemented using CMOS (complimentary metal-oxide semiconductor) imagers or CCD (charge-coupled device) cameras that are interfaced to digital processing systems operating with serial algorithms. These systems often consume too much power, are too large, and compute with too high a cost.[1]

Though neuromorphic technology has advantages in these areas, there are some disadvantages too: current implementations have less-programmable architectures, for example, than digital processing technologies. In addition, digital image processing has a long history and highly-developed hardware and software for pattern recognition are readily available. We therefore think it is practical—at least at the current stage of progress in neuromorphic engineering—to combine neuromorphic sensors with conventional digital technology to implement, for robot vision, the computational essence of what the brain does. On this basis, we designed a neuromorphic vision system consisting of analog VLSI (very-large silicon integration) neuromorphic chips and field-programmable gate array (FPGA) circuits.

Figure 1 shows a block diagram of the system, which consists of silicon retinas, 'simple-cell' chips (named after the simple cell in the V1 area of the brain) and FPGA circuits. The silicon retina is implemented with active pixel sensors (conventionally-sampled photo sensors)[2] and has a concentric center-surround Laplacian-Gaussian-like receptive field.[2] Its output image is transferred to the simple-cell chips serially. These chips then aggregate analog pixel outputs from the silicon retina to generate an orientation-selective response similar to the simple-cell response in the primary visual cortex.[3] The architecture mimics the feed-forward model proposed by Hubel and Wiesel,[4] and efficiently computes the two dimensional Gabor-like receptive field using the concentric center-surround receptive field.

The signal transfer from the silicon retina to the simple cell chip is performed using analog voltage, aided by analog memories embedded in each pixel of the simple cell chip. The output image of the simple-cell chip is then converted into a digital signal and fed into the FPGA circuits, where the image is further processed with programmable logic in parallel (not in serial). An example of this real-time computation is shown in Figure 3.

The system emulates the responses of complex cells that are tuned to particular binocular disparities based on the disparity energy model.[5] Here, the emulation is carried out with a binocular platform as shown in Figure 2. In the experiment, a hand was

Shimonomura, *continued on p. 4*

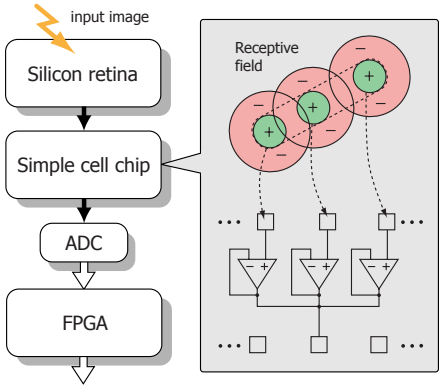


Figure 1. Block diagram of the mixed analog-digital system and analog multi-chip circuit for an orientation-selective response.

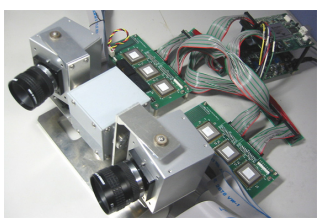


Figure 2. Photograph of the binocular vision system.

In this issue:

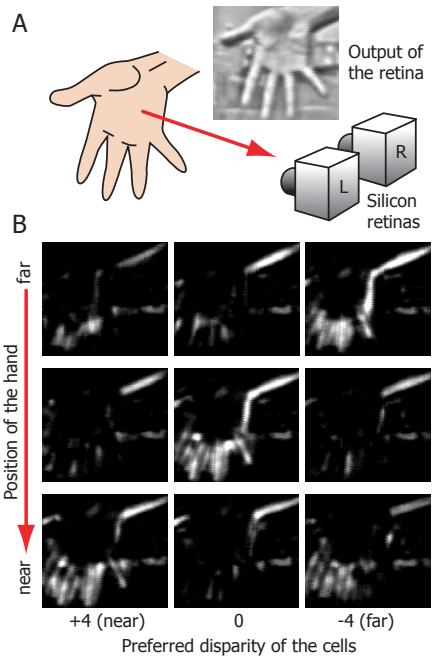


Figure 3. Real time disparity computation with the binocular vision system. (A) Sketch of the experimental setup. (B) Neural images of the complex cells tuned to the disparity of near, fixation and far.

THE NEUROMORPHIC
ENGINEER

is published by the

INSTITUTE OF
NEUROMORPHIC
ENGINEERING

Editor
Sunny Bains
Imperial College London
sunny@sunnybains.com

Editorial Assistant
Stuart Barr
newsletters@sunnybains.com

Layout
Freddy B-Apeagyei
Kut and Payste Studios, London

Editorial Board
Avis Cohen
Ralph Etienne-Cummings
Timothy Horiuchi
Giacomo Indiveri
Shih-Chii Liu
André van Schaik

This material is based upon work supported by the National Science Foundation under Grant No. IBN-0129928. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

The Institute of Neuromorphic Engineering
Institute for Systems Research
AV Williams Bldg.
University of Maryland
College Park, MD 20742

http://www.ine-web.org

Telluride Workshop: Highlights from 2005

This year's three-week summer workshop included background lectures (from leading researchers in biological, computational and engineering sciences), practical tutorials (from state-of-the-art practitioners), hands-on projects (involving established researchers and newcomers/students), and special interest discussion groups (proposed by the workshop participants). Two daily lectures (90 minutes each) covered issues important to the community in general, presenting the established scientific background for the areas and providing some of the most recent results. These lectures spanned most of the diverse disciplines comprising neuromorphic engineering. Furthermore, additional lectures were presented by group work participants in the afternoons on topic germane to the projects.

There were 84 attendees, composed of eight organizers, seven administrative and technical staff members, 27 invited speakers, 10 neurobiologists (a joint workshop organized by Terry Sejnowski) and 32 applicants/students/invitees (as detailed at our web-site[1]). Participants were encouraged to become involved in as many activities as interest and time permitted. They were free to explore any topic, choosing among the 10 workgroups and approximately 30 hands-on projects, and five interest/discussion groups.

Some of the highlights of the work and discussion groups are detailed below.

Telluride Grand Challenge

This first time event *(see more details on next page–Ed.)* was modeled after the DARPA (Defense Advanced Research Program Agency) Grand Challenge. It required groups of participants to equip robots with neuromorphic sensors and sensory processing systems to navigate a cluttered environment. Three quarters of the robot had to cross the finish line. Similar to the first time the DARPA Grand Challenge was held, the Telluride Grand Challenge did not produced a complete winner, although one robot covered 99% of the course before crashing into the finish line! A video showing the highlights of this race is available online.[3]

Bias generators

Participants learned why it is important to build chips that can be manufactured and distributed to users, how generating controlled bias currents that span a very large range is the main necessity for neuromorphic chips, how these circuits work, and how a design kit for Tanner computer-aided-design tools is used to automatically generate layout for a given set of bias currents. All the participants generated their own bias generator layout using the design kit.

Computational neural systems

This workgroup covered a wide range of topics that overlapped with several others. The main emphasis was on address-event-representation multi-chip systems. However, within this theme, the projects covered auditory processing, learning theory, FPGA (field-programmable gate array) and CPLD programming, and neural-network dynamics.

My first AER chip

The goal in this first-time workgroup was to have participants leave with a chip design (finished layout) that incorporates a core circuit from their particular research area into a word-serial AER transmitter (and/or receiver) frame. They were provided the participants with ChipGen, a tool written in the Neuroengineering lab at the University of Pennsylvania. An L-Edit macro, ChipGen compiles the core circuit layout into a fully-verified chip. Two teams designed a conductance-based neuron, while the third pursued a switched-capacitor-based design.

Stochastic synapse transmission in VLSI AER systems

Also presented for the first time, this workgroup discussed the stochastic nature of synaptic transmission, its possible role in information transfer and learning, and (mainly) ways to implement it in AER VLSI (very large silicon integration) systems. It studied possible ways of implementing stochastic synaptic transmission in AER VLSI systems following three main approaches: implementation of stochastic mechanisms via the AER communication infrastructure; implementation of stochastic mechanisms using mixed-mode analog/digital circuits, and on/off chip solutions; and implementation of stochastic mechanisms at the basic synaptic circuit level, using sub threshold MOSFETs (metal-oxide semiconductor field-effect transistors).

Vision chips

This group mainly focused on tutorials of

Etienne-Cummings, *continued on p. 4*

The Telluride 'Grand Challenge' Robot Race

This year's Telluride Neuromorphic Engineering Workshop (June-July, 2005) featured a new event: the *Telluride 'Grand Challenge' Robot Race*. While this was a simple event, our goal was to create a fun, focused activity that would guide the many different participant and workgroup projects towards a concrete demonstration.

Our chosen task was a combination of goal-seeking and obstacle avoidance behavior that challenged robots to negotiate an unknown obstacle course autonomously. We encouraged the integration of many different sensor modalities (e.g., visual, olfactory, auditory, tactile), decision-making models, and locomotion platforms (wheeled, slithering, and walking robots) to provide diversity and encourage innovative solutions. The competition took place in a simple rectangular arena with the start and finish points in opposite corners, and with obstacles (boxes and trashcans) placed at the last minute.

The rules of competition were straightforward:

1. Robot entries must run 'autonomously'. Judges may allow a reset intervention for up to three tries.
2. 'Dumb' beacons of various types may be used only at the end point. Any beacon you create must be considered public (i.e., usable by everyone, if desired).
3. At least 1/3 of the original mass of the robot entering the race must cross the finish line.
4. Robot entries may not damage the racecourse, judges, or bystanders.
5. Water, pyrotechnics, and radiation sources should not play any role in the robot's operation or construction (see also rule 4).
6. Obstacles will be finalized (but not their position) and shown to the competitors at least two days prior to the race.
7. All robots must begin with zero initial kinetic energy.
8. Robots must not move obstacles out of their initial locations.

In support of the competition, the companies K-Team and Wow-Wee Toys generously donated development and robot hardware, both for use in the competition and as prizes. K-Team sent the KoreBot development board, compatible for mounting on Koala mobile robots, but also good for general use. WowWee sent RoboSapiens and RoboRaptor platforms. Two RoboSapiens were used in the competition while we reserved a brand-new RoboRaptor as a prize.

Competitors

By racetime we had 12 entries that ranged from simple locomoting devices with little to no sensing capabilities to fully sensor-equipped robots with goal-sensors and obstacle avoidance behaviors. There were many notable entries:

Garcia, the Antennanator (Noah Cowan, Mitra Hartmann, Jennifer Laine, Vincent Chan, and Kate Williams) was a robot

Garcia, the Antennanator

platform that employed two whisker sensors to detect obstacles and trigger movements that allow extended wall-following behavior. This robot succeeded in progressing about 80% of the way towards the finish line, negotiating two box obstacles before getting stuck against the wall. They did not implement any goal-sensing capabilities.

Kave/Koala robot (Olivier Rochel, Peng Xu, Christian Faubel, André van Schaik) was a binaural cochlea chip that generated AER spikes from two microphones mounted on a Koala robot platform. For stimulation, they used a sonar 'ping' beacon at the finish line, and the spikes were used to generate an estimate of the interaural time difference of the sound. The Koala robot already implements infrared-based obstacle avoidance.

The Kave/Koala Team

RoboSensation (Tobi Delbruck, Hisham Abdalla, Jonathan Tapson, Deborah Gunning, Gong Boonsobhak and Joseph Lin) was a RoboSapien platform that used a wireless network device to send commands and data between the robot and a laptop. The built-in microphones were used to estimate the beacon direction. Both a visual tracking chip and a commercially-available solvent sensor were interfaced to use an odor plume for goal guidance.

RoboSensation

AudioSapiana (Antje Ihlefeld, Mounya Elhilali, Malcolm Slaney, Nima Mesgarani, Tara Hamilton, Jonathan Tapson, Stephen David, Sue Denham, David Anderson, Shihab Shamma, and Andreas Andreou) was constructed from a network of two computers that communicated sound data through the workshop's WiFi-accessible file system. *(Full details in article on next page–Ed.)* Using a binaural auditory saliency model, the two functions of goal-direction finding and recognition of the beacon sound were handled across two computers. This team suffered the ill-fortune of a town-wide power failure

Horiuchi, *continued on p. 9*

Shimonomura, *continued from p. 1*

moving from far to near, crossing the fixation point (see Figure 3A). Figure 3B shows the disparity energy computed by FPGA circuits. Three complex cell layers, each of which is tuned to disparities of *near, fixating point* and *far,* were prepared in parallel. As shown in Figure 3B, the maximum response appears in each corresponding disparity zone as the hand approaches the silicon retina. The neural image of disparity energy model is visualized in real-time.

The simple cells are known to exhibit more-or-less linear receptive fields.[5,6] From this point of view, it is thought to make sense to emulate the simple cell with analog chips without using the spike representation. However, computation in the brain significantly deviates from the linear representation beyond the simple cell. We can therefore use digital technology to compute the nonlinear properties of the complex-cell receptive field, making it compatible with conventional machine-vision techniques. Another possible area to explore is computation using the spike representation. Together, both of these approaches will further robot vision research. Which approach we use for a given application will depend on the type of visual cortex computation we need to implement.

Kazuhiro Shimonomura and Tetsuya Yagi
Department of Electronic Engineering
Osaka University, Japan
E-mail:{kazu, yagi}@ele.eng.osaka-u.ac.jp
http://brain.ele.eng.osaka-u.ac.jp/index_e.html

References
1. G. Indiveri and R. Douglas, *Neuromorphic vision sensors* **Science 288,** pp. 1189-1190, 2000.
2. S. Kameda and T. Yagi, *An analog VLSI chip emulating sustained and transient response channels of the vertebrate retina* **IEEE Trans. on Neural Networks, 14** (5), pp. 1405-1412, 2003.
3. K. Shimonomura and T. Yagi, *A multi-chip aVLSI system emulating orientation selectivity of primary visual cortical cell,* **IEEE Trans. on Neural Networks 16** (4), pp. 972-979, 2005.
4. D. Hubel and T. Wiesel, *Receptive fields, binocular interaction and functional architecture in the cat's visual cortex,* **J. of Physiol. (London) 160,** pp. 106-154, 1962.
5. I. Ohzawa, G. DeAngelis and R. Freeman, *Stereoscopic depth discrimination in the visual cortex: Neurons ideally suited as disparity detectors,* **Science 249,** pp. 1037-1041, 1990.
6. J. A. Movshon, I. D. Thompson and D. J. Tolhurst, *Spatial summation in the receptive fields of simple cells in the cat's striate cortex,* **J. Physiol. (London) 283,** pp. 53-77, 1978.

Etienne-Cummings, *continued from p. 2*

successful vision chips developed by workshop attendees. In particular, Tobi Delbruck talked about his 'physiologist's friend' orientation-detection chip, Timothy Horiuchi talk about his 'horizon detection' vision chip, while André Van Schaik described his 'LogiTek trackball mouse' chip. The latter is one of the few neuromorphic chips that has been commercially-successful.

Distributed neuromorphic sensor nets
Wireless sensor networks have been enabled by development of powerful wireless digital radio components, along with operating systems like TinyOS that support networks of wireless sensors. Low-power neuromorphic chips are very interesting for this application because they can provide significant preprocessing of sensory information without the necessity of active digital processor computation. In addition, the sensor information captured by (for instance) a silicon retina, is preprocessed to greatly reduce redundancy, so that when nodes send wireless information, they have much less processing to do themselves and the data transmission rate is reduced. This year's projects included online learning using TinyOS, giving commercial robots—like RoboSapien—eyes and ears.

Locomotion workgroup
This group continued their work of miniaturizing locomotion controller chips. Jocob Vogelstien, brought a new 24-neuron floating gate (FG) central-pattern-generator (CPG) chip that was tested in one of the workgroup projects. Also, a 10-neuron CPG Chip with digital synapses was used to control a biped, RedBot, with hip and knee muscles. A new version of RedBot was introduced by Tony Lewis, which also had ankle muscles and springy toes. The CPG chip that we had available could not control all the muscles in this new robot, but next year the group plans to control all the muscles using the new FG-CPG chip. Lastly, this group started investigating the muscle synergies, using Vivian Mushahwar's 'IF-THEN' algorithms, to make a RedBot dance.

Auditory work group
The audition group provided the 'enabling technologies' for various projects in the computational neuroscience, distributed sensors and Telluride Grand Challenge (TGC) groups: two of the three most successful robots in the TGC used auditory cues to find the finish line.

Sensors and robotics work group
This year saw increased activities in tactile sensing, particular due to the collaboration of a newcomer to the neuromorphic community—Noah Cowan—with Mitra Hartmann. The projects in this work group included: tactile sensing for robotic locomotion; cockroach antennae and rat whiskers; Garcia (the orange touchy-feely robot); whisker sensors for obstacle avoidance; and dynamical systems approach to robotic navigation. Two projects lead to Telluride Grand Challenge entries, and they were quite successful, winning one of the top prizes!

Conclusion
The lectures were again of the highest quality, with some of the top neuroscientists, computer scientists and engineers occupying the same room and discussing some of the most timely research questions. The participants were particularly engaged this year, showing the depth and breath of their knowledge. We had a number of lectures that had to be spread over four hours because of discussion and questions. A list of the speakers and talk titles is posted online.[3] We are already planning next year's Workshop and hope to continue the great tradition of the last 10 years.

Ralph Etienne-Cummings
Johns Hopkins University
Baltimore, MD
E-mail: retienne@jhu.edu

References
1. http://www.ini.unizh.ch/telluride
2. http://www.theredplanet.co.uk/Telluride2005
3. http://ine.ini.unizh.ch/telluride/current/index.html

AudioSapiana: an auditory robot that uses salience and binaural hearing

AudioSapiana is a listening and walking robot designed by the audio group at the 2005 Neuromorphic Workshop. We added ears to a RoboSapien robot from WowWee Entertainment to help it navigate in a difficult, obstacle-strewn environment. Previously, robots that oriented themselves towards a target beacon were designed to work in quiet environments. Audiosapiana, however, was designed so she could navigate in a noisy, multi-source acoustic environment.

The underlying algorithm was based on a model of human auditory perception: specifically, how listeners identify and attend to novel sounds. We combined a simple model of auditory saliency with a model of sound localization that allowed AudioSapiana to attend to novel sounds, turn towards her mating call, and to navigate around obstacles. Our approach is novel in that information from both monaural and binaural pathways is integrated in a psychophysically-plausible way. This is an important step towards a model of how listeners understand speech in a cocktail-party environment.[2]

Motivation

The processing architecture and time-scale for perceptual integration of monaural and binaural properties into primitive groups are still open research questions. One possible strategy that the brain may use is to analyze the location of each frequency channel separately and attend to the frequency components that have common interaural cues. Alternatively, the brain may initially group the auditory scene into different objects based on harmonicity, onset cues etc., and then determine the location of each object.

Traditionally, speech recognition algorithms that use both monaural and binaural information filter the signal first through a binaural pathway followed by a monaural recognizer. However, some psychophysical data suggest that monaural stream segregation precedes the binaural processing part when the signals are very short.[1] This idea was implemented in our computational model.

Computational model for segregation using monaural and binaural integration

Upon entering the ears, the ongoing sound signal is segregated into auditory objects by monaural pathways (see also Figure 2). Simultaneously, the auditory scene is analyzed by a binaural processor that groups the scene into frequency patches of coherent location and similar onset. The monaural processor sends the estimated target frequency channels to the binaural processor. The latter determines the corresponding target locations, detecting spatial mismatches within the proposed frequency channels, and feeds these back into the monaural processor along with all non-target frequencies that originate from the estimated target location.

Figure 1. AudioSapiana on race day, ready to search for her mate.

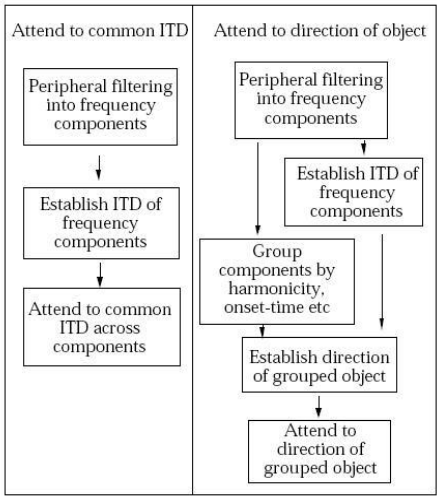


Figure 2. Two models of monaural and binaural processing. AudioSapiana implemented the model on the right.

Only the first 10msec after the detected onset of the target are used for location estimation. This is advantageous in two ways. Firstly, it limits the processing time. Secondly, it avoids reverberation. Human listeners use a similar processing strategy. When localizing sounds in reverberant environments, more weight is typically given to transient than steady-state cues (the precedence effect).

The binaural processor also has a decay-constant that smoothes the estimated target location across time. Perceptually this corresponds to the binaural sluggishness effect.

In summary, the monaural and binaural pathways work together and enhance the estimated target frequency channels, allowing more processing power of the central processor to be used for detailed analysis of the target signal.

Practical implementation

AudioSapiana was designed to navigate towards her mating call: an acoustic beacon provided by Andreas Andreou. Two small microphone ears were mounted on the robot's feet. This unusual placement permitted detection of acoustic shadows behind short obstacles as well as tall ones.

The neuromorphic algorithms were implemented in MATLAB. The mating call detection was accomplished using Gaussian mixture models (GMMs) trained on some of the calls. Time-frequency bins with significant acoustic-call energy were passed to the binaural system.

Audiosapiana navigated through the environment based on interaural level difference (ILD) cues based on a simple decision maker: for ILDs with a magnitude of 1dB or larger, the robot turned by 45° towards the louder ear. For smaller ILDs, the robot continued to walk straight. ILDs were averaged across the five estimated target frequency bands with the highest signal-to-noise ratio, and these bands were weighted by the average signal energy in each channel.

We chose ILD cues over interaural phase differences (IPD) for four reasons. Firstly, ILDs are easier to calculate. Secondly, the robot noise was mostly in the low-frequency range, therefore masking the IPDs. Thirdly, the different phase delays

Ihlefeld, *continued on p. 9*

High-speed and high-precision current winner-take-all circuit

A winner-take-all (WTA) circuit, which identifies the highest signal intensity among multiple inputs, is one of the most important building blocks in neural networks, non-linear filters, and various fuzzy and neuromorphic systems. Many implementations have been proposed in the literature,[1-5] since the first WTA circuit was introduced by Lazzaro *et al.*.[1] The performances of all these can be measured in terms of speed, precision, and power consumption. While power consumption is very important in some applications, there are others where high speed and high precision are more important.

Here we describe a novel, high-speed, and high-precision current-mode WTA circuit. The circuit employs inhibitory and local excitatory feedback based on input currents average computation, and this enhances both the precision and speed of the circuit. Local excitatory feedback provides a hysteretic mechanism that prevents the selection of other potential winners unless they are stronger than that selected by a set hysteretic current. This circuit can be useful for integration with others operating in the strong inversion region and supplying input currents of 3μA-50μA, as well as for sub-threshold applications with inputs of below 50nA.

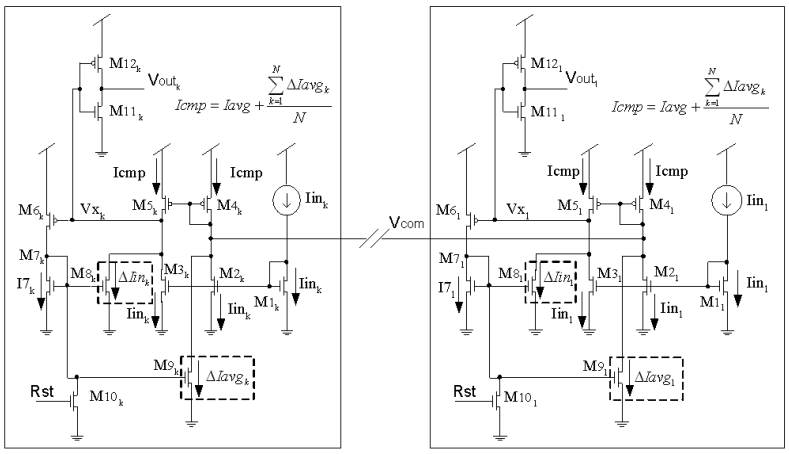
Unlike previously-presented WTA circuits, ours achieves very high speed: 32ns for high measured currents of 3μA-50μA, and 34ns for simulated sub-threshold currents. This applies when a very small difference between two input currents is applied (30nA for high currents and 1.8nA for sub-threshold applications). Such circuit performance is the direct result of very strong feedback. However, this leads to higher power dissipation than existing low-performance WTA circuits that operate in the sub-threshold regime and optimize power consumption:[1,4] 87.5μW for high input currents and 22.5μW for sub-threshold currents. The power consumption of our circuit can be significantly reduced by decreasing feedback values, but this, in turn, degrades circuit performance.

Figure 1 shows cells 1 and *k* (of N interacting cells) of the WTA circuit. Cell *k* receives a unidirectional input current, Iin_k, and produces an output voltage $Vout_k$. This output has a high digital value if the input current I_k is identified as winner, and is low otherwise.

The WTA circuit operates as follows: the drains of the M2 transistors of all N cells of the array are connected to the drains of the M4 transistors by a single common wire with voltage Vcom. The circuit starts competition by applying Rst = '1' for a short period of time. This way, the excitatory feedback ΔIin_k and inhibitory feedback $\Delta Iavg_k$ are cancelled. Assuming that all N cells in the array are identical and Rst = '1' is applied, the current Icmp, through $M4_k$, is equal to the average of all the input currents of the array, neglecting small deviations in the referenced input currents. Icmp is copied to $M5_k$ by the PMOS (positive-channel metal-oxide semiconductor) current mirror ($M4_k$ and $M5_k$) and is compared with the input current Iin_k copied by the NMOS (negative-channel MOS) current mirror ($M1_k$ and $M3_k$). An increase in input current Iin_k relative to Icm_p causes a decrease in Vx_k value due to the Early effect. This way, during the reset phase, input currents of all cells are compared to the average of all input currents of the array, producing a unique output Vx_k for every cell. The cell having the highest input current value produces the smallest Vx_k voltage.

With the completion of the reset phase, i.e. Rst = '0', the excitatory feedback ΔIin_k and the inhibitory feedback $\Delta Iavg_k$ are produced. The Vx_k node inputs to the gate of $M6_k$ PMOS transistor, thus the cell with smaller Vx_k (highest input current) produces higher current $I7_k$ through $M6_k$ and $M7_k$. This current is copied by the NMOS current mirror ($M7_k$ and $M8_k$), creating the excitatory feedback ΔIin_k. On the other hand, $I7_k$ is copied by the NMOS current mirror ($M7_k$ and $M9_k$), resulting in inhibitory feedback $\Delta Iavg_k$. The ΔIin_k is added to the Iin_k flowing through $M3_k$ and $\Delta Iavg_k$ is added to the average of all input current by connection $M9_k$ transistor to COM node, increasing the Icmp value. This way, every cell produces a new Vx_k voltage value, according to the comparison between current $Iin_k + \Delta Iin_k$ and a new value of Icmp current. For the cell with the highest input current, the difference between $Iin_k + \Delta Iin_k$ and Icmp grows: this decreases the Vx_k value. At the same time, cells, with small input currents have their Vx_k increased. The computation phase is finished when only one cell is identified as a winner, producing $Vout_k$ = '1' at the inverter output. All other cells are identified as losers with $Vout_k$ = '0'.

Figure 1: Cells 1 and k (out of n) of the proposed WTA circuit.



A. Fish,* Vadim Milirud,† and Orly Yadid-Pecht*†
*VLSI Systems Center, Israel
†Dept of Electrical and Computer Engineering
University of Calgary, Alberta, Canada
E-mail: afish@ee.bgu.ac.il,
vmilirud@ucalgary.ca, oyp@ee.bgu.ac.il

References

1. J. Lazzaro, S. Ryckebusch, M. A. Mahowald, and C. A. Mead, **Winner-take-all networks of O(n) complexity,** Technical Report, California Institute of Technology, San Mateo, CA, 1988.
2. A. G. Andreou, K. A. Boahen, A. Pavasovic, P. O. Pouliquen, R. E. Jenkins, and K. Strohbehn, *Current-mode subthreshold MOS circuits for analog VLSI neural systems,* **IEEE Trans. on Neural Networks 2** (2), pp. 205-213, March 1991.
3. S. P. DeWeerth and T. G. Morris, *CMOS current mode winner-take-all circuit with distributed hysteresis,* **Electronics Lett. 31** (13), June 1995
4. G. Indiveri, *A Current-Mode Hysteretic Winner-take-all Network, with Excitatory and Inhibitory Coupling,* **Analog Integrated Circuits and Signal Processing 28,** pp. 279-291, 2001.
5. T. Serrano and B. Linares-Barranco, *A modular current-mode high-precision winner-take-all circuit,* **IEEE Trans. on Circuits and Systems II 42** (2), pp. 132-134, February 1995.

Biologically-inspired image processing for a robotic grasping task

Visual processing in mammals is adapted to their behavioral needs: likewise, in visually-guided robots, image processing needs to be suitable for a desired behavior. Thus, the function of the mammal brain may be a good guideline for choosing the right image-processing techniques for machines. In our work, we make robots learn through experience and thereby study which learning and image-processing techniques lead to a good performance for a given task.

Here, we describe a study in which our goal was to make a robot arm grasp an object presented visually.[1] The robot learned to associate the image of an object with an arm posture suitable for grasping. Learning an association means that there are no world coordinates and there is no tedious calibration of the vision system, instead, the robot learns by randomly exploring different arm postures and by observing the appearance of objects put on a table. Though the emphasis of our work is on learning techniques, here we will focus on the image processing.

We used a robot arm with six joints and a gripper: the vision system was a stereo camera head mounted on a pan-tilt unit (see Figure 1). This setup was located behind a table, which was the operational space and which was visible to the cameras. In training, the robot placed a red brick on the table in random positions and, for each position, recorded an image of the scene after removing the arm. Thus, the training set contains corresponding pairs of grasping postures and object images.

An image can be interpreted as a point in a high-dimensional space (with the number of dimensions equal to the number of pixels). A mapping from such a space to an arm posture suffers from the so-called 'curse of dimensionality': the distance between pair-wise different images is almost constant, and the orientation of the target gets lost under the dominance of the positional information.[2] Therefore, the image must be pre-processed.

The processing technique that was eventually successful was inspired by the function of the visual cortex. The image processing was split into two parts: one for the object's location and one for its orientation (see Figure 1). To decode the location, the image was first blurred and sub-sampled. Since here the target (the brick) was almost point-like within the camera image, the blurred image is like a population code of the brick's position. In a population code, many neurons encode a parameter: such a code for the retinal location of a stimulus exists also in the primary visual cortex.[3]

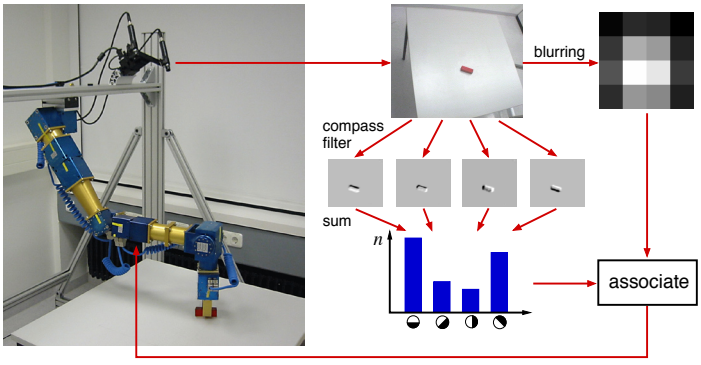


Figure 1. Shown is the information flow in the grasping task. The processing of the camera image is split into two parts. First, to extract position information, the image is blurred and sub-sampled. Second, to extract orientation information, four different compass filters (directional edge filters) extract edges in different directions. The sum of the white pixels in each of the four filtered images results in a histogram of edge distribution. This histogram, together with the blurred image, is associated with an arm posture that enables the robot to grasp the observed object.

To decode the orientation, image filters were used to extract edges in different directions: for each, we counted the edge pixels within the image. This sum was invariant of the brick's position and was a measure of how close the brick was to a given orientation. Position invariance and orientation tuning are also properties of V1 complex cells.[4]

The resulting visual information could be used to first learn and then to recall the association with an appropriate arm posture for grasping (Figure 2). Specifically, the decomposition of the image processing into two parts and the use of population codes kept the grasping errors low.[1,2] This robot experiment demonstrated that brain functions can provide guidelines for robotic control, but also robots can help us to understand the brain. This is done by first demonstrating that certain (often hypothetical) functions actually work and then showing the advantages of certain data-processing techniques in a behavioral context.

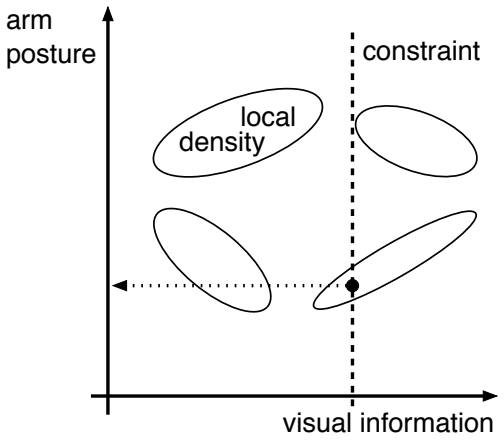


Figure 2. Pattern association. Training patterns lie in the product space of arm posture and visual information. The density of the pattern's distribution is modeled by a mixture of Gaussian functions (ellipses are iso-density curves). To map the visual information onto an arm posture, we define the output space as a constrained space anchored at the input. On this subspace, the highest local density gives the desired output.

Heiko Hoffmann
Max Planck Institute for Human Cognitive and Brain Sciences
Munich, Germany
E-mail: mpi@heikohoffmann.de
http://www.heikohoffmann.de/publications.html

References

1. H. Hoffmann, W. Schenck and R. Möller, *Learning visuomotor transformations for gaze-control and grasping*, **Biological Cybernetics 93,** pp. 119-130, 2005.
2. H. Hoffmann, **Unsupervised Learning of Visuomotor Associations,** Vol. 11 of MPI Series in Biological Cybernetics, Logos Verlag, Berlin, 2005.
3. D. Jancke, W. Erlhagen, H. R. Dinse, A. C. Akhavan, M. Giese, A. Steinhage and G. Schöner, *Parametric population representation of retinal location: Neuronal interaction dynamics in cat primary visual cortex*, **J. of Neuroscience 19** (20), pp. 9016-9028, 1999.
4. T. W. Kjaer, T. J. Gawne, J. A. Hertz, and B. J. Richmond, *Insensitivity of V1 complex cell responses to small shifts in the retinal image of complex patterns*, **J. of Neurophysiology 78,** pp. 3187-3197, 1997.

Capture and display in wide dynamic-range imaging

Introduction

It is commonly accepted[1] that the dynamic range (DR) of an image, i.e. the image color definition, is equal to the image maximum signal-to-noise ratio (max SNR). We also know that the maximum amount of image information is given by the product of resolution (cross-image definition) and color definition. Therefore, increasing the DR of images is by no mean less important than increasing their resolution. It is worth remembering however that DR and resolution cannot possibly be traded-off against each other, since they live in orthogonal sub-spaces of the image. Below, we analyze the two most commonly-used methods to increase dynamic range.

Purely logarithmic photo-response

Working with a purely log-response practically eliminates saturation altogether, thereby making the maximum signal unlimited. It can be shown however[2] that to first order, the DR at the output of any 1-1 analytic mapping equals the DR at its input.

To realize this in the log case we recall that the small-signal gain of the log function is inversely proportional to the average input signal. This implies that, at this sensor's output, a local contrast whose magnitude is usually small relative to the average luminance, maintains a pretty-much constant SNR for all levels of average luminance. At the same time, however, such a local contrast becomes progressively 'washed-out' as the average luminance increases: a familiar and most undesirable property of the log transmission.

The multiple-exposure method

Multi-exposure (ME) means to apply shorter exposure times to originally saturated pixels, such that saturation is altogether avoided. This is generically illustrated in Figure 1, where the exposure periods are determined according to the simplest normalized series of

$$1,\ \frac{1}{2},\ \frac{1}{3},\ \cdots\ \frac{1}{k},\ \cdots,$$

where k is a natural number. Note that the overall sensor's gain to light is proportional to the exposure period. Therefore, to avoid contrast reversals, the gain of segment 2 must be elevated by a factor of 2, and so on.

The RMS of the camera shot noise, however, is proportional to the square root of the exposure time. Therefore the RMS of the camera output noise in the continuous transmission arrangement above is elevated by a factor of √2 in segment 2, a factor √3 in segment 3, and so on. The DR of this arrangement is therefore given by

$$DR_{ME} = DR(1+\frac{1}{\sqrt{2}}+\frac{1}{\sqrt{3}}+\cdots+\frac{1}{\sqrt{k}})$$

and the added DR in bits is

$$\log_2(\sum_{i=1}^{k} i^{-1/2})$$

One can see that there is not much point in having the shortest exposure period less than a 1/3 of the default, since e.g., in 1/4-default the output noise is twice that of the default, thereby amplifying the minimum detectable local contrast by a factor of 2 and causing an **already-significant 'noise-washout' effect.** This limits the added DR of the ME arrangement, in practice, to just a single bit.

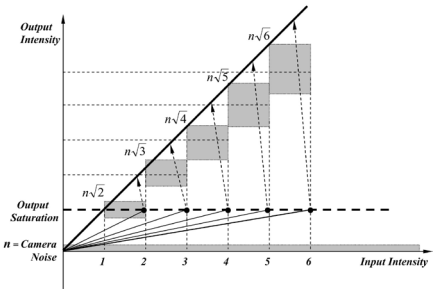


Figure 1. Using multiple exposures to increase dynamic range.

The role of display DR

As it turns out,[3] in normal indoor lighting we cannot see more than 5 to 6 bits per color (bpc) with most cathode-ray-tube, liquid-crystal, and plasma displays. The effect of this bottleneck is demonstrated by the upper half of Figure 2, an 8bpc RGB picture *Room* (courtesy of Vincent Laforet, Photographer in Residence at the New-York Times) where the maximum amount of lost visual content due to the display is about nine bits per pixel (24 minus 15).

To get around the display bottleneck, we consider a similar situation that takes place inside our retina.[4] The neural channels that carry all sensory information to the cortex cannot possibly support more than 7 bits of DR. Nevertheless, the DR we actually perceive can easily exceed 30bpc. The mechanism that makes this possible has been analytically modeled[5] and is called retinal dynamic-range compression (DRC).

Figure 2. The 8bpc version of Room *before (upper half) and after (lower half) retinal dynamic-range compression to 5bpc.*

Reconstruction of retinal DRC

The basis of retinal DRC is the spatial feedback automatic-gain-control (fb-AGC), described in Figure 3. Here, the acquired image E_i is multiplied by the scalar forward gain K and fed into the first input of an image pixel-wise multiplier. The output is then fed back into the multiplier's second input after having passed through a linear spatial low pass filter (LPF, to average) and being subtracted from unity.

The average transmission (DC) of this model[5] is

$$E_o = \frac{E_i}{1/K + E_i}$$

which is known as Michaelis' equation. Its graph (see Figure 4) is known as Weber's law.

The DR compression ratio (CR) is defined as the output/input ratio for $E_i = 1/K$ (the knee-point). We then have: $CR = K/2$. The compression ratio is thus readily controlled through the parameter K. The fb-AGC gain for variations of low spatial frequencies (the 'DC-gain') is given by

$$G_{DC} \equiv \left.\frac{dE_o}{dE_i}\right|_{DC} = \frac{d}{dE_i}\left(\frac{E_i}{1/K+E_i}\right) = \frac{1/K}{(1/K+E_i)^2}$$

The fb-AGC gain for local contrast variations (the 'AC-gain') is obtained by assuming

Shefer, *continued on p. 11*

Horiuchi, *continued from p. 3*

that put the computer network in a funny state. After getting the system functional again, the robot headed out, but did not operate properly. (There is video of the robot elegantly completing the course soon after the competition was over. See: www.theredplanet.co.uk/Telluride2005)

RoboCup (Hisham Abdalla) was a hand-built robot that detected ultrasonic signals transmitted by two beacons in the room to estimate x-y position in the room. This allowed the system to virtually 'sense' the direction of the finish line. A prominent feature of this robot was the plastic cup used to hold up one of its ultrasonic transducers.

It (Chris Twigg and David Graham) was a small toy car on which two binary whiskers were installed. This car achieved a high speed and would turn when one of the whiskers ran into an obstacle. When both whiskers are deflected, a reverse command was issued after some delay. The car managed to careen off of two obstacles and strike a wall extremely close to the finish line.

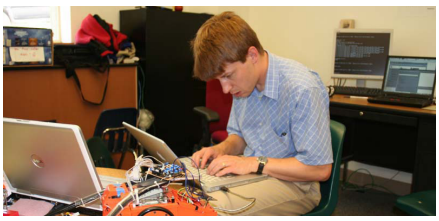
Kave/Koala.

And the winners were…

The two main prizes, provided by K-Team and WowWee, were won by *Garcia, the Antennanator,* for the 'Coolest Approach Attempted' award, and by *Kave/Koala,* for the 'Most Successful Robot' award. Additional awards were given for the following categories: 'Fastest Entry' was It; 'The Rube Goldberg Contraption' was *RoboCup;* 'Best Looking Robot' was *AudioSapianna;* 'Most Pathetic Performance on the Field of Battle' was *RoboSensation;* and the 'Most Integrated Robot' was *RoboSapianna.*

Closing thoughts

Overall, we feel that the robot race was a great success in drawing the interest and efforts of many of the participants. This year we saw the emergence of the RoboSapien as a new robot platform and the effective use of wireless network devices in various systems. We also received lots of feedback that perhaps too much of everyone's time was spent in making the most basic hardware operational, leaving less time for algorithm development and systems-integration issues. In future competitions, we intend to provide more infrastructure and ready-to-use robotic and development platforms to encourage the use of common locomotion platforms and communications. We encourage all of you to contact us about your thoughts on the competition and ways to make this an even better event for the future.



Noah J. Cowan.

Timmer Horiuchi and Giacomo Indiveri*
Department of Electrical Engineering
Institute for Systems Research
Neurosci. and Cognitive Sci. Program
University of Maryland
College Park, MD
*Institute of Neuroinformatics
Uni/ETH Zurich
Zurich, Switzerland
E-mail: timmer@isr.umd.edu

Ihlefeld, *continued from p. 5*

in the microphones caused an IPD bias that we could not remove with our limited recording equipment. Lastly, for navigation of AudioSapiana through the maze, ILD cues are advantageous because they contain information about the acoustic shadows from the obstacles.

In the real-time implementation, AudioSapiana listened for her mating call, made a binaural decision, and then moved a few steps towards the left, right or straight. While moving, AudioSapiana listened again, and repeated the process. AudioSapiana did not have any route-planning or other artificial intelligence. Instead, when she bounced into an obstacle, she backed up a few steps (a pre-programmed feature of RoboSapien) and listened again for a new direction. This strategy proved successful, enabling AudioSapiana to consistently finding a path to her mate.

In the future, the audio team hopes to investigate more realistic implementations of the auditory-saliency binaural model. We intend to close the feedback loops, and study how this model fits human perception.

The people listed below all contributed to AudioSapiana's success.

- *Dave Anderson (Gatech): Noise suppression model.*
- *Jay Kadis (Stanford): Microphones.*
- *Jonathan Tapson (Cape Town): Hardware hacking.*
- *Mark Tilden (WowWee): Father of our girl.*
- *Mounya Elhilali (Maryland): Crew chief and audio wrangler.*
- *Nima Mesgarani (Maryland): Wireless audio.*
- *Sue Denham (Plymouth)—Salient object detector*
- *Shihab Shamma (Maryland): Random (mostly great :-) ideas.*
- *Steven David (Maryland): Control and hardware.*
- *Tara Hamilton (Sydney): Wireless and hardware.*
- *Toby Delbruck (ETH): Wireless and control.*
- *Malcolm Slaney (Yahoo!): Primitive auditory front end.*
- *Antje Ihlefeld (Boston Univ.): Binaural Model.*

Antje Ihlefeld and Malcolm Slaney*
Department of Cognitive and Neural Systems
Boston University, Boston, MA
*Yahoo! Research, Sunnyvale, CA
E-mail: antje1@gmail.com, malcolm@ieee.org

References

1. C. J. Darwin and R. W. Hukin, *Auditory objects of attention: the role of interaural time-differences,* **J. Exp. Psychol. 25,** pp. 617-629, 1999.
2. A. Bregman, **Auditory Scene Analysis: The Perceptual Organization of Sound,** The MIT Press, 1990.

LABORATORY NOTES

The USB revolution

We have all heard the quote, "Revolutions do not happen overnight". Indeed, the USB revolution was years in the making: but soon after the USB (universal serial bus) board fired its first spike, nearly every member of our lab had abandoned their chip-testing setups to join the insurgency. This custom tool had ousted logic analyzers, pattern generators, digital I/O cards, and other bourgeois equipment. My comrades, we wish to spread the revolution.

We have two aims in this article: first, we want to demonstrate that, with the right components and driver, USB 2.0 offers a high-performance (7 million spikes/sec), portable (compatible with any PC), inexpensive (under $500 in parts and printed circuit board), and flexible (coded in C++) computer interface. Second, we want to describe how this interface has revolutionized the way in which we interact with our neuromorphic systems. Our obvious enthusiasm stems from the sentiment that our chips are no longer limited by the surrounding technology, but instead by our own creativity.

Hardware

To achieve the interactivity that we envision, the USB link must seamlessly integrate with our neuromorphic chips. A typical neuromorphic chip with 10,000 neurons[1] can demand bandwidths exceeding 500,000 address-events[2]/s (considering both the transmitter and receiver), and proposed multi-chip systems[3] can increase this demand by ten-fold. In theory, USB2.0 can support 480Mb/s: to blast address-events at these rates, we must enlist the appropriate chip set.

We chose the USB Cypress FX2 transceiver chip because its configurable datapath allows us to meet our performance requirement of streaming millions of address events per second. Specifically, we can bypass the FX2's on-chip microcontroller (through a firmware assembly configuration file), which would otherwise introduce a devastating bottleneck for real-time data transfers. We use a pair of 4×256×16 bit (quadruple buffered) FIFOs (first in, first outs) on the FX2 for streaming address events to and from our chip (one for each direction).

The transfers are asynchronous and we match the address-event representation protocol to the FX2's protocol using a CPLD (complex programmable logic device from *Lattice Semiconductor*). And indeed, we realize a high-performance bidirectional link, achieving 7 million 16bit events per second, or 112Mb/s (see Figure 1).

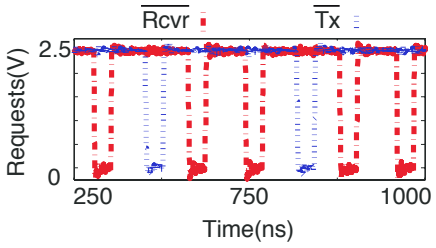


Figure 1. USB 2.0 communicates between a PC and a neuromorphic system at 7 million 16bit events per second (duplex).

Software

With our high-performance USB2.0 link, we now have the ability to stream spikes (and other data) in real-time between a neuromorphic system and a software application (GUI, graphical user interface): the challenge is ensuring that the software can keep up. Since coding drivers is anathema to us hardware guys and gals, we decided to use the commercial USB driver from Thesycon. Naturally, Thesycon is not the only game in town, but by following their Microsoft Visual C++ examples we were able to integrate real-time events into our software apps in no time. Essentially what Thesycon provides is the ability to probe the status of the FX2's FIFOs (both input and output). For instance, when an input FIFO buffer is full, the driver copies it to memory. At low event rates, this takes a while, so the CPLD injects timing events every 50μs. For the majority of the time, however, our GUI is free to process, plot, and save spike data.

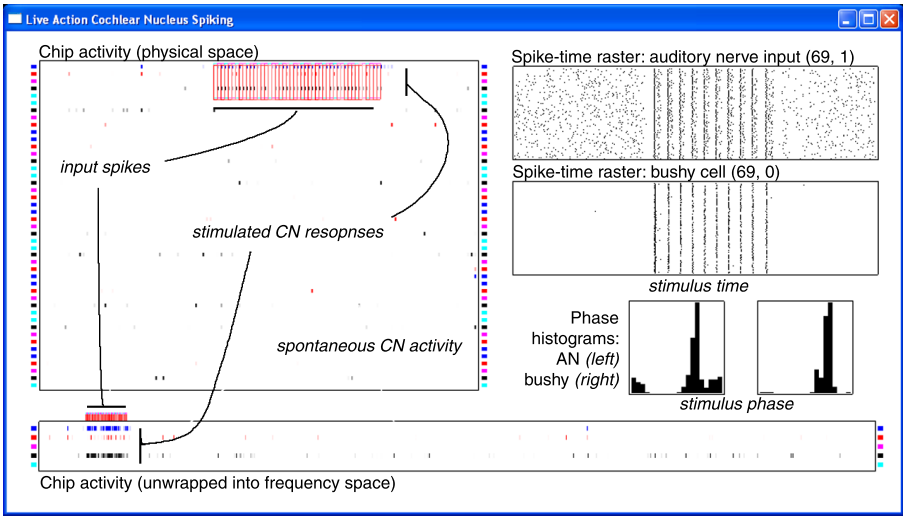
The GUI enables us to visualize chip activity, processing and plotting incoming spikes in real-time. Visualization, however, has the potential to hold up the link if FIFOs fill up faster than the screen can update. For this reason, we use openGL (the open graphics language), which plots extremely quickly.[4] Of course, some of the more advanced plotting features (e.g., in three dimensions) are still too slow: it is not a general panacea. In practice, openGL suffices for the two dimensional plots we use.

¡Viva la revolución!

The USB2.0 interface has become an essential part of testing, visualizing, and interacting with the neuromorphic systems in our lab. We briefly highlight some of the systems we have visualized using this link in Table 1. Note that, for each project, the GUI was specifically tailored to monitor the essential computation in that system. For instance, if we are interested in the phase-locking

Merolla, *continued on p. 11*

Figure 2. The cochlear nucleus chip responds to sound encoded by auditory nerve spikes (inputs spikes). Chip activity *displays neurons that fired recently according to chip layout (physical space) and cochlear mapping (frequency space).* Spike rasters *display 100 trials of a neuron's spike times during a 50ms stimulus, while* phase histograms *quantify spike-time precision. The selected bushy cell (address 69,0) phase-locks better than its inputs: its phase-histogram is tighter.*



Shefer, *continued from p. 8*

that, for such variations, the LPF output remains constant. Denoting this constant by $\overline{E}_o$ we have

$$G_{AC} \equiv \left.\frac{dE_o}{dE_i}\right|_{AC} = K(1-\overline{E}_o) = K(1-\frac{E_i}{1/K+E_i}) = \frac{1/K}{1/K+E_i}$$

from which we get

$$G_{AC}/G_{DC} = 1 + KE_i$$

Taking the quotient (G_{AC}/G_{DC}) as a measure of the detail enhancement of the retinal DRC, we see that the amount of this detail enhancement, or the 'effective visual acuity' increases linearly with the average luminance of the viewed scene: a well known property of human and animal vision. The lower half of Figure 2 is *Room* after retinal DRC to 5bpc.

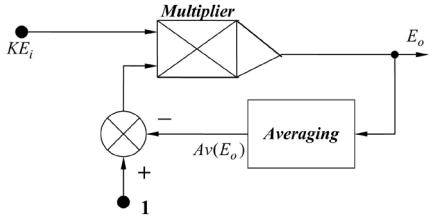


Figure 3. The spatial feedback automatic-gain-control (fb-AGC) model.

Conclusions

On the capture side of wide-dynamic-range imaging, we have demonstrated that dealing with saturation alone cannot significantly increase the DR of image sensors. We therefore conclude that this can only be achieved via significant noise reduction. On the display side of the problem, Figure 2 above demonstrates how retinal DRC can increase the potential number of perceived distinguished colors—by a factor of 512 in the current example—and shows how this affects our watching experience.

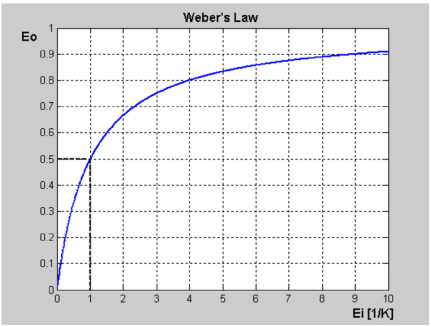


Figure 4. The Weber's law curve.

Thanks to Dr. Jacques Benkoski for his helpful remarks.

Moti Shefer
Trusight Ltd.,
Haifa, Israel
E-mail: moti@trusight.com

References
1. A. Belenky, E. Artyomov, A. Fish and O. Yadid-Pecht, *Wide dynamic Range (WDR) Imaging,* **The Neuromorphic Engineer 1** (1), p. 4, Spring 2004.
2. M. Shefer, **Increasing the dynamic range of image sensors: Critiques of three commonly practiced methods,** www.trusight.com/motis_place.htm, Sept. 2005.
3. M. Shefer, **What's DR?,** www.trusight.com/motis_place.htm, Sept. 2005.
4. M. Shefer, **Neuromorphic-retinal dynamic range compression as applied to image display,** www.trusight.com/motis_place.htm, Sept. 2005.
5. M. Shefer, **AGC models for Retinal signal processing,** M.Sc. Thesis, EE Dept., Technion, I.I.T., November 1979.
6. S. Polyak, **The Retina,** University of Chicago Press, 1941.
7. F. Ratliff, **Mach Bands: Quantitative studies on neural networks in the Retina,** Holden-Day Inc., 1965.
8. G. Westheimer, *Visual acuity and spatial modulation thresholds,* **Handbook of sensory physiology 7** (4), pp. 170-187, Springer Verlag, 1972.
9. J. E. Dowling, B. Ahinger and W. L. Hadden, *The interplexiform cell: A new type of Retinal neuron,* **Investigations in Ophthalmology 15,** pp. 916-926, 1976.
10. S. Chen and R. Ginosar, *Adaptive sensitivity CCD image sensor,* **Proc. SPIE 2415,** pp. 303-309, 1995.

Merolla, *continued from p. 10*

property of neurons, we can directly plot a metric that quantifies phase-locking: the task of finding neuron parameters that enhance phase-locking can now occur interactively (see Figure 2). The ostensible real-time advantage of neuromorphic systems over other modeling techniques is now apparent.

The high-performance USB2.0 interface we have described here has revolutionized the way in which we interact with our neuromorphic systems. Our strategy for designing this interface was to optimize the hardware so that the much of the computational burden was transparent to the software. This strategy has paid off, as evident from the rapid development (three weeks) of the five customized real-time applications.

Table 1: Doing it on the fly[5]

Revolutionary	Chip	Visualization (all real time)
Bo Wen	Inner ear	Cochleagram (spectrum vs. time) with real sound input
Brian Taba	Axon Guidance	Activity and histograms of ON-OFF retinal inputs for balanced topographic refinement
John Arthur	Hippocampus	Rasters and chip activity phase-coded in color
John Wittig Jr	Cochlear Nucleus	Rasters and histograms showing phase locking to sound waveforms.
Paul Merolla	Visual Cortex	Orientation maps generated with drifting gratings

Paul Merolla, John Arthur, and John Wittig Jr
University of Pennsylvania
Bioengineering Department
Philadelphia, PA
E-mail: {pmerolla, jarthur, jwittig}@seas.upenn.edu

References
1. P. Merolla and K. Boahen, *A Recurrent Model of Orientation Maps with Simple and Complex Cells,* Advances in Neural Information Processing Systems 16, S. Thrun and L. Saul, Eds., MIT Press, pp. 995-1002, 2004.
2. K. A. Boahen, *Point-to-Point Connectivity Between Neuromorphic Chips using Address-Events,* IEEE Trans. on Circuits & Systems II 47 (5), pp. 416-434, 2000.
3. T. Choi, P. Merolla, J. Arthur, K. Boahen and B. Shi. *Neuromorphic Implementation of Orientation Hypercolumns,* IEEE Trans. on Circuits & Systems I 52 (6), pp. 1049-1060, 2005.
4. http://nehe.gamedev.net
5. http://www.neuroengineering.upenn.edu/boahen/index.htm

Development of a cortically-inspired active binocular-vision system

A joint project between my laboratory at the Hong Kong University of Science and Technology, Ning Qian at Columbia University, and Meihua Tai at the Polytechnic University (both NY) has been set up to develop an active binocular-vision system where visual control is based upon the distributed populations of cortical neurons. This system consists of a six-degree-of-freedom binocular-vision head and custom hardware for rapidly computing, communicating and combining the outputs of retinotopic maps of model cortical neurons.

The development of neuromorphic systems for cortically-inspired visual processing leads naturally to the incorporation of active gaze control. For example, the disparity-selective model neurons that are constructed using the disparity energy model are only accurate within a small spatial range due to effects such as phase wraparound. As a result, the active control of camera gaze can bring different parts of the image into the required disparity range. Gibson anticipated this when he argued that perception arises through an active process that involves adjustments of the perceptual organ.[1] Appropriately, he likens active senses to tentacles or feelers. Although the visual senses have the potential to acquire environmental information purely passively—as evidenced by our ability to engineer algorithms using stereo heads with fixed camera parameters to extract environmental depth—there are many computational advantages of incorporating active gaze control into perceptual processing.[2]

Our binocular active-vision head (see Figure 1) has three degrees of freedom for each eye: horizontal and vertical rotation, as well as rotation around the line of sight (torsion). Because the most rapid eye movements are associated with saccades, during the design phase we took care to ensure that saccadic eye movements performed by the head can match or exceed those observed in primates. However, since it appears that visual perception is shut down during a saccade, we were not particularly concerned with matching exact trajectories.

The addition of torsional control distinguishes this binocular vision head from most of those previously developed. In humans, the eyes cannot only rotate horizontally or vertically, but also within about 10° around the line of sight. Active neural control of this torsional component may be important in reducing the motion

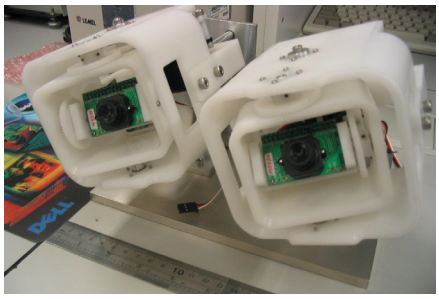


Figure 1. The six-degree-of-freedom active vision head.

of epipolar lines to enable stereopsis with smaller retina-fixed disparity search zones,[3] as well as in quickly stabilizing the retinal image during gaze shifts where both eyes and head move.[4]

The system also includes custom-designed hardware for computing the outputs of retinotopic arrays of artificial neurons (maps). These model the responses of populations of neurons within the visual cortex that are tuned to respond to different combinations of spatial/temporal frequency, orientation, and binocular disparity. For maximum expandability, we adopted a modular architecture. Computation is distributed among a number of identical boards, each of which (see Figure 2) contains a high-speed fixed-point digital signal processor (DSP) chip (the TI 6414 DSP) operating at 600MHz for computing the responses of the model neurons. Intra-

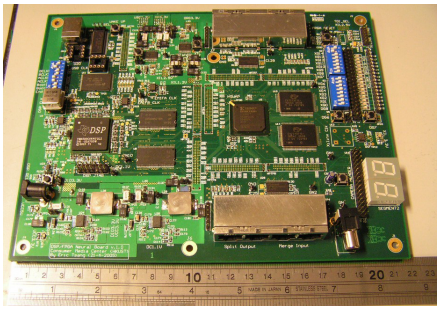


Figure 2. One of the boards used to compute cortical maps.

board communication is handled by a Xilinx Spartan III field-programmable gate array (FPGA) chip connected with low-voltage differential signalling serializers/deserializers. Each board supplements the on-chip memories of the DSP and FPGA with 8MB of synchronous dynamic random access memory (SDRAM) and 4MB of static RAM.

On these boards, computation and communication is similar to that previously developed for neuromorphic models of the retinotopic arrays of neurons tuned to different orientations in the primary visual cortex. In these earlier implementations, computation was performed using custom-designed mixed-signal analog-digital chips.[5] However, our current system uses digital processing to enable rapid reconfiguration of the processing performed by each board, sacrificing low power consumption for enhanced flexibility. This enables more rapid experimentation with different models of bio-inspired processing. However, because the structure of the overall system is similar to that used by multi-chip neuromorphic networks, we expect that the processing performed by each board will be easily mapped onto mixed-signal neuromorphic VLSI (very large silicon integration) chips. Thus, we view this system as an intermediate step between software simulations on a standard personal computer and multi-chip networks of custom-designed chips.

The project described here is supported by the Institute of Neuromorphic Engineering and the Hong Kong Research Grants Council. The author would like to acknowledge the efforts of E. Tsang, S. Lam, Y. C. Meng and C. H. Fung at the Hong Kong University of Science and Technology in developing the electronic hardware and the active-vision head described here.

Bertram Shi
Department of Electrical and Electronic Engineering
Hong Kong University of Science and Technology
Kowloon, Hong Kong
E-mail: eebert@ust.hk

References

1. J. J. Gibson, **The Senses Considered as Perceptual Systems,** Greenwood Press, Westport, CT, 1966.
2. Y. Aloimonos, I. Weiss and A. Bandyopadhyay, *Active vision,* **Int'l J. Comput. Vision 1,** pp. 333-356, 1988.
3. K. Schreiber, J. D. Crawford, M. Fetter and D. Tweed, *The motor side of depth vision,* **Nature 410,** pp. 819-22, 2001.
4. D. Tweed, T. Haslwanter and M. Fetter, *Optimizing gaze control in three dimensions,* **Science 281,** pp. 1363-1366, 1363.
5. T. Y. W. Choi, P. A. Merolla, J. V. Arthur, K. A. Boahen and B. E. Shi, *Neurmorphic Implementation of Orientation Hypercolumns,* **IEEE Trans. on Circuits and Systems-I: Regular Papers 52,** pp. 1049-1060, 2005.