Research in Progress

Photographic Memory

Neural network research is one of the hottest topics in computing nowadays, but the field is still in its infancy. The very features that give neural nets their unique attributes make them hard to simulate on "regular" serial computers, and harder still to model as physical devices. A number of esoteric devices are being researched that could someday revolutionize computer design. In the meantime, Aharon Agranat, a research fellow in Amnon Yariv's applied physics group, is developing a system to model the flexible complexity of neural nets using mature technologies. Caltech has several research groups working with neural nets, and the faculty includes some of the field's leading scientists.

Neural nets are loosely modeled after the brain's own structure, where every brain cell, or neuron, is connected to thousands of others. The connection strengths (or "weights") differ. Data (memories) and programs (such as how to process visual information to recognize a face) are stored as connection patterns of varying weights. When the brain learns, the patterns change.

Building a brain, with its approximately 10^{12} neurons, each with up to 10,000 connections, is still beyond anyone's wildest dreams. Scientists today are working with a handful of neurons, modeling systems from a dozen or so up to a few hundred. Even these small nets present complex problems.

Although microprocessing technology is pretty sophisticated, trying to squeeze a highly interconnected, three-dimensional structure onto an essentially two-dimensional chip is a tough proposition. With the additional condition that the connections be made variable, the problem becomes well-nigh impossible.

An alternative approach, using optical technology, shows promise. Since light beams pass through each other unhindered, and beam intensity is easily controlled, the connection problem vanishes. In fact, small neural nets have been built entirely in optics; but optical information processing technology is not nearly as advanced as its silicon analog.

Agranat decided to put the best-developed features of silicon and optical technology together. The result is a system where silicon neurons process data (connection strengths) from an optically loaded memory. He was assisted this past summer by SURF (Summer Undergraduate Research Fellowship) student Chuck Neugebauer, a senior interested in VLSI physics.

Agranat's system codes the connection weights for a chip with N neurons as an N × N matrix. Each matrix element codes the weight from the neuron in the element's row to the neuron in its column. Transformed into optical data, the matrix becomes a pattern of bright and dark dots. The brighter the dot, the greater the weight. A personal computer from a home electronics store generates the pattern according to any of several algorithms. ("It's not standard, but we got it really cheap, and it's got good graphics capabilities," Neugebauer said.) Once in the computer, the pattern can be modified at will. For example, the chip could be "taught" by modifying the pattern between successive runs until a desired output is produced.

But how does the pattern get on the chip? That's the key to Agranat's off-the-shelf design. The computer's video drive is connected to one of those tiny LCD (liquid crystal display) TV sets—the kind that have recently become the gift of choice for consumer electronics addicts. A system of lenses reduces the image 20-fold and focuses it on a charge-coupled device, or CCD. A CCD is the image-sensing device in a television camera. The dot matrix becomes puddles of electric charge on the CCD, and the chip reads them off just as a camera's chips would read a video image. "Basically, we can put any pattern of dots on the CCD we need, using the LCD screen as a spatial light modulator," Neugebauer said.

"People are using LCD-TVs in all sorts of optical computing setups," Agranat notes. "They are cheap, about $200, and readily available. There are high-quality spatial light modulators built specifically for scientific use, but they're very expensive. There's a mageneto-optical device with a 48 × 48 matrix. It's a very good device, but it costs about $5,000, and it can't make shades of gray. We need shades of gray for the weights. With the LCD-TV, we can make a 100 × 100 matrix having more information. The optical setup is routine, but it's not trivial. The image has to be placed accurately to within 10 microns everywhere on the CCD, and the brightness has to be uniform as well."

The chip assembly is enclosed in a light-tight box with a shuttered aperture, like a still camera. A single flash exposure suffices to load the pattern.

"My background is in optics," Agranat said. "I had the idea, but I didn't have the expertise to build it. Then Chuck showed up. I built the optics, and he designed the chips. Usually this would be done as a collaboration of two professionals—I would find a postdoc to share the work. But here one of the professionals is an undergraduate, and that's unique. I gave him the architecture—the CCD, the integrators, and so forth, and he developed the designs completely on his own. Without him, this project would still be on paper."

In fact, the group developed three architectures over the summer. The CCD system is a semi-parallel, synchronous architecture: Just as creating a video image requires a number of sweeps equal to the number of rows on the screen, so processing the CCD's contents takes a number of computational cycles equal to the number of neurons. The other two architectures can calculate the network's output in either a single cycle (fully parallel synchronous processing) or continuously without cycling (parallel asynchronous processing). "Chuck dreamed up the the third one himself," Agranat said. "He got the idea for it, and he went back and figured out exactly what engineering principles and what components it would take to make it work. It's actually the easiest of the three to fabricate, so it may be the first one we'll see working."

All three designs have been sent out for fabrication. Some of the chips have returned recently, and the testing process is about to begin. Several variations on each architecture will be examined, and one design each will be chosen for the next scale-up. The largest chip in the current batch has 32 neurons, enough gray matter to perform simple computations.

A 1,000-neuron chip could be built with current technology, Agranat said, but there aren't as yet spatial light modulators big enough to load them. In the meantime, the 100-neuron range accessible through the LCD-TV is sufficient for much network research that would otherwise be impossibly expensive or time-consuming to run.□—*DS*

Computer's-Eye View

IF A PROPOSED Mars rover can find its way through Congress, it may use a Caltech-designed vision system to find its way around Mars. The vehicle would have to recognize and avoid obstacles unaided, as signals from an earthbound driver would take as long as twenty minutes to arrive.

Assistant Professor of Computation and Neural Systems Christof Koch, along with Brian Wilcox of the JPL vision group, and Carl Ruoff, a graduate student in mechanical engineering, are evaluating Koch's design for space applications. The project is supported by NASA funds. Current computer vision systems are based on so-called "expert system" programs. "Expert systems are very brittle," Koch said. "You have to have rules for everything. If you built a vision system that way, it would be made up of if-then rules like, 'If you have a red blob at a one-meter height in an office, it's probably a telephone.' But it could be a cup. So you add another rule: 'If it's round, it's a cup.' But maybe it's really an apple. Our approach is much more flexible."

"You never think about vision," Koch said. "You open your eyes and the world is there. But your eyes just make a big array of numbers—they're actually voltages, but conceptually they correspond to numbers. Out of that array you have to infer that one object is in front of another, that things are moving, how they're moving. This is called early vision—getting three-dimensional information from a two-dimensional intensity array."

Koch heads a research group working on early vision and related computational problems. The group has developed an algorithm to separate objects from background based on their relative motion. Since a computer's field of vision consists of a matrix of discrete points (pixels), a velocity vector can be assigned to each pixel. The velocity vector field is generated by comparing two successive images of the same scene and determining how each point moved between images. The system actually traces the apparent motion of pixels of constant brightness. It works from two assumptions. Neighboring points in the field of view move the same way unless they are on different objects. When the points have different motions, they are separated by discontinuities (sharp changes between adjacent vectors) coinciding with the object's edges.

Koch's algorithm poses the problem in terms of minimizing an energy function (an optimization problem). The problem resembles a landscape of hills and valleys. A boulder dropped on one of the hillsides should eventually roll to the lowest point in the valley. Computing the boulder's path mathematically is time-consuming, as thousands of iterations may be required over the entire landscape.

However, the problem is directly analogous to the behavior of a network of simple resistors. The junctions, or nodes, correspond to the pixels. An external voltage fed to each node corresponds to the brightness change at that pixel. Once the system has reached a stable state (minimum power dissipation), the voltage at each node corresponds to the velocity. "You can exploit the physics of the network to get the solution, instead of going to a computer and using a set of logical rules," Koch said.

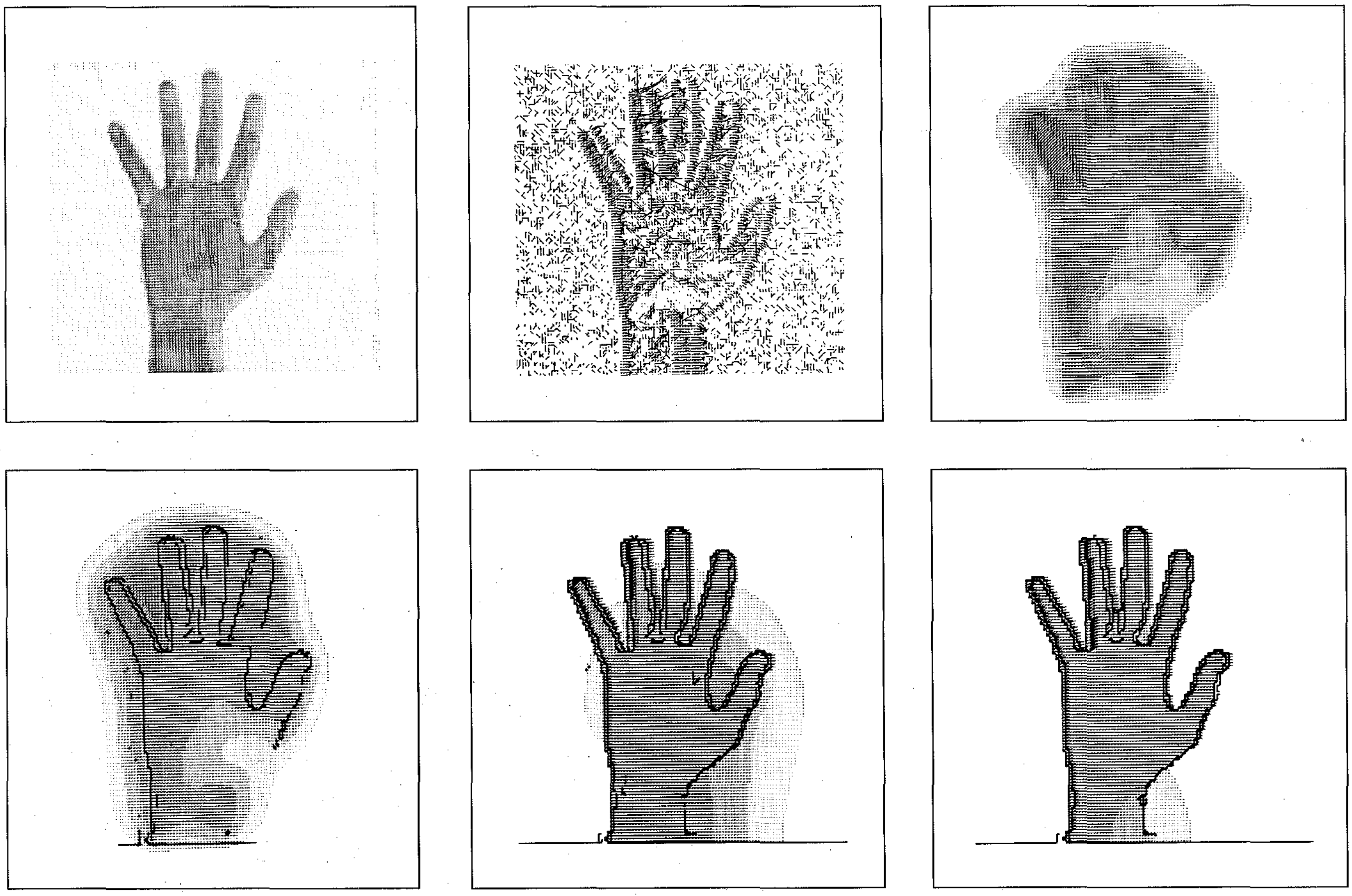
Other people have proposed relative-motion algorithms that can determine an object's direction and velocity. Unfortunately, these methods generate a velocity field that varies smoothly over the entire visual field. The smoothness hides an object's edges, leaving an unrecognizable blob. Furthermore, if overlapping objects move in opposite directions, the network averages their velocities, generating a zero-velocity region where the objects meet.

Koch's network design includes an edge simulation process. This so-called "line process" has recently been introduced in a variety of vision applications. The line process is a binary switch in each resistive link. When the line process is "off," the switch is closed and current flows—smoothing the velocity field. When it's "on," the switch is open. No current flows between the nodes, creating a velocity field discontinuity—an edge.

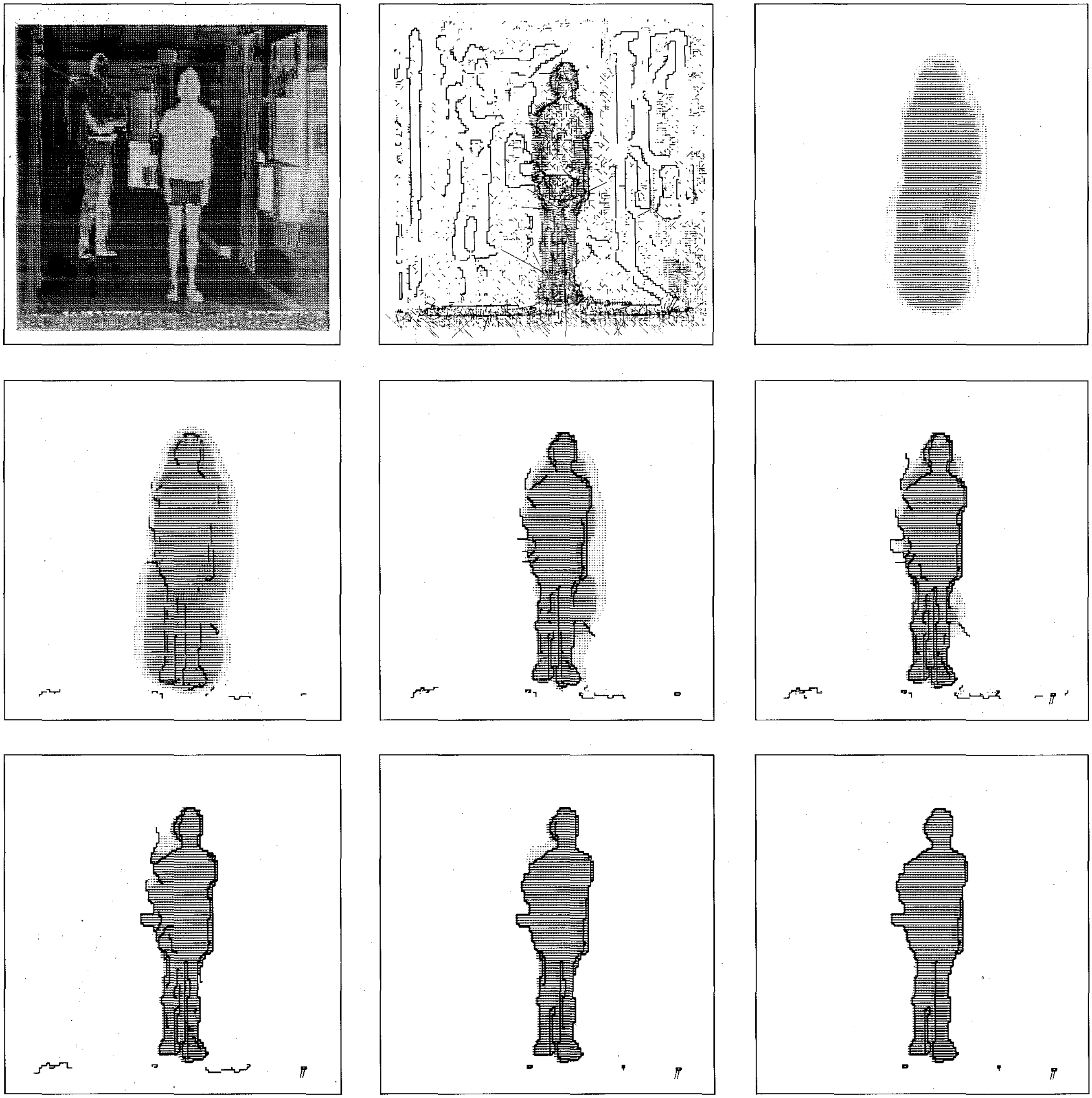
The result is a hybrid chip design with analog and digital components connecting each node. The analog components are variable resistors designed by Carver Mead, the Gordon and Betty Moore Professor of Computer Science (*E&S* June '87). The chip actually has two identical networks— one each for the x and y components of motion. Each pixel's x and y nodes are linked by a variable resistor.

The system operates in cycles. Initially all line processes are "off" and current flows until the field stabilizes. Next, an external unit evaluates each node. If a neighboring node has a significantly higher voltage, the line process between them turns "on," creating an edge. Evaluation is fast, even for large networks, because system-wide information isn't needed. Once all the nodes have been checked, the active line processes break their respective connections, and the current flows until another field stabilizes. The system re-evaluates the nodes, turning on new line processes as needed, and the cycle repeats. "In its final state, you can read off the velocities and discontinuities directly," Koch said. "This thing has almost solved the problem of object segmentation—deciding what surfaces belong to what objects. It's a very difficult problem."

The group has simulated the system on JPL's 32-node Mark III Hypercube parallel computer, and on smaller 4- and 16-node NCUBE computers on campus. The most recent work used 128 × 128 pixel "before" and "after" images from a video camera. Given a moving hand on a featureless background, the system generated an easily recognizable image, complete with fingers, in only seven cycles. (1,000 iterations of the resistive net alone generated a smooth-flow image resembling



"Seeing" a moving hand. Top row, from left: 128 × 128 pixel image from video camera; the computer compares it to a second image (not shown) in which the hand has moved as much as two pixels. Initial velocity data; "snow" in background also registers. Smooth-flow field after 1,000 iterations, with grayness proportional to velocity. Bottom row, from left: State of the hybrid network after one, three, and seven cycles respectively. Line processes grow toward each other along edges, and the velocity aura around the hand disappears. After thirteen cycles (not shown) it vanishes altogether.



In a hallway. Top row, from left: 128 × 128 video image; person in foreground is moving right and person in background is standing still. Discontinuities superimposed on initial velocity data; note band of camera noise across bottom of both images. Smooth-flow field after 1,000 iterations; noise vanishes as random variations cancel. Middle row: Hybrid network after one, three, and five cycles respectively. Bottom row: After seven, ten, and thirteen cycles. Line processes not supported by the velocity field wither, while line processes coinciding with edges grow.

a mitten.) A person walking down a hallway filled with stationary objects was cleanly separated from the background in thirteen cycles.

"Even though it's a Hypercube, it still takes 20 minutes to simulate two frames which may be a small fraction of a second apart. The chips, we think, will be able to do it in real time or very close," Koch said.

A 30 × 30 chip built to Koch's design has recently returned from the fabricator, and network testing is about to begin. If this is successful, a 128 × 128 chip could be built in two to three years with existing technology.

"The common housefly, *Musca domestica*, has 24,000 photoreceptors in each eye, or about 50,000 photoreceptors covering almost 360°. That's around 220 × 220. And *Musca domestica* is a very good flier—try to catch it. So we think with 128 × 128 we can do all the navigation we want. It may not be able to read, but our Mars rover doesn't really need to do that," Koch concluded. Even if the Mars rover never gets off the ground, this rugged, lightweight technology will find plenty of earthbound uses.□—*DS*