Vision by Man and Machine

5

How does an animal see? How might a computer do it? A study of stereo vision guides research on both these questions. Brain science suggests computer programs; the computer suggests what to look for in the brain.

• • •

Tomaso Poggio
April, 1984

The development of computers of increasing power and sophistication often stimulates comparisons between them and the human brain, and these comparisons are becoming more earnest as computers are applied more and more to tasks formerly associated with essentially human activities and capabilities. Indeed, it is widely expected that a coming generation of computers and robots will have sensory, motor and even "intellectual" skills closely resembling our own. How might such machines be designed? Can our rapidly growing knowledge of the human brain be a guide? And at the same time can our advances in "artificial intelligence" help us to understand the brain?

At the level of their hardware (the brain's or a computer's) the differences are great. The neurons, or nerve cells, in a brain are small, delicate structures bound by a complex membrane and closely packed in a medium of supporting cells that control a complex and probably quite variable chemical environment. They are very unlike the wires and etched crystals of semiconducting materials on which computers are based. In the organization of the hardware the differences also are great. The connections between neurons are very numerous (any one neuron may receive many thousands of inputs) and are distributed in three dimensions. In a computer the wires linking circuit components are limited by present-day solid-state technology to a relatively small number arranged more or less two-dimensionally.

In the transmission of signals the differences again are great. The binary (on-off) electric pulses of the computer are mirrored to some extent in the all-or-nothing signal conducted along nerve fibers, but in addition the brain employs graded electrical signals, chemical messenger substances and the transport of ions. In temporal organization the differences are immense. Computers process information serially (one step at a time) but at a very fast rate. The time course of their operation is governed by a computer-wide clock. What is known of the brain suggests that it functions much slower but that it analyzes information along millions of channels concurrently without need of clock-driven operation.

How, then, are brains and computers alike? Clearly there must be a level at which any two mechanisms can be compared. One can compare the tasks they do. "To bring the good news from Ghent to Aix" is a description of a task that can be done by satellite, telegraph, horseback messenger or pigeon post equally well (unless other constraints such as time are specified). If, therefore, we assert that brains and computers function as information-processing systems, we can develop descriptions of the tasks they perform that will be equally applicable to either. We shall have a common language in which to discuss them: the language of information processing. Note that in this language descriptions of tasks are decoupled from descriptions of the hardware that perform them. This separability is at the foundation of the science of artificial intelligence. Its goals are to make computers more useful by endowing them with "intelligent" capabilities, and beyond that to understand the principles that make intelligence possible.

In no field have the descriptions of information-processing tasks been more precisely formulated than in the study of vision. On the one hand it is the dominant sensory modality of human beings. If we want to create robots capable of performing complex manipulative tasks in a changing environment, we must surely endow them with adequate visual powers. Yet vision remains elusive. It is something we are good at; the brain does it rapidly and easily. It is nonetheless a mammoth information-processing task. If it required a conscious effort, like adding numbers in our head, we would not undervalue its difficulty. Instead we are easily lured into oversimple, noncomputational preconceptions of what vision really entails.

Ultimately, of course, one wants to know how vision is performed by the biological hardware of neurons and their synaptic interconnections. But vision is not exclusively a problem in anatomy and physiology: how nerve cells are interconnected and how they act. From the perspective of information processing (by the brain or by a computer) it is a problem at many levels: the level of computation (What computational tasks must a visual system perform?), the level of algorithm (What sequence of steps completes the task?) and then the level of hardware (How might neurons or electronic circuits execute the algorithm?). Thus an attack on the problem of vision requires a variety of aids, including psychophysical evidence (that is, knowledge of how well people can see) and neurophysiological data (knowledge of what neurons can do). Finding workable algorithms is the most critical part of the project, because algorithms are constrained both by the computation and by the available hardware.

Figure 5.1 STEREO VISION BY A COMPUTER is shown in aerial photographs (provided by Robert J. Woodham). They were made from different angles so that objects in each have slightly different positions. The images were made by a mosaic of microelectronic sensors, each of which measures the intensity of light along a particular line of sight, as do the photoreceptor cells of the eye. The map at the bottom was generated by a computer programmed to follow a procedure devised by David Marr and the author and further developed by W. Eric L. Grimson. The computer filtered the images to emphasize spatial changes in intensity. Then it performed stereopsis: it matched features from one image to the other, determined the disparity between their positions and calculated their relative depths in the three-dimensional world. Increasing elevations in the map are coded in colors from blue to red.

Here I shall outline an effort in which I am involved, one that explores a sequence of algorithms first to extract information, notably edges, or pronounced contours in the intensity of light, from visual images and then to calculate from those edges the depths of objects in the three-dimensional world. I shall concentrate on a particular aspect of the task, namely stereopsis, or stereo vision (see Figure 5.1). Not the least of my reasons is the central role stereopsis has played in the work on vision that my colleagues and I have done at the Artificial Intelligence Laboratory of the Massachusetts Institute of Technology. In particular, stereopsis has stimulated a close investigation of the very first steps in visual information processing. Then too, stereopsis is deceptively simple. As with so many other tasks that the brain performs without effort, the development of an automatic system with stereo vision has proved to be surprisingly difficult. Finally, the study of stereopsis benefits from the availability of a large body of psychophysical evidence that defines and constrains the problem.

The information available at the outset of the process of vision is a two-dimensional array of measurements of the amount of light reflected into the eye or into a camera from points on the surfaces of objects in the three-dimensional visual world. In the human eye the measurements are made by photoreceptors (rod cells and cone cells), of which there are more than 100 million. In a camera that my colleagues and I use at the Artificial Intelligence Laboratory the processes are different but the result is much the same. There the measurements are made by solid-state electronic sensors. They pro-