10

"Hand–Eye" Research

THE MOTIVATION FOR MUCH OF THE COMPUTER VISION RESEARCH THAT I HAVE described during this period was to provide information to guide a robot arm. Because the images that could be analyzed best were of simple objects such as toy blocks, work was concentrated on getting a robot arm to stack and unstack blocks. I'll describe some typical examples of this "hand–eye" research, beginning with a project that did not actually involve an "eye."

10.1 At MIT

A computer-guided mechanical "hand" was developed by Heinrich A. Ernst in 1961 as part of his Electrical Engineering Sc.D. work at MIT.[1] (His advisor was Claude Shannon.) The hand, named MH-1, was a "mechanical servomanipulator [an American Machine and Foundry model 8] adapted for operation by the TX-0 computer." It used tactile sensors mounted on the hand to guide it because, as Ernst wrote, "organs for vision are too difficult to build at the present time." The abstract of Ernst's thesis describes some of what the system could do:

> [O]ne program consisting of nine statements will make the hand do the following: Search the table for a box, remember its position, search the table for blocks, take them and put them into the box. The position of the objects is irrelevant as long as they are on the table. If as a test for the built-in mechanical intelligence, the box should be taken away and placed somewhere else while the hand searches for blocks, MH-1 will remember the new position of the box and continue to work with it as soon as it has realized the change in the situation, that is, has bumped into the box while looking for blocks. This will be done automatically, without any need to mention it in the specific program for this block-and-box performance.

Actually, MH-1 was not the first computer-guided hand, although it was the first to employ touch sensors to guide its motion. One was developed and patented in 1954 by George Devol, an American engineer. Based on this invention, he and another engineer, Joseph F. Engelberger, founded Unimation, Inc. Soon after, they installed a prototype of their first industrial robot, called a "Unimate," in the General Motors Corporation Ternstedt Division plant near Trenton, New Jersey.

Back at MIT, Ph.D. students Patrick Winston (later an MIT professor and director of its AI Laboratory), Thomas O. Binford (later a Stanford professor), Berthold K. P. Horn (later an MIT professor), and Eugene Freuder (later a University of New Hampshire professor) developed a system that used an AMF Versatran robot

https://doi.org/10.1017/CBO9780511819346.013 Published online by Cambridge University Press

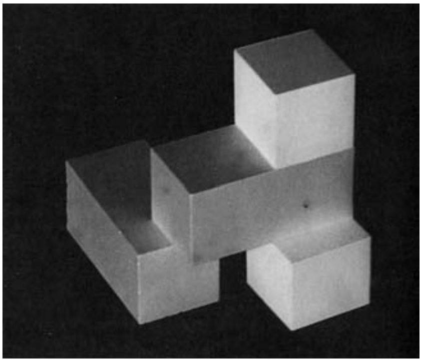

Figure 10.1. A block arrangement for the MIT copy demo. (Used with permission of Berthold Horn.)

arm to "copy" a configuration of blocks. The scene consisting of the blocks was first scanned, and lines were extracted from the image using a "line-finder," which was under development by Binford and Horn.[2] Using these lines, objects in the image were identified, and a plan was made for the robot arm to disassemble the blocks in the scene. The robot arm then carried out this plan and reassembled the blocks in their original configuration. The system was demonstrated in December 1970 for various configurations of blocks. An example block configuration successfully handled by their "copy demo" is shown in Fig. 10.1. (A film, called *Eye of the Robot*, showing the copy demo in action is available at http://projects.csail.mit.edu/films/aifilms/digitalFilms/9mpeg/88-eye.mpg.)

The system depended on precise illumination and carefully constructed blocks. Attempts to extend the range of computer vision to less constrained scenes led to further concentration at MIT and elsewhere on the early stages of vision processing. I'll describe some of the ensuing work on these problems in more detail later.[3]

10.2 At Stanford

Meanwhile at John McCarthy's SAIL, a team led by Professor Jerome Feldman (1938–) pursued work on hand–eye projects using the PDP-1 and later the PDP-6 and PDP-10 computers.[4] McCarthy later told me that he got interested in robots because of his interest in computer vision. He was not very excited about the work in pattern recognition – it was "discrimination" rather than "description." "If you want to pick something up, you have to describe it not merely recognize it."[5]

In 1966, SAIL had acquired a Rancho Los Amigos Hospital electromechanical prosthetic arm. By the spring of 1967, a hand–eye system was developed by Karl Pingle, Jonathan Singer, and Bill Wichman that could use a TV camera and primitive vision routines to locate blocks scattered on a table. Using the information thus obtained, it could control the arm to sort the blocks.[6] According to the authors,

> One section of the system scans the TV image to find the outer edge of an object, then traces around the outside edges of the object using a gradient operator [an edge detector] to find the location and direction of the edge. Curve fitting routines fit straight lines to a list of points found on the edges and calculate the position of the corners. . . . A second section of the

https://doi.org/10.1017/CBO9780511819346.013 Published online by Cambridge University Press

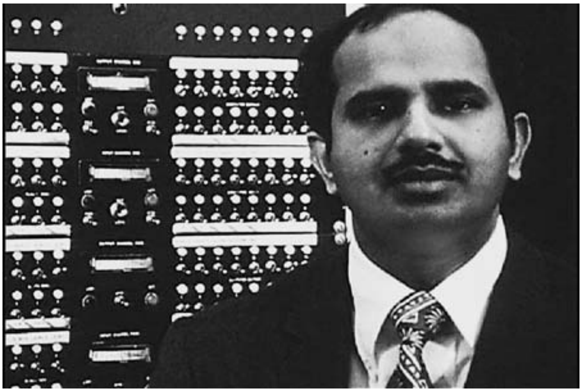

Figure 10.2. Raj Reddy. (Photograph courtesy of Raj Reddy.)

system is devoted to control of the arm.... The sections which control it consist of a solution program which calculates the angular position required at each actuator and a servo program which drives the arm to the desired positions ...

Les Earnest wrote that this was "the first robotics visual feedback system," although "only the outer edges of the blocks were observed, [and] the hand had to be removed from view when visual checking was done...."[7]

Several versions of this block-sorting and stacking system were demonstrated. In one, the system located colored blocks on a table and placed them in separate stacks of red and blue blocks.[8]

By 1971, a vision-guided block stacking system solved the "instant insanity" puzzle.[9] In that puzzle, four cubes, each face of which has one of four different colors, must be arranged in a tower such that each side of the tower shows four different colors. The system, running on SAIL's PDP-10 computer, used a TV camera equipped with a turret of four lenses and a color wheel to locate four cubes on a table top. The arm picked up and turned each cube to expose all faces to the camera. Then, knowing the color of each face and having found a solution to the puzzle, the arm stacked the cubes in a tower exhibiting the solution. [A silent, 16-mm color six-minute film, titled *Instant Insanity*, was made by Richard Paul and Karl Pingle in August 1971 and shown at the second International Joint Conference on Artificial Intelligence (IJCAI) in London. The film can be seen at http://www.youtube.com/watch?v=O1oJzUSITeY.]

Dabblal Rajagopal "Raj" Reddy (1937– ; Fig. 10.2) was the first Ph.D. student of Stanford's new Department of Computer Science. His thesis research was on speech recognition. After obtaining his Ph.D. in 1966, Reddy joined Stanford's faculty and continued research on speech recognition at SAIL. While there he participated in a project to control a hand–eye system by voice commands.[10] As stated in a project review, "Commands as complicated as 'Pick up the small block in the lower lefthand corner,' are recognized and the tasks are carried out by the computer controlled arm." (The system was demonstrated in a 1969 fifteen-minute, 16-mm color, sound film showing some of Reddy's results on speech recognition. It is titled *Hear Here*

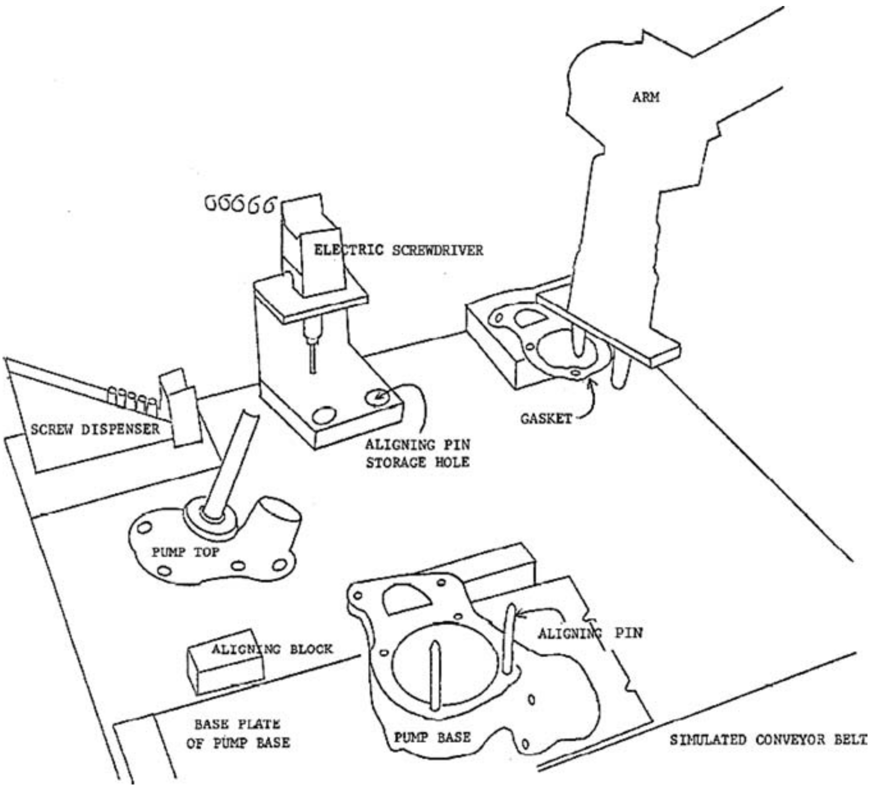

Figure 10.3. Diagram of a water pump assembly workspace. (Illustration used with permission of Robert Bolles.)

and was produced by Raj Reddy, Dave Espar, and Art Eisensen. The film is available at http://www.archive.org/details/sailfilm_hear.) In 1969, Reddy moved to CMU where he pursued research in speech recognition and later became Dean of CMU's School of Computer Science.

In the early 1970s, the Stanford team used a vision system and a new electromechanical hand designed by mechanical engineering student Vic Scheinman[11] to assemble a model "T" Ford racing water pump.[12] An industrial-style setup was used with tools in fixed places, screws in a feeder, and the pump body and cover on a pallet. A diagram of the workspace is shown in Fig. 10.3.

The hand–eye system executed the following complex set of steps that was computed previously:

> locate the pump base, move it into standard position, determine the final grasping position by touch, place the pump base in its standard position, insert two pins to guide the alignment of the gasket and cover, put on the gasket, visually check the position of the gasket, locate the cover by touch, put on the cover over the guide pins, pick up a hex head power screw driver, pick up a screw from the feeder, screw in the first two screws, remove the aligning pins, screw in the last four screws, and finally check the force required to turn the rotor.

https://doi.org/10.1017/CBO9780511819346.013 Published online by Cambridge University Press

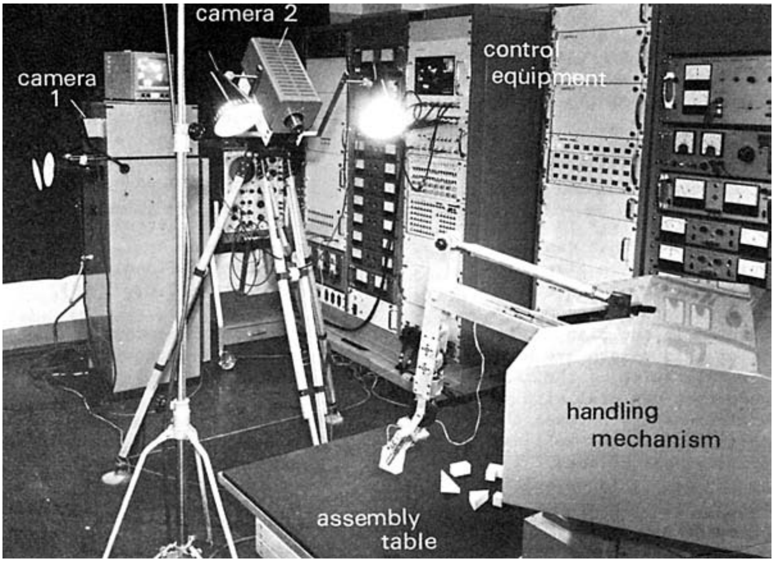

Figure 10.4. Hitachi's HIVIP robotic assembly system.

A film of the water pump assembly can be seen at http://www.archive.org/details/sailfilm_pump. It is also available at http://www.saildart.org/films/, along with several other Stanford AI Lab films.

10.3 In Japan

Hand–eye work was also being pursued at Hitachi's Central Research Laboratory in Tokyo. There, Masakazu Ejiri and colleagues developed a robot system called HIVIP consisting of three subsystems called EYE, BRAIN, and HAND. (See Fig. 10.4.) One of EYE's two television cameras looked at a plan drawing depicting an assembly of blocks. The other camera looked at some blocks on a table. Then, BRAIN figured out how to pick up and assemble the blocks as specified in the drawing, and HAND did the assembly.[13]

10.4 Edinburgh's "FREDDY"

During the late 1960s and into the 1970s, researchers under the direction of Professor Donald Michie in the Department of Machine Intelligence and Perception at the University of Edinburgh developed robot systems generally called "FREDDY."[14] The best known of these was the hand–eye system FREDDY II, which had a large robot arm and two TV cameras suspended over a moving table. Even though the arm did not move relative to the room, it did relative to its "world," the table. The setup is shown in Fig. 10.5.

A demonstration task for FREDDY II was to construct complete assemblies, such as a toy car or boat, from a kit of parts dumped onto the table. The aim was to develop AI techniques that could provide the basis for better industrial assembly robots, that

https://doi.org/10.1017/CBO9780511819346.013 Published online by Cambridge University Press

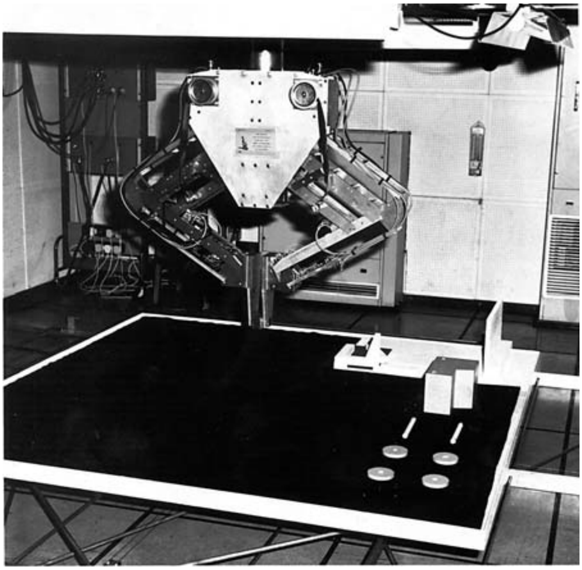

Figure 10.5. FREDDY II, the University of Edinburgh robot. (Photograph courtesy of University of Edinburgh.)

is, robots that were more versatile, more reliable, and more easily programmed than those in operation at that time.

At the beginning, the component parts were in an unorganized jumble. FREDDY had to find and identify them and then lay them out neatly. Once it had found all the parts needed, FREDDY could then perform the assembly sequence, using a small workstation with a vice.

Isolated parts were recognized from features of their outline (corners, curves, etc.), their holes, and their general properties. These were taught to FREDDY by showing it different views of each of the parts in a prior training phase.

To deal with heaps of parts, FREDDY applied several tactics: It could try to find something protruding from the heap, which it could grasp and pull out; it could attempt to lift something (unknown) off the top; or it could simply plough the hand through the heap to try to split it into two smaller ones.

Constructing the assembly was performed by following a sequence of instructions that had been programmed interactively during the training phase. Some instructions were simple movements, but others were much more sophisticated and used the force sensors in the hand. For example, in a "constrained move," the hand would slide the part it held along a surface until it hit resistance; in "hole fitting," the hand would fit one part (such as an axle) into a hole in another (such as a wheel) by feel, as humans do.

FREDDY could assemble both the car and the boat when the two kits were mixed together and dumped on the table. It took about four hours to do so, primarily because of the limited power of FREDDY's two computers.[15] The main computer was an Elliot 4130 with 64k 24-bit words (later upgraded to 128k) and with a clock

https://doi.org/10.1017/CBO9780511819346.013 Published online by Cambridge University Press

speed of 0.5 MHz. It was programmed in the POP-2 language. The robot motors and cameras were controlled by a Honeywell H316, with 4k 16-bit words (later upgraded to 8k words, at a cost of about $8,000 for the additional 8k bytes!).

Harry Barrow (1943–), a key person involved in the work, later gave this account of FREDDY's operation:[16]

> [A]cquiring an image from the TV camera took quite a few seconds and processing took even longer, and in a single run FREDDY took between 100 and 150 pictures! It took a picture every time it picked up an object to check it has successfully lifted it and not dropped it, and it took a picture every time it put an object down to verify the space was empty. It also scanned the entire world (which required multiple pictures) several times to make a map, and it looked at each object from two different cameras to do some stereo-style estimation of position and size. In fact, the system made the most intensive use of image data of any robot system in the world. The Stanford and MIT systems only took a very small number of pictures to perform their tasks, and relied heavily on dead reckoning and things not going wrong. We, on the other hand, assumed that things were likely to go wrong (objects dropped, rolling, etc.) and made our system highly robust. I really believe that in many ways it was probably the most advanced hand–eye system in existence at the time.

FREDDY is now on permanent exhibition in the Royal Scottish Museum in Edinburgh, with a continuous-loop movie of FREDDY assembling the mixed model car and boat kits.

Hand–eye research at Edinburgh was suspended during the mid-1970s in part owing to an unfavorable assessment of its prospects in a study commissioned by the British Science Research Council. (I'll have more to say about that assessment later.)

Notes

1. Heinrich A. Ernst, "MH-1, A Computer-Operated Mechanical Hand," Sc.D. thesis, Massachusetts Institute of Technology, Department of Electrical Engineering, 1962. Available online at http://dspace.mit.edu/bitstream/handle/1721.1/15735/09275630.pdf?sequence=1. [141]
2. Reported in Berthold K. P. Horn, "The Binford–Horn Line Finder," MIT AI Memo 285, MIT, July 1971 (revised December 1973). Available online at http://people.csail.mit.edu/bkph/AIM/AIM-285-OPT.pdf. [142]
3. Patrick Winston gives a nice description of the MIT programs just mentioned (including the copy demo and Winston's own thesis work on learning structural descriptions of object configurations) in Patrick H. Winston, "The MIT Robot," *Machine Intelligence 7*, Bernard Meltzer and Donald Michie (eds.), pp. 431–463, New York: John Wiley and Sons, 1972. [142]
4. For a brief review, see Lester Earnest (ed.), "Final Report: The First Ten Years of Artificial Intelligence Research at Stanford," Stanford Artificial Intelligence Laboratory Memo AIM-228 and Stanford Computer Science Department Report No. STAN-CS-74-409, July 1973. (Available online at http://www-db.stanford.edu/pub/cstr/reports/cs/tr/74/409/CS-TR-74-409.pdf.) For more details see Jerome A. Feldman *et al.*, "The Stanford Hand–Eye Project," *Proceedings of the IJCAI*, pp. 521–526, Washington, DC, 1969, and Jerome A. Feldman *et al.*, "The Stanford Hand-Eye Project – Recent Results," presented at IFIP Congress, Stockholm, 1974. [142]
5. John McCarthy, private communication, August 11, 2007. [142]

6. The system is described in Karl K. Pingle, Jonathan A. Singer, and William M. Wichman, "Computer Control of a Mechanical Arm through Visual Input," *Proceedings of the IFIP Congress (2)*, pp. 1563–1569, 1968. The vision part of the system is described in William Wichman, "Use of Optical Feedback in the Computer Control of an Arm," Stanford Electrical Engineering Department Engineers thesis, August 1967 (and also appears as Stanford Artificial Intelligence Memo AIM-56, 1967.) [142]
7. Lester Earnest, *op. cit.* [143]
8. *Butterfinger*, an 8-minute, 16-mm color film showing a version of this sorting system in operation was produced and directed by Gary Feldman in 1968. The film is available at http://projects.csail.mit.edu/films/aifilms/digitalFilms/1mp4/09-robot.mp4. [143]
9. The system is described in Jerome Feldman *et al.*, "The Use of Vision and Manipulation to Solve the 'Instant Insanity' Puzzle," *Proceedings of the IJCAI*, pp. 359–364, London: British Computer Society, September 1971. [143]
10. The system is described in Les Earnest *et al.*, "A Computer with Hands, Eyes, and Ears," *Proceedings of the 1968 Fall Joint Computer Conference*, Washington, DC: Thompson, 1968. [143]
11. Victor D. Scheinman, "Design of a Computer Manipulator," Stanford AI Memo AIM-92, June 1, 1969. [144]
12. This task is described in Robert Bolles Bolles, Robert and Richard Paul, Paul, Richard "The Use of Sensory Feedback in a Programmable Assembly System," Stanford AI Laboratory Memo AIM-220, Stanford Computer Science Department Report STAN-CS-396, October 1973, which is available online at ftp://reports.stanford.edu/pub/cstr/reports/cs/tr/73/396/CS-TR-73-396.pdf. [144]
13. See Masakazu Ejiri *et al.*, "An Intelligent Robot with Cognition and Decision-Making Ability," *Proceedings of the IJCAI*, pp. 350–358, London: British Computer Society, September 1971, and Masakazu Ejiri *et al.*, "A Prototype Intelligent Robot That Assembles Objects from Plan Drawings," *IEEE Transactions on Computers*, Vol. 21, No. 2, pp. 161–170, February 1972. [145]
14. This is an abbreviation, according to Donald Michie, of Frederick, an acronym of Friendly Robot for Education, Discussion and Entertainment, the Retrieval of Information, and the Collation of Knowledge. [145]
15. http://www.aiai.ed.ac.uk/project/freddy/. The key reference is A. P. Ambler, H. G. Barrow, C. M. Brown, R. M. Burstall, and R. J. Popplestone, "A Versatile Computer-Controlled Assembly System," *Artificial Intelligence*, Vol. 6, pp. 129–156, 1975. [146]
16. E-mail note from Harry Barrow of January 3, 2009. [147]