Robots, After All

Hans Moravec

Carnegie Mellon University
Robotics Institute

August 2003

Computers have invaded everyday life, and networked machines are worming their way into our gadgets, dwellings, clothes, even bodies. But if pervasive computing soon handles most of our information needs, it will still not clean the floors, take out the garbage, assemble kit furniture or do any of a thousand other other essential physical tasks. The old dream of mechanical servants will remain mostly unmet.

Robot inventors in home, university and industrial laboratories have tinkered with the problem for most of the century. While mechanical bodies adequate for manual work can be built, artificial minds for autonomous servants have been frustratingly out of reach. The problem's deceptive difficulty fooled generations of workers who attempted to solve it using computers.

The first electronic computers in the 1950s did the work of thousands of clerks, seeming to transcend humans, let alone other machines. Yet the first reasoning and game-playing programs on those computers were a match merely for single human beginners, and each only in a single narrow task. And, in the 1960s, computer-linked cameras and mechanical arms took hours to unreliably find and move a few white blocks on a black tabletop, much worse than a toddler. A modest robot industry did appear, but consisted only of arms and vehicles following predetermined trajectories. The situation did not improve substantially for decades, and disheartened waves of robotics devotees.

But things are changing. Robot interactive behavior wildly impossible in the 1970s and 1980s became experimental demonstrations in the 1990s: mobile robots mapped and navigated unfamiliar office suites [1], robot vehicles drove themselves, mostly unaided, across entire countries [2], computer vision systems located textured objects and tracked and analyzed faces in real time. Programs that recognized text and speech became commercially successful. Market success extended to physical robots as the 2000s began: Sony has sold hundreds of thousands of the AIBO robot pet despite its over-$1000 price, and several small robot vacuum cleaners, especially the affordable iRobot Roomba, seem to gaining customer acceptance. Not far behind, dozens of companies, established and new, are developing cleaning and transport robots using new sensors, leading edge computers and algorithms licensed from research efforts. Emerging capabilities include the ability of mobile robots to navigate ordinary places without special markers or advance maps. Some systems map the surroundings in 2D or even 3D as they travel, enabling the next step of recognizing structural features and smaller objects. Why suddenly now?

Trick of Perspective

The short answer is that, after decades at about 1 MIPS (**M**illion **I**nstructions **P**er **S**econd, each instruction representing work like adding two ten-digit numbers), computer power available to research robots shot through 10, 100 and now 1,000 MIPS in the 1990s. This is odd because the cost-effectiveness of computing rose steadily all those decades. In 1960 computers were a new and mysterious factor in the cold war, and even outlandish possibilities like artificial intelligence (AI) warranted significant investment. In the early

1960s AI programs ran on the era's supercomputers, similar to those used for physical simulations by weapons physicists and meteorologists. By the 1970s the promise of AI had faded, and the effort limped for a decade on old hardware. In contrast, weapons labs upgraded repeatedly to new supercomputers. In the 1980s, departmental computers gave way to smaller project computers then to individual workstations and personal computers. Machine costs fell and their numbers rose, but power stayed at 1 MIPS. By 1990 the research environment was saturated with computers, and only then did further gains manifest in increased power rather than numbers.

Mobile robot research might have blossomed sooner had the work been done on supercomputers, but pointlessly. At best, a mobile robot's computer could substitute for a human driver, a function worth perhaps $10 an hour. Supercomputer time cost at least $500 per hour. Besides, dominant opinion in the AI labs, dating from when computers did the work of thousands, was that, with the right program, 1 MIPS could encompass any human skill. The opinion remained defensible in the 1970s, as reasoning and game-playing programs performed at modest human levels.

For the few researchers in the newborn fields of computer vision and robotics, however, 1 MIPS was obviously far from sufficient. With the best programs, single images crammed memory, simply scanning them consumed seconds, and serious image analysis took hours. Human vision performed far more elaborate functions many times a second.

Hindsight enlightens. Computers calculate using as few gates and switching operations as possible. Human calculation, by contrast, is a laboriously learned, ponderous, awkward, unnatural behavior. Tens of billions of neurons in our vision and motor systems strain to analogize and process a digit a second. If our brain were rewired into 10 billion arithmetic circuits, each doing 100 calculations a second, by a mad computer designer with a future surgical tool, we'd outcompute 1 MIPS computers a millionfold, and the illusion of computer power would be exposed. Robotics, in fact, gave us an even better exposé.

Though spectacular underachievers at the wacky new stunt of longhand calculation, we are veteran overachievers at perception and navigation. Our ancestors, across hundreds of millions of years, prevailed by being frontrunners in the competition to find food, escape danger and protect offspring. Existing robot-controlling computers are far too feeble to match this massive ultra-optimized perceptual inheritance. But by how much?

The vertebrate retina is understood well enough to be a kind of Rosetta stone roughly relating nervous tissue to computation. Besides light detectors, the retina contains edge- and motion-detecting circuitry, packed into a little tenth-millimeter-thick, two-centimeter-across patch that reports on a million image regions in parallel about ten times a second via the optic nerve. In robot vision, similar detections, well coded, each require the execution of a few hundred computer instructions, making the retina's 10 million detections per second worth over 1,000 MIPS. In a risky extrapolation that must serve until something better emerges, this implies it would take about 50,000 MIPS to functionally imitate a gram of neural tissue, and almost 100 million MIPS (or 100 trillion instructions per second) to emulate the 1,500 gram human brain. By that measure PCs in 2003 are just a match for insect nervous systems, or the 0.01 gram brain of a guppy. Coordinated insectlike behavior in robots is probably best exhibited in the exciting field of Robocup robot soccer.

An international community of researchers began in 1993 to organize an effort to develop fully autonomous robots that could eventually compete in human soccer games just as chess computers compete in human chess tournaments. The incremental development would be tested in annual machine/machine tournaments. The first "RoboCup" games were held at a 1997 Artificial Intelligence conference in Nagoya, Japan. Forty teams entered in three competition categories, simulation, small robots and middle-size robots (the next size step, human scale, was reserved for the future). The small robot teams (of about five coffee-can-sized players) were each controlled by an outside computer that viewed the billiard-table-sized playing field

through an overhead color camera. To simplify the problem, the field was uniformly green, the ball was bright orange and the players top surfaces each had a unique pattern of large dots, relatively easy for programs to track. The middle size robots, about the size of breadboxes, had cameras and computers onboard, and played on a similarly colored but larger field. Action was fully preprogrammed, no human intervention was allowed during play. In the first tournament, merely finding and pushing the ball was a major accomplishment (never mind the goal location), but the conference encouraged participants to share developments, and play improved in subsequent years. In 1998 Sony provided some AIBO robot dogs for a new competition category. Almost 400 teams signed up for RoboCup 2003, in Padua, Italy, and regional tournaments were introduced to cull the final tournament competitors by more than half. AIBOs have became increasingly popular. Remarkably cute in play, they provide a standard, reliable, prebuilt hardware platform that needs only soccer software. In recent tournaments, the best teams frequently exhibit effective coordinated (goal directed!) behavior, intelligent blocking, even passing.

Though PCs in 2003 are still a daunting 100,000 times too weak, the goal of human performance is probably not impossibly far away. Computer power for a given price roughly doubled each year in the 1990s, after doubling every 18 months in the 1980s, and every two years prior. Twenty or thirty more years at the present pace would close the gap. Or, estimating the design effort ahead, the first multicelled animals with nervous systems appeared about 550 million years ago, ones with brains as advanced as guppies' perhaps 200 million years later. Self-contained robots covered similar ground in about 20 years. If we accept that evolutionary time roughly estimates engineering difficulty, at that pace the remaining 350 million years of our ancestry could be paralleled in robots in about 35 years.

Better yet, sufficiently useful robots don't need full human-scale brainpower. Commercial and research experiences convince me that mental power like a small guppy, about 1,000 MIPS, will suffice to guide mobile utility robots reliably through unfamiliar surroundings, suiting them for jobs in hundreds of thousands of industrial locations and eventually hundreds of millions of homes. Such machines are less than a decade away, but have been elusive so long that only a few dozen small research groups pursue them.

One Track Minds

Industrial mobile robots first appeared in 1954. In that year a driverless electric cart made by Barrett Electronics Corporation began pulling loads around a South Carolina grocery warehouse. Such machines, dubbed AGVs (Automatic Guided Vehicles) since the 1980s, originally, and still commonly, navigate by following signal-emitting wires entrenched in concrete floors. AGVs range from very small, carrying a few tens of kilograms, to very large, transporting many tons. Built for specific tasks, they often are equipped with specialized loading and unloading mechanisms like forks and lifters. In the 1980s, AGVs acquired microprocessor controllers allowing more complex behavior than afforded by simple electronic controls. New navigation techniques emerged. One uses wheel rotations to approximately track vehicle position, correcting for drift by sensing the passage of checkerboard floor tiles or magnets embedded along the path. In the 1990s a method became popular that triangulates a vehicle's position by sighting three or more retroreflectors mounted on walls and pillars with a scanning laser on the vehicle.

In five decades about a hundred thousand self-guided vehicles have found work in industry worldwide, but lighter "service robots" have yet to match even that modest success. These are intended for human-service tasks like delivery of mail in offices, linens and food in hospitals, floor cleaning, lawn mowing and guard duty. The most successful service robot to date is the Bell and Howell Mailmobile, which follows a transparent ultraviolet-fluorescent track spray-painted on office floors. About 3,000 have sold since the late 1970s. A few dozen small AGVs from several manufacturers have been adapted to transport linens or food trays along hospital corridors. In the 1980s several small US companies were formed to exploit suddenly-available microprocessors to develop small transport, floor-cleaning and security robots that navigated by sonar, beacons, reflectors and clever programming. The units were expensive, often costing

over $50,000, and required expert installation. No company managed to sell more than a few dozen a decade, and all slowly expired.

Larger AGVs and service robots must today follow carefully prearranged routes, greatly limiting their uses. Emerging techniques, utilizing increased computer power, are poised to loosen that restriction by letting the robot do the routing, surely greatly expanding the market. Customers will be able, unassisted, to put a robot to work where needed, enabling casual transport, floor cleaning and other mundane tasks that cannot bear the cost, time and uncertainty of expert installation. Though much freer in their wanderings, these new robots will (must) retain the reliability of their tracked predecessors. In my experience, customers routinely rejected transport and security robots that, after a month of flawless operation, wedged themselves in corners, wandered away lost, rolled over employees' feet or endangered themselves on stairs. Six months of successful operation, however, earned a sick day.

Sense of Space

Experimental robots that chart their own routes emerged from laboratories worldwide in the mid 1990s, as microprocessors reached 100 MIPS. Most built two-dimensional maps from sonar arrays to locate and route themselves, and the best were able to navigate office hallways for days between confusions. Those using sonar sensors fell far short of the six-month commercial criterion. Too often different locations in coarse 2D maps resemble one another, or the same location, scanned at different heights, looks different, or small obstacles or awkward protrusions are overlooked. A scanning laser sensor from German company Sick, that scans 180 degrees in quarter degree steps, and gives reliable ranges out to several tens of meters with centimeter accuracy, greatly improved 2D mapping performance in the early 2000s. Many experimental mobile robot now sport one or more Sick scanners (blue, yellow or white, with conical scanning window, resembling a compact coffee maker) and some seem to travel reliably. Commercial applications are emerging. Siemens offers a navigation package incorporating a Sick scanner for mapping and multiple sonar units for obstacle detection. It has been incorporated into a floor cleaning machine from Hefter that cleans the interior of an area after a human guides it around the perimeter. An AGV from Swisslog follows a "virtual guidepath" defined by scanner-sensed wall outlines.

Sick 2D scanning laser rangefinders are providing a first solution to the problem of freely navigating robots, but they're unlikely to be the final word. The maps are 2D, oblivious to hazards or opportunities above or below the scanning plane. The lasers are complex, precision electro-opto-mechanical devices, that emit a powerful infrared beam, and their price is likely to fall only slowly from its current $5,000 a unit. For over thirty years, I've worked towards practical 3D perception for robots from a variety of sensors, including inexpensive ones, to enable not only very reliable navigation but also abilities like object recognition. In the 1980s my lab devised a way to distill large amounts of noisy sensor data into reliable maps by accumulating statistical evidence of emptiness or occupancy in each cell of a grid representing the surroundings. The approach worked well in 2D, and guided many sonar-equipped robots in the 1990s. Three-dimensional maps, a thousand times richer, promised to be even better, but for years seemed computationally out of reach. In 1992 we found economies of scale and other tricks that reduced 3D grid costs a hundredfold, and by 1996 demonstrated a test program that accumulated thousands of measurements from stereoscopic camera glimpses to map a room's volume down to centimeter-scale. With 1,000 MIPS, now common in PCs, the program digests over a glimpse per second, adequate for slow indoor travel. The program was further developed from 1999 to 2003 with DARPA support, greatly improving its quality. A key addition was a learning process that tunes the sensor models by which stereoscopic (and other) readings are interpreted. Multiple camera images of the actual scene are projected from corresponding positions into trial 3D grids produced using particular sensor model settings: in good maps the occupied cells correspond to things in the physical scene, and receive similar colors form the multiple views, so average color variance per cell is low. The learning process tunes the sensor model in the direction of decreasing average color variance. The latest results are very good, as can bee seen from Figure 2, a 3D map constructed entirely from panoramic

stereoscopic images obtained in a single traverse down the center of the L-shaped hallway. These results, and prior experience in navigating from poorer 3D and 2D data, convinced me that, after thirty years, the techniques were finally ready for commercial development.

In February 2003 we founded Seegrid Corporation to to the job. Our first product will be a light duty self-navigating cart which a customer installs by pushing it once through a facility, stopping at important locations to add their names to a menu. Once trained, the cart can be loaded and directed to any menu destination. It will drive to the location, stop, and wait to be unloaded, ready for the next trip. As it travels, it observes and records its surroundings in rich 3D, plans safe routes, and localizes its position relative to a map from its training tour that is incrementally extended and updated in subsequent trips. We anticipate collaborations that apply the techniques to commercial floor cleaning machines, allowing them to map and select their own cleaning routes for indicated rooms or corridors. A custodian would supervise a flock of such semi-autonomous cleaners. We are also seeking an early entry to security robots that patrol warehouses and other large facilities detecting intrusions. We expect to expand these applications with routines that scan the 3D maps to recognize large features like walls, doors, corridors and rooms, and smaller objects including furniture and people. The hardware cost, for processors, cameras and other sensors, is several thousand dollars in the short run, but the component costs are falling at a rate that will bring the system into consumer price range within five to ten years. Imagine small, patient and competent robot vacuum cleaners that automatically learn their way around a home, explore unoccupied rooms and clean when everyone is away, recharging and emptying their dust loads at a docking station.

Fast Replay

Commercial success will provoke competition and accelerate investment in manufacturing, engineering and research. Vacuuming robots should beget smarter cleaning robots with dusting, scrubbing and picking-up arms, followed by larger multifunction utility robots with stronger, more dexterous arms and better sensors. Programs will be written to make such machines pick up clutter, store, retrieve and deliver things, take inventory, guard homes, open doors, mow lawns, play games and on. New applications will expand the market and spur further advancements, when robots fall short in acuity, precision, strength, reach, dexterity, skill or processing power. Capability, numbers sold, engineering and manufacturing quality, and cost effectiveness will increase in a mutually reinforcing spiral. Perhaps as by 2020 the process will have produced the first broadly competent "universal robots," as big as people but with lizardlike 10,000 MIPS minds that can be programmed for almost any simple chore.

Like competent but instinct-ruled reptiles, first-generation universal robots will handle only contingencies explicitly covered in their current application programs. Unable to adapt to changing circumstances, they will often perform inefficiently or not at all. Still, so much physical work awaits them in businesses, streets, fields and homes that robotics could begin to overtake pure information technology commercially.

A second generation of universal robot with a mouselike 300,000 MIPS will adapt as the first generation does not, and even be trainable. Besides application programs, the robots would host a suite of software "conditioning modules" that generate positive and negative reinforcement signals in predefined circumstances. Application programs would have alternatives for every step small and large (grip under/over hand, work in/out doors). As jobs are repeated, alternatives that had resulted in positive reinforcement will be favored, those with negative outcomes shunned. With a well-designed conditioning suite (e.g. positive for doing a job fast, keeping the batteries charged, negative for breaking or hitting something) a second-generation robot will slowly learn to work increasingly well.

A monkeylike 10 million MIPS will permit a third generation of robots to learn very quickly from mental rehearsals in simulations that model physical, cultural and psychological factors. Physical properties include shape, weight, strength, texture and appearance of things and how to handle them. Cultural aspects include a

thing's name, value, proper location and purpose. Psychological factors, applied to humans and other robots, include goals, beliefs, feelings and preferences. Developing the simulators will be a huge undertaking involving thousands of programmers and experience-gathering robots. The simulation would track external events, and tune its models to keep them faithful to reality. It should let a robot learn a skill by imitation, and afford a kind of consciousness. Asked why there are candles on the table, a third generation robot might consult its simulation of house, owner and self to honestly reply that it put them there because its owner likes candlelit dinners and it likes to please its owner. Further queries would elicit more details about a simple inner mental life concerned only with concrete situations and people in its work area.

Fourth-generation universal robots with a humanlike 300 million MIPS will be able to abstract and generalize. The first ever AI programs reasoned abstractly almost as well as people, albeit in very narrow domains, and many existing expert systems outperform us. But the symbols these programs manipulate are meaningless unless interpreted by humans. For instance, a medical diagnosis program needs a human practitioner to enter a patient's symptoms, and to implement a recommended therapy. Not so a third-generation robot, whose simulator provides a two-way conduit between symbolic descriptions and physical reality. Fourth-generation machines result from melding powerful reasoning programs to third-generation machines. They may reason about everyday actions by referring to their simulators like Herbert Gelernter's 1959 geometry theorem prover examined analytic-geometry "diagrams" to check special-case examples before trying to prove general geometric statements. Properly educated, the resulting robots are likely to become intellectually formidable.