Seeing Clearly and Moving Forward

Robert Laddaga, Massachusetts Institute of Technology
Mark L. Swinson, Defense Advanced Research Projects Agency
Paul Robertson, Oxford University and Dynamic Object Language Labs

AI RESEARCH IN THE NEAR FUTURE WILL TEND TOWARD DEVELOPMENT OF THE ABILITY TO PERCEIVE AND AFFECT THE WORLD THROUGH AUTOMATION, ROBOTICS, SPEECH, AND VISION—ALL ENHANCED BY SELF-AWARE AND CONTEXT-AWARE EMBEDDED SOFTWARE.

THIS SURVEY HIGHLIGHTS SOME important trends in AI research and development, focusing on perceiving and affecting the real world. We will primarily address robotics, but we certainly don't intend to imply that this is the only important area of AI research and development in the 21st century. We see tremendous importance in the continuing work in such areas as data mining, learning, knowledge representation, planning and scheduling, natural language understanding, expert systems, and deductive and inductive reasoning. Many of these areas will contribute substantially to the computers that see, hear, speak, and move, and that will be required for advances in ubiquitous, embedded computation. At the close of the article, we single out for special mention one area that contributes centrally to all of these technologies, software development technology.

Ubiquitous computing

We are moving swiftly toward a world of ubiquitous computing. Each year we produce more than one processor chip for each person on the planet, and the growth rate of chip production exceeds the population growth rate. So, by some definition, ubiquitous computing is surely upon us. Although for some people, ubiquitous computing means that we all carry devices that enable us to connect to the Internet wherever we happen to be, and although it does seem that this will happen, that is not what we mean by ubiquitous computing.

Our vision of ubiquitous computing is that devices under our control will heavily populate our environment, and that these devices will interconnect and interface with each other and with us. Many of the devices under computer control will be sensors that provide a window from the world of interconnected computation into the physical world around us. Today, cars are run by microcontrollers that can sense acceleration, engine performance, braking action, and so forth. Tomorrow's cars will have even more processors—connected with each other; sensing ambient temperature, weather conditions, other cars, and roadside facilities; and communicating with other cars, traffic controllers, roadside facilities, and us.

The major importance of ubiquitous computing will be the ability of computationally empowered devices to sense the world around us and to respond by changing that world. We will need these devices to take direction, explain their behavior, and modify their behavior to our directions. They won't need to communicate with us all the time, but they will always need to communicate when they are unsure or we wish to override them. Needless to say, it won't work (for a variety of reasons) to communicate with these devices with a keyboard on our person or dangling from the device. Our machines will need to use vision and auditory capability to see and hear what is going on in the background, and to understand and generate speech to engage in dialogue with humans.

Speech, vision, and language. Our view of ubiquitous computing makes clear that moving ahead in the sensor–affector arena is not sufficient; our computers must be capable of dialogue and able to understand both command and context. We understand our world and communicate with each other by seeing, hearing, speaking, and gesturing. If our ubiquitous computing environments are not to enslave us, they will need to communicate with us in the same manner and understand the context in which such dialogues occur.

We have made enormous progress in speech production but still have a way to go in introducing natural-sounding prosodics into generated speech. Similarly, we have made huge advances in the ability to present information in a visually compelling fashion but still have a long way to go with computer perception.

Visual and speech recognition are not only conceptual problems but also problems of computing in an embedded system context. Control of sensors such as cameras and microphones is a significant problem in itself, and the need to coordinate and fuse information from disparate sensors only exacerbates that control problem, although it offers hope of greater progress on the recognition side.

Generalized language understanding is an even more difficult technical problem. We have made significant progress with limited contexts and grammars, but general language-recognition capability might be even more elusive than general visual recognition. Computer vision started out as a core component of AI research, just as natural-language understanding did. In both cases, the fields have drifted away from the central theme of AI. Natural-language understanding has moved toward linguistics and computer vision toward image analysis. Both are huge and difficult undertakings, and in some sense, it is natural and necessary for them to have their own dedicated focus of research. They are also compelling problem areas for AI, in that they deal with the problem of making sense of the real world. We expect to see these disciplines come back into focus in the 21st century as core problems for AI research.

> *THE MAJOR IMPORTANCE OF UBIQUITOUS COMPUTING WILL BE THE ABILITY OF COMPUTATIONALLY EMPOWERED DEVICES TO SENSE THE WORLD AROUND US AND TO RESPOND BY CHANGING THAT WORLD.*

Real-world applicability. AI researchers have always desired to simplify problems. In the early days, they did this by building toy world problems. Such problems rightly fell out of favor when it became clear that solutions to them wouldn't scale up to solving real problems. For example, the early work on understanding images attempted to make sense of images consisting of blocks.[1–3] Rules for line-junction interpretation let us understand the lines in the images as 3D blocks. Unfortunately, the lines were not always visible in the images unless specially enhanced. Not all boundary lines are visible—and, of course, not all objects in the world are blocks. As analytically interesting as these pieces of work were and as influential as they were in helping to understand issues such as propagation of constraints, they would not lead directly to a solution to the computer-vision problem.

Since then, decades of low-level vision research have yielded many important algorithms for solving important parts of the computer-vision problem that can be applied to non-toy problem areas. Examples include algorithms for computing stereo disparities, extracting structure from motion, and analyzing and representing the rich textures found in the real world.[4–6]

In spite of the wealth of contributions in these areas and in low-level vision in general, advances in high-level visual interpretation have been slow to develop. Success stories in computer vision are few and far between. Computer vision that works is pretty much restricted to problem domains that have been carefully constrained. That is not to say that such accomplishments are not worthwhile. Applications in industrial inspection, medical vision, satellite image analysis, and most recently, visual information retrieval are all exciting constrained applications of computer vision on real images that have practical value unlike the early toy worlds.[7,8] Not only do they solve important problems in their restricted domains, but they also provide a focus for low-level vision research.

For example, focusing on the needs of robot vision has led to active vision.[9,10] The human visual system has some rather special capabilities, such as face recognition. Face identification and recognition are invaluable for many computer-vision problems.[11] Model-based vision succeeded in applications from medical image analysis to robotic vision, but constraint has been the key to success to date, and in that sense, we have failed to get away from toy worlds. The images used in these applications are certainly real, but the highly controlled nature of the problem domains is not.

The lure of the toy world or the artificially constrained world is that, by carefully controlling the complexity of the environment in which our programs must operate, we can construct complex algorithms that perform reasonably well. This is true, not just for computer vision, but also for speech recognition and even robot planning. Dealing with the complexity of the real world surely constitutes an important part of intelligence. At one end of the spectrum, by carefully controlling complexity, we can model the world with sufficient accuracy to allow complex algorithms to perform well. At the other end of the spectrum, illustrated by subsumption, we can build systems that have no model of the world but can *react* to it and thereby achieve a minimum level of competence.[12] Somewhere in between is intelligent, reasoned interaction with the world.

The 21st century will bring a proliferation of devices with embedded cameras. Some of these cameras will be embedded in robots, some in buildings, and some in vehicles. These cameras will often be deployed in environments that we cannot control. The challenge for AI will be to take what we have learned about low-level vision and design visual interpretation systems that can reason about the visual interpretation task at hand and the world in which it is operating to deliver robust visual interpretations. The new systems will have to *respond* to the world, not just react to it. Model induction will surpass hand-tailored models of the world, as the unrestricted world represents a large challenge compared to the restricted domains in which we have largely operated to date. Intelligent data-fusion will grow in importance as devices that support multiple sensory modalities proliferate. Many mobile robots already include a variety of sensors such as sonar, laser scanners, and cameras. Although researchers have worked much on interpreting the world through each of these sensors, the problem of building and maintaining a coherent world model by using the contributions of a variety of sensors is still largely unsolved. With so much momentum behind low-level vision, the availability of inexpensive cameras, and the ability to build in real-time support for many low-level vision routines using digital signal processors embedded within the devices, computer vision will become a central AI problem again.

Robotic beasts and machines

Many of our machines will have forms dictated by their tasks, environments, or simply our imaginations. We can expect to see robotic insects, dogs, and birds, as well as robotic cars, boats, and perhaps even washing machines. Robotics might represent the greatest unmet technological expectation of the 20th century, but many of those expectations will likely be realized early in the 21st. Critical to this progress are advances in AI technologies, such as situated machine learning.[13]

Increasing complexity. Reinforcement learning, for example, can serve for an unsupervised, learning-with-critic approach, in which mappings from precepts to actions are learned inductively through trial-and-error. Other approaches include evolutionary methods that begin with an initial pool of program elements and use genetic operators such as recombination and mutation to generate successive generations of increasingly successful controllers. By using these approaches, robots might learn by adjusting parameters, exploiting patterns, evolving rule sets, generating behaviors (and aggregations of behaviors), devising new strategies, predicting changes in the environment, and even exchanging this knowledge with other robots. Such robots can acquire new knowledge and adapt existing knowledge to new circumstances, and thereby solve problems in ways humans might not understand. Indeed, rather than something to be suppressed by careful design, emergent behavior might instead be encouraged by equally careful design.[14]

LIKE AUTOMOBILES, AS ROBOTS BECOME MORE PERVASIVE, THEY ARE LIKELY TO BECOME INCREASINGLY COMPLEX.

Like automobiles, as robots become more pervasive, they are likely to become increasingly complex. Some robots might be comprised of millions of parts. Fast, cheap, rapid manufacture of these robots might necessitate removal of humans from the process altogether. Jordan Pollack and his colleagues at Brandeis are using commercial CADCAM simulators together with a genetic algorithm to evolve the body and brains of simple robots.[15] They have succeeded in automatically designing, improving, and creating a real robot with only trivial human intervention. So far, the work has focused solely on creating a robot for locomotion, but eventually, this approach might allow cheap, near-perfect solutions to be evolved and deployed, even for complex tasks requiring unintuitive solutions.

Humanoid robots. The humanoid robotics field includes a rich diversity of projects in which perception, processing, and action are embodied in a recognizably anthropomorphic form to emulate some subset of the physical, cognitive, and social dimensions of the human experience to create a new kind of tool. Such a tool would work, not just *for* humans, but *with* them. Humanoids will be able to work safely alongside humans in everyday environments, as well as the more daunting space and undersea environments—thereby extending our capabilities in ways we cannot foresee.[16]

Indeed, humanoids might prove to be the ideal robot design for interacting with people. After all, humans naturally tend to interact with other human-like entities; the interface might well be hardwired in our brains.[17] Human-like bodies will let these robots blend seamlessly into environments already designed for humans. Although we humans have historically adapted to the limitations of our machines, here the machines will adapt to us. Humanoids will provide not only a new way for us to interact with machines, but an intuitive filter for humans to interact with an increasingly ubiquitous and pervasive information environment.

Humanoid robots that can acquire new knowledge incrementally from autonomous interactions with the environment will accomplish tasks by means their designers did not explicitly implement (or perhaps even conceive of), and will perhaps thereby be capable of adapting to the unanticipated circumstances of an unstructured, dynamic environment. Humanoid robots have already demonstrated basic task decomposition necessary to carry out complex commands given through gesture and speech.[18] They have also demonstrated the ability to adapt, orchestrate existing capabilities, and create new behaviors using a variety of machine learning techniques. As they adapt to their own unique experiences in the world, we might well see a population in which no two humanoid robots are exactly alike.

Humanoid robots might rekindle a new inspiration for AI as they motivate research toward intelligent, autonomous systems. Already, a growing number of robotics researchers have found that the human form provides an excellent platform for real-world learning. A similar body facilitates learning based on imitation by making it easier to map the human's action onto the robot.[19] In fact, it might be that human-like intelligence actually requires a human-like body.[20] At the very least, the anthropomorphism of these robots enables them to interface easily with existing technology and infrastructure with minimal disruption. Hence, humanoids appear to be a uniquely appropriate form with which to gradually introduce intelligent robots into new application domains.

Embedded and aware software

Advancing the embedded system agenda discussed earlier will require the introduction of AI technology in the process of developing and deploying software. Those same technologies will also advance the general AI agenda. Let's look briefly at these crucial interconnections.

There are two looming problems that software development researchers will need to address in the near future. One is the problem of software robustness in general. The second is the ability to program and maintain complex embedded systems—that is, to create and manage the software that powers physically instantiated systems, devices, and machines. Robotic devices, such as those we've discussed, are just one example of such complex, physically instantiated systems. Cars, homes, offices, factories, and medical facilities are more such examples. Lack of software robustness is already a significant problem, but it will become an even more crucial problem as computing becomes both ubiquitous and deeply physically embedded.

Robustness is more than simply conforming without bugs to a proper static specification. In this new world of ubiquitous computing, robustness will also mean that software will be responsive to environmental conditions and changing human requirements. For software to have these adaptive and embedded properties, it will need some understanding of its physical and information environment, including its users. It will also need understanding of its own makeup, including goals, designs, structure, and the functional and other properties of its modules or components.

The best way to incorporate both physical and structural properties is with models.[21] Complex modeling of physical processes, as well as of user needs and behaviors and of software structure and operation, is a significant task for the AI community. As we have said, such modeling will be necessary for the development of robust embedded systems. We will also need to bring many more technologies to bear, such as sophisticated understanding and dynamic modification of program structure and technology to treat software-based systems as closed loop control systems.[22,23] Agent technologies will also play an important role.[24]

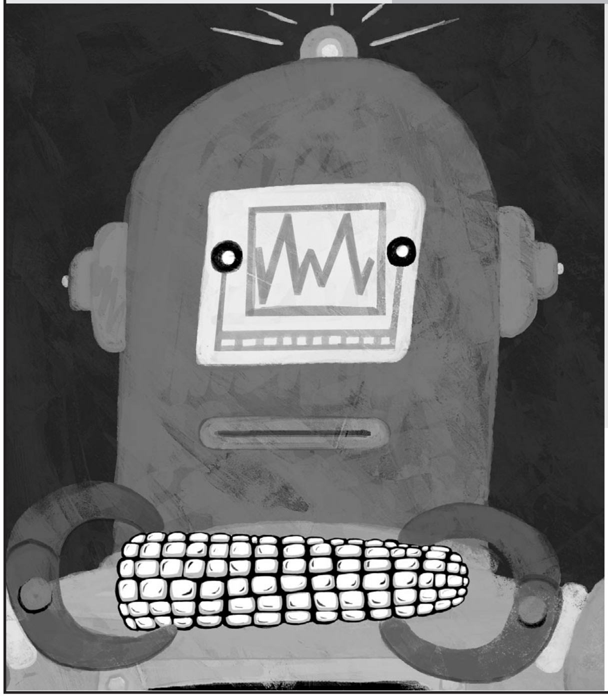
Many AI technologies will be crucial to the development of self-adaptive, physically embedded, software-integrated systems. Just as importantly, however, such software will play an important role in advancing AI technology. Self-aware and context-aware embedded software will be able to adapt, explain its behavior, request help, and respond to direction. The ability to produce such behaviors will establish self-aware and context-aware software as a crucial building block of artificially intelligent systems, as well as of the endeavor of AI itself.

THE DRIVE TOWARD PERVASIVE COMPUTATION IS CREATING A TREMENDOUS NEED FOR INTELLIGENT SYSTEMS THAT CAN PERCEIVE AND ACT AS WELL AS REASON.

THE DRIVE TOWARD PERVASIVE computation is creating a tremendous need for intelligent systems that can perceive and act as well as reason. AI is responding with work on speech recognition, vision automation, and robotics. Of special note are efforts in automating interpretation of visual data and building and automating human-form robots. The physically embedded nature of such intelligent systems requires new and better software tools for software with awareness of itself and its surroundings. AI researchers will have to provide self- and context-aware software, which will play a central role in AI research beyond the support it provides for intelligent perception and action.

References

1. D.L. Waltz, "Understanding Line Drawings of Scenes with Shadows in P.H. Winston," *Psychology of Computer Vision*, McGraw-Hill, New York, 1975, pp. 19–91.
2. M. Clowes, "On Seeing Things," *Artificial Intelligence*, vol. 2, no. 1, 1971, pp. 79–116.
3. D. Huffman, "Impossible Objects as Nonsense Sentences," *Machine Intelligence*, vol. 6, 1971, pp. 295–323.
4. D. Marr and T. Poggio, "A Computational Theory of Human Stereo Disparity," *Science*, vol. 194, no. 4262, 1979, pp. 283–287.
5. E.C. Hildreth, "Computations Underlying the Measurement of Visual Motion," *Artificial Intelligence*, vol. 23, no. 3, 1984, pp. 309–354.
6. S. Geman and D. Geman, "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 6, no. 6, 1984, pp. 721–741.
7. M. Nagao and T. Matsuyama, *A Structural Analysis of Complex Aerial Photographs*, Plenum Press, New York, 1980.
8. P. Robertson, "A Corpus-Based Approach to the Interpretation of Aerial Images," *Proc.7th IEEE Conf. Image Processing and its Applications*, IEEE CS Press, Los Alamitos, Calif., 1999.
9. A. Blake and A. Yuille, *Active Vision*, MIT Press, Cambridge, Mass., 1993.
10. M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active Contour Models," *Proc. 1st Int'l Conf. Comp. Vision*, IEEE Press, Piscataway, N.J., 1987, pp. 259–268.
11. T.F. Cootes, G.J. Edwards, and C.J. Taylor, "Active Appearance Models," H. Burkhardt and B. Neumann, eds., *Proc. European Conf. Computer Vision*, Springer-Verlag, Berlin, 1998, pp. 484-498.
12. R.A. Brooks, "Intelligence without Representation," *Artificial Intelligence*, vol. 47, pp. 139–159.
13. M. Swinson and D. Bruemmer, "Expanding Frontiers of Humanoid Robotics," *IEEE Intelligent Systems*, vol. 15, no. 4, July/Aug. 2000, pp. 12–17.
14. B. Scassellati, "Theory of Mind for a Humanoid Robot," to be published in *Proc. Humanoids 2000*, IEEE Press, Piscataway, N.J.
15. J. Pollack and H. Lipson, "Automatic Design and Manufacture of Robotic Lifeforms," *Nature*, vol. 406, no. 6799, 2000, pp. 974–978.
16. R.O. Ambrose et al., "Robonaut: NASA's Space Humanoid," *IEEE Intelligent Systems*, vol. 15, no. 4, July/Aug. 2000, pp. 57–63.



17. C. Breazeal and B. Scassellati, “How to Build Robots that Make Friends and Influence People,” *Proc. Int’l Conf. Intelligent Robots and Systems*, IEEE Press, Piscataway, N.J., 1999.

18. B. Adams et al., “Humanoid Robots: A New Kind of Tool,” *IEEE Intelligent Systems*, vol. 15, no. 4, July/Aug. 2000, pp. 25–31.

19. M. Mataric, “Getting Humanoids to Move and Imitate,” *IEEE Intelligent Systems*, vol. 15, no. 4, July/Aug. 2000, pp. 18–24.

20. R.A. Brooks et al., “The Cog Project: Building a Humanoid Robot,” *Computation for Metaphors, Analogy, and Agents: Springer Lecture Notes in Artificial Intelligence*, vol. 1562, Springer-Verlag, Berlin, 1998.

21. G. Karsai and J. Sztipanovits “A Model-Based Approach to Self-Adaptive Software,” *IEEE Intelligent Systems*, vol. 14, no. 3, May/June 1999, pp. 46–53.

22. P.M. Oreizy et al., “An Architecture-Based Approach to Self-Adaptive Software,” *IEEE Intelligent Systems*, vol. 14, no. 3, May/June 1999, pp. 54–62.

23. M. Kokar, K. Baclawski, and Y.A. Eracar, “Control Theory-Based Foundations of Self-Controlling Software,” *IEEE Intelligent Systems*, vol. 14, no. 3, May/June 1999, pp. 37–45.

24. K.R. Dixon, T.Q. Pham, and P.K. Khosla, “Port-Based Adaptable Agent Architecture,” to be published in *Self-Adaptive Software*, P. Robertson, R. Laddaga, and H.E. Shrobe, eds., Springer-Verlag, New York.

Robert Laddaga is a research scientist in the Artificial Intelligence Laboratory at the Massachusetts Institute of Technology. His research interests include intelligent systems and software, self-adaptive software, and semantically based collaboration. He served as a program manager in the DARPA Information Technology Office, where he assisted such programs as EDCS and Information Survivability and helped create the Active Software Composition and Autonomous Negotiating Teams programs. He has a BS in mathematics and an MA in philosophy from the University of South Carolina and a PhD in philosophy from Stanford University. Contact him at 200 Technology Square, NE43-804, Cambridge, MA 02139; rladdaga@ai.mit.edu.

Mark L. Swinson, a retired colonel in the US Army, is the deputy director of DARPA’s Information Technology Office. He is also the program manager for the Embedded Systems Program, the Mobile Autonomous Robot Software Program, and the Software for Distributed Robotics Program. His research interests include embedded software, distributed processing, domain-specific languages, and machine learning for robot programming. He has a BS in engineering from the US Military Academy at West Point, an MS in mechanical engineering from the University of Wisconsin, and a PhD in robot control systems from the University in Florida. He is a member of the IEEE, the American Society of Mechanical Engineers, the American Society for Engineering Education, and the Association for Unmanned Vehicle Systems International. Contact him at the Defense Advanced Research Projects Agency, Information Technology Office, 3701 Fairfax Dr., Arlington, VA 22203-1714; mswinson@darpa.mil; www.darpa.mil/ito/Personnel/mswinson.html.

Paul Robertson is president of Dynamic Object Language Labs and a visiting research scientist with the robotics research group at Oxford University. His research interests include self-adaptive software, computer vision, and architectures for artificial intelligence. He has 20 years’ experience in commercial applications of AI and in building tools and languages for building complex systems. He is completing a DPhil in image analysis at Oxford University and holds a BA in computer science. Contact him at DOLL, Inc. 9 Bartlet St. #334, Andover, MA 01810, probertson@doll.com.