

# Is Computer Vision Still AI?

Robert B. Fisher

■ Recent general AI conferences show a decline in both the number and the quality of vision papers, but there is tremendous growth in, and specialization of, computer vision conferences. Hence, one might conclude that computer vision is parting or has parted company with AI. This article proposes that the divorce of computer vision and AI suggested here is actually an open marriage: Although computer vision is developing through its own research agenda, there are many shared areas of interest, and many of the key goals, assumptions, and characteristics of computer vision are also clearly found in AI.

It is not easy to infer the relationship between the fields of computer vision and AI from their external appearances. Recent general AI conferences show a decline in both the number and the quality of vision papers, but there is tremendous growth in, and specialization of, computer vision conferences. Some computer vision or robotics researchers even claim that AI is unnecessary or irrelevant; special-purpose, dedicated, well-engineered, mathematics-based processes will lead to success.

Alternatively, we could consider the interests and directions of the two fields: An examination of many recent computer vision conferences and journals shows a marked inclination, especially in the more theoretical papers, toward complex mathematics (for example, geometric invariance, differential geometry, functional analysis, control theory); models of the physics of light, color, shape, motion appearance, texture, and so on; statistical models of the scene and other properties (for example, fractal, Markov random fields,

Bayesian); and nonsymbolic image-to-image transformations. Most successful practical vision systems are well-engineered combinations of special-purpose sensors, hardware, and algorithms that are tailored to solve specific visual problems (and are generally more successful in proportion to the narrowness of the task). Some texts (for example, Batchelor, Hill, and Hodgson [1985]) don't even mention AI or its methods. Example practical applications include assembly-line robot welding guidance, inspection of computer keyboards and integrated circuits, counting of tea bags, tomato grading, and optical tracking.

We can contrast computer vision's rather specific research directions with the outward appearance of AI,

---

*... the differences are  
largely illusory ...  
computer vision still has  
and will continue to have  
a strong relationship  
with AI.*

---

which has spent much time investigating more general methods, such as search controlled by domain-dependent constraints, models of (formal) logical and nonlogical reasoning, heuristic and uncertain reasoning, representation of general physical and world knowledge (for example, surface shape, elasticity, gravity) and human knowledge (for example, conceptual, belief), and learning (that is, inference of new relationships, self-organization).

Hence, one might conclude that computer vision is parting or has parted company with AI. This phenomenon is a more general problem of AI, and most subfields are specializing to the point that wasteful duplication is occurring, and loss of sight of the grand goal of reintegrating the different aspects of intelligence is evident. In addition, although there are many specializations of AI (for example, planning, natural language understanding, knowledge representation), no one questions the separateness of these fields from AI. However, the technology of most computer vision research is so non-mainstream AI (that is, it looks more like manufacturing engineering or applied physics) that the connection is no longer obvious.

Phrased in these terms, the differences seem extreme; however, I claim the differences are largely illusory and that computer vision still has and will continue to have a strong relationship with AI. Computer vision's specialization and unique preoccupations define it as a field of study but no more exclude it from the community of AI than do the distinctive formalisms of natural language grammars, the logics underpinning theorem proving and formal reasoning, or the numeric calculus of neural networks (and its learning algorithms) exclude their subfields.

This article proposes that the divorce of computer vision and AI suggested here is actually an open marriage: Although computer vision is developing through its own research agenda, there are many shared areas of interest, and many of its key goals, assumptions, and characteristics are also clearly found in AI. I take several views on this question, not because of insecurity but because many strands of connection imply a close coupling in the space of cross-disciplinary linkages (compared with a sparser coupling [in the sense of few links between concepts] between computer vision and, for example, architecture [coupled mainly through issues of shape, appearance, and visual aesthetics]). I hope to convince you that the relation will become even stronger in the future.

## AI and Computer Vision Share Methodological Approaches

There are many eloquent and intelligent discussions of the nature of AI (Boden 1988; Haugeland 1985) (and if it can even exist [Penrose 1989]), but lately, most arguments seem to center on the how of AI, not the what, that is, whether true AI can be realized through classical methods (that is, logic, representation, and search), pattern classification (for example, case and frame unification), behaviorist methods (that is, an organized conglomeration of individual goal-pursuing, goal-achieving behaviors), classical emergent-behavior methods (that is, global, epiphenomenal behavior arising from the aggregation of distributed actions of separate agents), or neural networks (that is, distributed, fault-tolerant, connectionist, emergent, numeric computations).

Loosely speaking, computer vision mirrors the methodological division of AI: There are certainly classical symbol manipulation (Brooks 1981), statistical pattern classification, distributed competence, and connectionist (Hinton 1981) paradigms. In addition, there are at least two other paradigms. The first paradigm is the *numeric processing of images*, where algorithms are linked to the geometric structure of the image and are constrained by underlying theories of the physical processes that gave rise to the sense data (as in a theory of surface shading [Horn 1975]). The image-to-image transformations typical of this class of processes are not in themselves intelligent (that is, they usually only do a limited amount of interpretation, as in suggesting possible edges). However, they can be considered a product of AI (Schank 1991). The processes might originally have been intelligent—that is, occurred as a result of explicit rule-based reasoning—but later were compiled into pure algorithmic form, or they might never have been, but the actions of the algorithms can be controlled by higher-level control input (such as focus of attention).

The second paradigm is active per-

ception (Aloimonos 1989), wherein the observer manipulates or maneuvers within the environment to produce constrained—and, thus, more easily interpreted—perceptual effects, such as the kinetic-depth process.

Computer vision also encompasses a large number of special-purpose application processes dedicated to extracting a single, perhaps obscure piece of information from a specialized type of image (for example, the parametric shape of a range image surface, the average intercellular distance in a particular tissue section, or the postal code on a letter). These specializations do not fit clearly into the general vision architecture, just as an expert chess player is generally not considered a component of a general AI program. Even so, the class of transformation that these applications represent does not differ significantly from the class of competences expected to be found in a general vision system.

## AI and Computer Vision Share Domain Assumptions

Broadly speaking, the following assumptions underpin AI and computer vision perspectives. This list is undoubtedly incomplete, but I tried to elucidate beliefs that underlie most research in the general sense—not the individual research areas but, rather, the themes that they share. I introduce the key topics from an AI perspective and then discuss how computer vision fits within this perspective.

As can be seen, part of what influences the distinctiveness of computer vision is the nature of raw sense data, which is often underconstrained, is always shaped by sensor characteristics, and is corrupted by noise. It has a regular geometric structure (for example, images) that is closely linked to the geometric ordering in the sensed domain. However, although both have a regular geometry, there is not necessarily an isomorphism because, for example, two-dimensional images arise from the projection of the three-dimensional world. In fact, the relationship can be

obscure (as in a holographic image) or inexact (as in a cartoon sketch).

The first AI topic is knowledge representation.

**Representable knowledge:** Knowledge is representable, usable, and communicable, although the details of how it is are unresolved as yet. Computer vision programs embody knowledge (for example, about shape, reflectance, apparent structure) and increasingly make the knowledge explicit in terms of geometric-object model bases and rule bases of information on how to recognize objects or when to apply various operators.

**Representation schemas:** Symbols can be used to represent some concepts, and the manipulation of the symbols can be used for reasoning about a domain. Most high-level vision programs manipulate symbolic representations of model and data features, usually attempting some form of search for a correspondence between the two. Intermediate-level image-interpretation processes more often have numeric image data as input but produce symbolic descriptions as their output.

**Multiple descriptions:** The characterization of real domains (or environments) requires many different descriptors and points of view as well as multiple levels of representation. The computer vision community has believed for at least a decade that the description of the visual world requires multiple descriptions, particularly through the concept of sketches (Marr 1982) and intrinsic images (Barrow and Tenenbaum 1978). These descriptions are focused, special-purpose representations of the world from the viewer's perspective that describe how it is moving, how it is illuminated, where its significant features lie, and so on.

**Underlying theories:** Mathematical theories underlie some aspects of knowledge and some domains (for example, theory of mechanics, physics, dynamics of a steam engine). The theories can be used to model and test hypotheses, interpret sense data, or predict effects. Mathematical theories of surface shape, motion and image flow, geometry and geometric

invariance, image formation, noise processes, combinatorics, and physical theories of shading, mutual illumination, color, and texture underlie much of the successful recent vision research. However, these theories tend to be *autonomous*, that is, require no insight, feedback, or control from later or other stages of the vision system. Expressing the complexities of algorithms that allow these external interactions in mathematical form can be difficult.

**Common sense:** Large amounts of commonsense knowledge are required for intelligent and effective behavior in the real world. A machine that can see the world as we see it will require a visual memory of thousands of objects and fragmentary shapes. To exploit this memory, a computer vision system will need to be able to reason about the way in which appearance can vary by position and illumination, depth ordering, interaction between various objects in a scene, and the way that members of a class of shapes can appear and deform (for example, clothing, trees).

The second AI topic is reasoning.

**Mechanization:** Reasoning can be mechanized, modeled, replicated, and experimented with, particularly through the use of computer systems. Any working computer vision system exemplifies the mechanization of visual reasoning in that it is clearly an algorithm implemented on a computer system (in the broad sense, allowing dedicated or hard-wired electronic implementations) (Brooks 1991; Rosenschein and Kaelbling 1986).

**Complexity:** Reasoning is complex and can require nondeterministic decisions. Because of the difficulty of interpreting data (because of noise or low resolution or because local constraints exist to fully explain the data), most real image-interpretation systems (that is, systems that label image features) embody some form of expert system reasoner (Draper et al. 1988) that pursues alternative hypotheses and quantifies the degree of verification.

**Heuristics:** Heuristics are often required to (1) model incompletely

understood phenomena, (2) simplify computationally intractable algorithms, or (3) provide a simple and reliable tool when exact methods are unnecessary or expensive. Most object recognition is still in the heuristic stage, except for simple geometric solids (for example, polyhedra). Even most edge detectors are based on the heuristic that all edges are intensity step edges, and most uses for the edge information assume that image edges correspond to object edges (ignoring lighting, shadows, specularities, and changes in reflectance).

**Uncertainty:** Reasoning involves uncertainty by virtue of incomplete knowledge, perceptual noise, and imperfect heuristics. Uncertain reasoning is needed to quantify the belief in a sensor measurement and to characterize the certainty of a hypothesis (Durrant-Whyte 1987). Recently, much computer vision research has been using statistical methods to represent measurement and hypothesis uncertainty, particularly through the use of the Kalman filter.

The third AI topic is behavior.

**Humans:** Human and other animal behavior can be studied and is underpinned by intelligible but as yet unknown computational processes.

Results in visual neurophysiology from studies conducted over the last 40 years have shown that a large number of neural structures exist for the purpose of extracting visual information, and testable theories for some of these structures have been developed (Marr 1982).

Because we are only well informed about the complexity of human visual perception, we cannot be sure of the sophistication of other intelligent systems. However, there is no doubting the visual intelligence of humans when we observe our arts, particularly the abstract, surreal, or cartoon forms. Through these forms, we move from a literal description of shape and reflectance to a reductionist symbolic representation of the world expressed in functionally useful concepts to a cultural dialogue in which the form and content of the image (or sculpture) are as much a

# Back Issues of selected Issues of *AI* Magazine Are Still Available!

For a brochure listing issues available, subjects covered, and costs, send your name and address to:

AAAI  
445 Burgess Drive  
Menlo Park, CA 94025

info@aaai.org

response to current historical, social, and artistic context as a description of a possible reality. As for accessibility, we know a little about human perception from observers' reports and visual agnosia ("the inability to recognize objects even though elementary visual functions remain unimpaired" [Farah 1990]) studies; about other species, we know almost nothing. Although the early stages of animal vision, the knowledge of the active neurology in other species might be greater because of the infor-

mation obtained from live animal experiments.

**Complex behavior:** Complex, intelligent behavior is a consequence of the complexity of the domain in which the behavior occurs (for example, the richness of human experience and imagination generates linguistic complexity).

The complexity of visual behavior is reflected in the complexity of the programs that implement the behavior. Almost any vision system that does anything of consequence has thousands of lines of code, at a minimum. Systems that aspire to even limited degrees of general-purpose capability (Draper et al. 1988) have more like hundreds of thousands of lines of code and often involve the use of complex reasoning mechanisms, such as blackboards. Even extracting simple information from a real image (for example, counting cells in a microscope slide view) is complex because of the detail and variability of the real world (for example, different cell sizes, optical constraints, placements, debris, adjacencies, variations in boundary appearance and shape).

**Intelligent perception:** Intelligent perception requires integrating many different sources of information plus using knowledge about what is being perceived.

Intelligent vision, in particular, requires integrating many different sources of visual information, as generating a full understanding of complex scenes seems to require different representations of shape, position, color, motion, and so on. For example, although stereo gives information about distance to, and shape of, regions containing a lot of visual texture, extending the understanding to nearby nontextured regions needs other information, from, say, the shading. We can also actively use knowledge about the nature of the world, such as when we reason about how a pile of books that partially hides a water glass affects the appearance of the glass and how the optics of the light passing through the glass affects what we see through the glass itself.

**Multiple theories:** No single theo-

ry explains all intelligent behaviors, and an agent can use different behaviors, as is appropriate.

Computer vision is a perfect example of where multiple types of behavior are needed: Much low-level image work is data driven, but most high-level systems embody both data- and model- driven reasoning. Most low-level programs (that is, image-to-image processing) are numeric, but most high-level programs are purely symbolic. Classical symbol manipulation high-level vision systems can interface with neural network low-end modules. Active, or multicamera, vision systems can sometimes acquire data far more easily or reliably than passive monocular vision systems.

**Multiple skills:** Truly intelligent systems are capable of perception; communication; memory; learning; self-analysis; self-knowledge; decision making; acting; planning; attention focusing; and, undoubtedly, other skills (in varying degrees). These skills need to be integrated into a cohesive system to generally be useful.

In computer vision, perceptual results need to be encoded for use, the shape and appearance of objects must be known, and the shape and appearance of new objects must be learned. A vision system that can reason and then move to obtain a better viewpoint is more effective. Knowing what results can be trusted can guide when to proceed, and knowing how well a set of computer vision processes works when applied to different domains can guide the selection of which process to apply.

**Learning:** Learning is required because (1) the amount of knowledge available is too immense for explicit encoding by human designers, (2) domains change over time, (3) new concepts enter discourse, and (4) agents enter new domains.

Model-based vision needs models; these models are mostly constructed by humans and, for complex objects, rather slowly. Self-acquisition is the only possibility for acquiring the description of large numbers of objects that must be known by competent active agents. Visual descriptions must change over time as a consequence of seasons, aging, wear,

growth, or the encounter of new objects.

Some models might be learned using classical structural learning methods adapted for visual representations. However, experience with medical imaging suggests that some patterns (for example, biological) are too complex for compact description; hence, the communication of these patterns by other than example will nearly be impossible. Thus, much as a medical specialist learns to interpret a class of image data by repeated exposure over six months, so too will machine vision systems learn, perhaps in a connectionist manner.

## AI and Computer Vision Share Goals

Here I outline the goals of both AI and computer vision. The main goals of AI research are as follows:

First is to characterize intelligence and intelligent behavior in general. This characterization includes theories of the architectures for integrating the various intelligent skills into single autonomous agents and cooperating systems of agents.

Second is to understand human competence and computational processes, in part by providing a methodology for developing testable theories.

Third is to develop tools that need less human attention, embody greater capability and compiled experience, and extend human control over our environment (that is, *environment* in the sense of our social, biological, physical, and intellectual context). The intelligence of the tools is, in part, a consequence of their complexity and reactive flexibility; however, the real breakthrough is in the embedding of intent in the tool.

Fourth is to develop epistemologies suitable for representing different knowledge domains.

Fifth is to develop tools for advanced computer science and engineering (for example, distributed processing, programming languages, correctness proving, automatic programming).

Sixth is to extend the philosophical view of humanness to

reconcile the physical operation of the brain and its computational structure with the concept of the mind and the mental universe (for example, Dennett [1992]).

The goals of computer vision are as follows:

First is to understand the human and other biological vision systems through the building of testable models. This goal must also include understanding the purposes of vision (Gibson 1979; Sloman 1988), which consists of multiple information-extraction modules, providing visual control (for example, motion) feedback as well as information extraction.

Second is to provide machines that extend human perceptual abilities into new domains or heighten them in normal domains.

Third is to provide tools that embody autonomous information-extraction abilities (for example, need not employ a human).

Fourth is to determine the key representations needed for a visual and spatial description of the real world and to discover computational processes that can reliably infer them.

Hence, it can be seen that computer vision's goals are largely specializations of AI's goals. (Computer vision's goal 1 is a specialization of AI's goal 2, computer vision's goals 2 and 3 are specializations of AI's goal 3, and computer vision's goal 4 is a specialization of AI's goal 4.)

## AI and Computer Vision Share a Common Intellectual Context

We can also consider the relation of AI and computer vision from the perspective of three aspects of their shared intellectual context: (1) philosophical support, (2) biological and psychological support, and (3) shared computational methodology.

### Philosophical Support

No issue is more central to AI than world representation, and perception, through its direct external input from the world, provides the basis for constructing these first, basic representations. The philosophy of perception

and mind shows strong connections between AI and computer vision in the attempts to understand the relationship between a physical reality and our perceived understanding of this world. Fundamental to the perceptual view is that what we embody is not a literal description of the world as it is but an internalized and abstracted representation of the world encoded in terms that form part of the internal states of the agent. Without such an encoding, the description would be of no use; we would still need to analyze the new description to extract useful information (Sloman 1988). The full nature of the actual representations is unclear and is perhaps incompletely communicable between different humans. What is clear is that it is a representation, as opposed to the real thing and, hence, impoverished.

These abstractions are necessarily

piece is the apple. Considering some subset of the world as a distinct nameable entity is often simply a human convention and, thus, is suitable for AI. Of course, not all entities are defined purely by convention because biological and physical processes clearly also play a major role in physically defining the world.

Conceptual entities need not only be objects; they can also be actions (for example, when does a jog become a run or a sprint) and attributes (blue versus cyan), for example. Of course, names are linked to function; so, one physical object might be referenced by different names, according to the use that the object has (for example, as a bowl, cup, ashtray, paper clip holder). It is clear that one must also know a lot about human society and its conceptual structures to interpret the meaning of the visual input at any level

## *Computer vision will need to explore more case-based and opportunistic reasoning.*

interpretations. As sensors are tuned to sense particular modalities, the results they report are biased by what they expect to see. A sensor that detects green must decide what is or is not green, whereas the physical spectrum is dense, and the light coming off objects usually has components at all frequencies. This complication is particularly acute with the color brown, the perception of which is affected by the relative lightness of the surrounding colors. Hence, any conceptual representation of the world must have a relation to the world but is not the world.

AI has its strongest linkage with computer vision in those aspects of vision that are distinctly human. Perhaps the most distinct competence is the partitioning of the sensed environment into conceptual, named entities. For example, an apple on a table is a relatively distinct entity, yet when still attached to a tree, it is part of a larger entity. We decide which

deeper than a largely empirical, physical description.

### Biological and Psychological Support

Vision is often naively considered a monolithic sense, whereas actually, many different information-extraction and information-interpretation processes are involved. Neurophysiologists have identified separate regions of the visual cortex that appear to extract shape, motion, color, and edges from our visual input. Altogether, about 9 (including the retina) distinct regions of visual processing have been identified so far, many with several processing layers; so, something like 20+ different but as yet unknown representations of the world can be extracted. The neural connectivity of some regions in the early stages of processing is broadly known (Zeki 1993; Hubel 1988), but on the whole, their function is not. Some activities are fully autonomous

(for example, processing the full visual field input), whereas other activities involve input from other centers of the brain (for example, attention focusing). Output of visual processing are used in many places—helping maintain balance, tracking moving objects, or stabilizing our representation of the world while in motion; ducking or blinking when danger approaches; making local maps of our environment; discriminating between alternatives; and, the obvious, labeling the world.

From this list of visual functions, it is clear that many biological visual processes contribute to and define the behavior of intelligent beings. The linkage also goes in the reverse direction in that intelligent processes are needed to select what to attend to, to record and extract visual memories, to provide the motion needed for active vision (for example, through motion parallax), to do spatial reasoning, to make visual aesthetic judgments, and so on.

### Shared Computational Methodology

The foundation behind the experimental methodology of both AI and computer vision is the computer. However, many fields use computers, so we must look deeper to the more fundamental notion of the computer as a tool for theory testing, a tool for empirical explorations, and a vehicle for the embodiment of theories to create usable artifacts (compared to more conventional numeric calculating, text processing, and database-transaction machines). Although theories of mechanics or energy flow, for example, lie behind the function of other machines, the theory of computational information processing lies behind machines built for AI and computer vision (Marr 1982). This theory entails internal representations; reasoning based on, and transformations between, representations; constructions of these representations; and actions based on them.

The use of the computer is so pervasive in AI and computer vision (and now in most other fields) that it is taken for granted. More important are the many shared tools and tech-

niques that exploit the computer capabilities to achieve both AI and computer vision. A nonexhaustive list must include frame-representation techniques (Brooks 1981), expert system machinery (Matsuyama 1990), probabilistic and uncertain reasoning, search tree exploration (Grimson 1990), generate-and-test algorithms, constraint-satisfaction systems (Brooks 1981; Waltz 1975), hierarchical representations and reasoning methods (Fisher 1989), symbolic and neural network (Hinton 1981) perspectives, and the embedding of much domain-specific knowledge. These shared techniques are used mostly for what is loosely called *high-level vision*, which is preoccupied with symbol-to-symbol transformations.

Also shared by both domains is the methodology of experimental programming, in that the computer is the ideal tool for performing experiments to validate or explore intelligent information-processing processes.

### Conclusion

It is notable how easily the more general characteristics of AI relate to the more specific characteristics of computer vision. Such a straightforward characterization surely suggests that the relationship is still strong. The largely shared goals, plus the additional support of the shared interests in philosophical underpinnings and neurophysiological mechanisms and processes, also strengthen the conclusion.

To complete this essay, I look at a few aspects of computer vision that are intertwined with some of the central issues of AI. To start with, let's consider nameability.

Computer vision at the highest levels addresses recognition of objects and actions. Because these entities are not distinguishable merely by appearance, this activity necessarily moves into other areas of AI, such as natural language for naming conventions, commonsense reasoning, and robotics (behavior). As the nameability of the human world depends on the humans that inhabit it, computer

vision must be based on the modes of intelligence that provide the names.

However, once we have the names, we also need some way to invoke mental structures connected with the names, that is, to select their model from the visual description base that potentially explains a set of visual data. The details of how this selection is done are not clear, but it is interesting that the selection process regularly appears in other subfields of AI, such as the invoking of schema for dialogue understanding, cases in case-based reasoning, or appropriate metalevel search heuristics.

Computer vision also requires the ability to generalize and reason about similarity. Suppose we encounter a person whom we have never seen before. We don't have to go sequentially: "Now, is this a house, a dog, an apple, or ...?" We can directly generalize from the specific person's appearance about the general nature of the humanness. We are not troubled much by new haircuts or missing limbs. Somehow we abstract into a space that compares the generalizations.

Computer vision will need to explore more case-based and opportunistic reasoning. It is well known that people need a period of training before they can achieve expert-level performance at new, nonintuitive, visual-interpretation tasks, such as x-ray interpretation and radar-display interpretation. It is also clear that people sometimes use specific features for search or identification—*cues* (for example, the color of a book or a scar distinguishing identical twins). Some visual learning seems to be largely iconic (for example, alphabetic letters, word groups). These examples suggest the use of case methods, with problem-specific compiled processes, rather than the use of generic, high-level visual processes. In addition, it is well-known that people can compile explicit reasoning into intuitive procedures, suggesting that future vision systems might have to apply their case-based reasoning in both forms.

Once our intelligent AI system is capable of performing, it will need a large visual knowledge base (a specialization of the general knowledge base search [Lenat and Feigenbaum 1991]). For general expert performance, a large corpus of common-sense visual knowledge will be needed, for example, how classes of objects typically appear; how to discriminate between specific classes; and what an object's material properties, likely contexts, and associations are. The vision system will need to build, extend, and generalize the database from new examples. Models of time, causality, general physical principles, and uncertainty will be needed to interpret the observations.

Through these themes, I am trying to second-guess the future direction for computer vision research. One common thread is that there will be a greater dependence on the methods being developed for general AI systems: case-frame matching, case-frame invocation, truth maintenance systems, generalization, learning, control of combinatorial search, and so on. Some form of self-understanding will be needed for feedback on performance. Planning and focus-of-attention mechanisms will be needed to focus computational resources. At the same time, computer vision will become essential for a truly intelligent autonomous AI machine, if only for the ability to learn for itself (let alone the philosophical contention that an artifact cannot know the world if it is incapable of acting in it and, hence, sensing it). Hence, I must conclude that the connections can only grow stronger as the two fields develop.

What is central to computer vision are issues of how to represent what is known and observable, how to reason with this represented information, and how to act on this knowledge (controlling both internal and external behavior). Because these will always be three of the main foundations of AI, there is no chance that computer vision will ever drift far. In the past decade, much computer vision research has concentrated on developing competent processes that reliably extract useful low-level

descriptions of the world. As this research matures, there will then be a major increase in research that relates these descriptions to stored representations of objects and situations, again making clear the association with AI.

### Acknowledgments

I would like to thank many people for their thoughts on this theme but particularly Alan Bundy, Howard Hughes, Chris Malcolm, Aaron Sloman, Manuel Trucco, and Mike Uschold.

### References

- Aloimonos, J. 1989. *Integration of Visual Modules: An Extension of the Marr Paradigm*. San Diego, Calif.: Academic.
- Barrow, H. G., and Tenenbaum, J. M. 1978. Recovering Intrinsic Scene Characteristics from Images. In *Computer Vision Systems*, eds. A. Hanson and E. Riseman, 3–26. San Diego, Calif.: Academic.
- Batchelor, B. G.; Hill, D. A.; and Hodgson, D. C. 1985. *Automated Visual Inspection*. Kempston, U.K.: IFS Publications.
- Boden, M. A. 1988. *Computer Models of Mind*. Cambridge: Cambridge University Press.
- Brooks, R. A. 1991. Intelligence without Reason. In *Proceedings of the Twelfth International Joint Conference on Artificial Intelligence*, 569–595. Menlo Park, Calif.: International Joint Conferences on Artificial Intelligence.
- Brooks, R. A. 1981. Symbolic Reasoning among 3-D Models and 2-D Images. *Artificial Intelligence* 17:285–348.
- Dennett, D. C. 1992. *Consciousness Explained*. London: Allen Lane.
- Draper, B.; Collins, R.; Brolio, J.; Hanson, A.; and Riseman, E. 1988. Issues in the Development of a Blackboard-Based Schema System for Image Understanding. In *Blackboard Systems*, eds. R. Englemore and T. Morgan, 189–218. Reading, Mass.: Addison-Wesley.
- Durrant-Whyte, H. F. 1987. Uncertain Geometry in Robotics. In *Proceedings of the IEEE Conference on Robotics and Automation*, 851. Washington, D.C.: IEEE Computer Society.
- Farah, M. J. 1990. *Visual Agnosia*. Cambridge, Mass.: MIT Press.
- Fisher, R. B. 1989. *From Surfaces to Objects: Computer Vision and Three-Dimensional Scene Analysis*. New York: Wiley.
- Gibson, J. J. 1979. *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin.
- Grimson, W. E. L. 1990. *Object Recognition by Computer: The Role of Geometric Constraints*. Cambridge, Mass.: MIT Press.
- Haugeland, J. 1985. *Artificial Intelligence: The Very Idea*. Cambridge, Mass.: MIT Press.
- Hinton, G. 1981. A Parallel Computation That Assigns Canonical Object-Based Frames of Reference. In *Proceedings of the Seventh International Joint Conference on Artificial Intelligence*, 683–685. Menlo Park, Calif.: International Joint Conferences on Artificial Intelligence.
- Horn, B. 1975. Obtaining Shape from Shading Information. In *The Psychology of Computer Vision*, ed. P. Winston, 115–155. New York: McGraw-Hill.
- Hubel, D. H. 1988. *Eye, Brain, and Vision*. New York: Freeman.
- Lenat, D. B., and Feigenbaum, E. 1991. On the Thresholds of Knowledge. *Artificial Intelligence* 47:185–250.
- Marr, D. 1982. *Vision*. New York: Freeman.
- Matsuyama, T. 1990. *SIGMA: A Knowledge-Based Aerial Image Understanding System*. New York: Plenum.
- Penrose, R. 1989. *The Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics*. Oxford: Oxford University Press.
- Rosenschein, S. J., and Kaelbling, L. P. 1986. The Synthesis of Machines with Provable Epistemic Properties. In *Proceedings of the Conference on Theoretical Aspects of Reasoning about Knowledge*, ed. J. Halpern, 83–98. San Mateo, Calif.: Morgan Kaufmann.
- Schank, R. C. 1991. Where's the AI? *AI Magazine* 12(4): 38–49.
- Sloman, A. 1988. Why Philosophers Should Be Designers. *Behavioral and Brain Sciences* 11(3): 529–530.
- Waltz, D. 1975. Understanding Line Drawings of Scenes with Shadows. In *The Psychology of Computer Vision*, ed. P. Winston, 19–91. New York: McGraw-Hill.
- Zeki, S. 1993. *A Vision of the Brain*. Oxford: Blackwell.



**Robert Fisher** is a senior lecturer in the Department of Artificial Intelligence at the University of Edinburgh. He has been a member of the staff since 1984. He received his Ph.D. from the University of Edinburgh in 1987 and has been researching three-dimensional scene understanding and other topics in computer vision for more than 10 years.