17

Adequacy Criteria for Visual Knowledge Representation

Alan K. Mackworth*

Department of Computer Science
University of British Columbia

INTRODUCTION

The proper study of artificial intelligence is the design of computational systems that represent, use, and acquire knowledge to perceive, reason, communicate, and act. Under that definition, knowledge representation is the heart of artificial intelligence. Past and future success in building systems for vision, problem solving, planning, and language depends critically on progress in knowledge representation. Workers in the field have been prolific in proposing and exploiting a variety of knowledge representation schemes such as grammars, semantic nets, programs, logics, schemas, rules, constraints, and neural nets. However, as we explore in the world of knowledge representation we need navigational tools: the analogs of chart, compass, log, and sextant. In this paper a framework for evaluating knowledge representation schemes is presented.

A BRIEF HISTORY

Chomsky's (1965) early work on syntactic structures was explicitly motivated by adequacy criteria. In developing the theory of transformational grammar the devices of anomaly, paraphrase, and ambiguity were constantly exploited as experimental probes. By regard-

* Fellow, Canadian Institute for Advanced Research

ing a particular grammatical framework as a representation of grammatical knowledge Chomsky was asking, "Is this representation sufficiently powerful to capture these phenomena?"

In fact, he was able to go beyond sufficiency to necessity. For a structurally ambiguous sentence *any* adequate grammatical framework must necessarily provide multiple syntactic interpretations. For a structurally anomalous sentence *any* adequate grammar must fail to provide an interpretation. Two sentences that are mutual paraphrases must share common structure in their syntactic interpretations. The imaginative use of these probes allowed the delimitation of the necessary boundary conditions of any adequate grammar. In particular, the many varieties of anomaly were particularly fruitful.

Chomsky proposed the problematic competence/performance distinction but concentrated the adequacy arguments on the competence of the "ideal" speaker/hearer's generative (rule-based) mental representations. This approach directly inspired Clowes (1971) and Huffman (1971) to specify representations for simple blocks world scenes that would similarly reject ill-formed scenes and discover multiple interpretations of ambiguous pictures. Well-chosen examples highlight distinctions, representations, and chains of reasoning that any adequate system must make, possess, and follow to be able to discriminate well-formed from ill-formed scenes, for example, as illustrated in Figure 1.

McCarthy and Hayes (1969) further distinguished three kinds of adequacy for mental representations of the world: "metaphysical," "epistemological," and "heuristic." The focus of their paper is on epistemological adequacy: "A representation is called epistemologically adequate for a person or machine if it can be used practically to express the facts that one has about the world." On this basis they argued for various logical systems as declarative representation languages.

In the early 1970s, other researchers were building and advocating systems that exploited procedural representations of knowledge; the ensuing declarative/procedural controversy featured a loud debate between two entrenched camps. Despite attempts to find a synthesis of the two positions (Winograd, 1975) the best perspective on the controversy is afforded by realizing that the two camps were implicitly relying on criteria that emphasized different forms of adequacy of knowledge representation, thus ensuring the incoherence of the debate.

In computer science the distinction between the specification level and the implementation level of description of a computational task is common. It is also common to identify these levels, roughly,

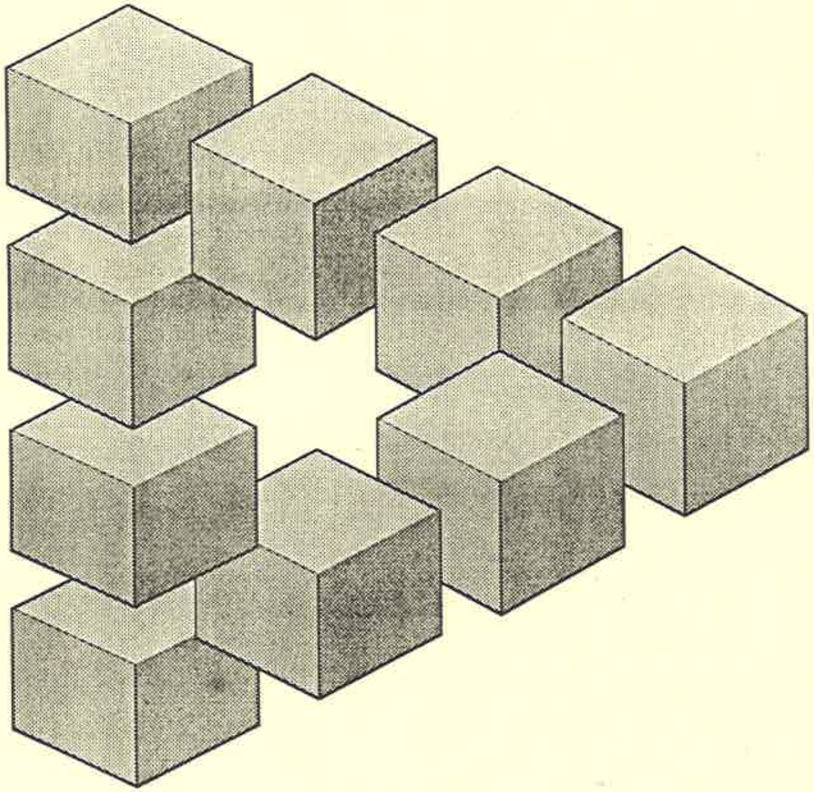

Figure 1. What's wrong with this scene?

with "what" and "how." The implementation level is often recursively described as a series of levels such that each level is implemented or realized on top of the virtual machine provided by the level immediately below it. This recursion terminates at the physical machine level—the real hardware—that provides primitive computational operations.

Marr (1982) discussed three levels:

1. Computational theory
2. Representation and algorithm
3. Hardware implementation

He emphasized the one-many relationships from level 1 to level 2 and from level 2 to level 3. Marr said that level 1 describes the

"what" and "why" of the computational task, level 2 specifies "how" at an abstract level, and level 3 specifies "how" at the concrete level of an artificial or biological machine. One can criticize this approach (surely representations are required at level 1 and 3 as well as at level 2) but the crucial point that Marr makes is that we cannot say that we understand how a machine is carrying out an information processing task until we have adequate descriptions at all levels.

ADEQUACY CRITERIA

These attempts to clarify the notion of adequacy often interact in oblique ways: each author emphasizes a different idiosyncratic aspect. We have not yet arrived at a coherent theory with commonly accepted set of terms and criteria. This paper is an attempt to establish a framework for that theory.

The claim is that adequacy criteria can be categorized as *descriptive adequacy* and *procedural adequacy* criteria (Mackworth, 1977) and that both must necessarily be satisfied by an adequate knowledge representation scheme.

DESCRIPTIVE ADEQUACY

Descriptive adequacy criteria are concerned with the extent to which the mental representation adequately describes or represents situations in the world. Eleven descriptive adequacy criteria are outlined here.

D1. *Capacity*. If the computational task requires that an unbounded number of possible situations in the world be distinguished then a descriptively adequate finite representation scheme (Hayes, 1974) must necessarily embody a generative and recursive set of rules (in some language) that can generate an unbounded number of configurations.

D2. *Primitives*. The set of rules must generate descriptions of the legitimate primitive objects in the world and their possible attributes and relationships.

D3. *Composition*. Composition rules generate descriptions of structured objects, their components and their attributes and relationships.

D4. *Specialization.* Specialization rules generate possible refinements of object classes.

D5. *Subworlds.* If the world of interest consists of two or more distinct subworlds then the representation scheme must maintain that distinction. For example, for visual knowledge representation, the two-dimensional image domain and the three-dimensional scene domain are mutually exclusive. A descriptively adequate visual knowledge representation can only avoid elementary category errors, such as confusing an edge with the line that depicts it, by categorizing or typing objects as belonging to one domain or the other. (Any real visual knowledge representation will make several finer distinctions.) If the distinction between image and scene is maintained then there must be two sets of rules describing the primitives in both domains and the composition rules needed to form composite objects.

D6. *Depiction.* A visual knowledge representation must also carry information about the *depiction relation:* how objects in the scene domain appear in the image domain. (The use of the term "relation of representation" for the depiction relation is misleading, as we shall see.)

D7. *Equivalence Classes.* For three-dimensional worlds the depiction relation is a many-to-one mapping function from the scene domain to the image domain, confounding the subdomains of lighting and surface reflectance, orientation, shape and position, with viewpoint and other attributes of the imaging situation (Mackworth, 1983b).

Even for a simple image like that in Figure 2 there are an infinite number of scenes that could serve as legal interpretations. The representation, to serve descriptive adequacy, must provide a finite representation of all and only those scenes which, technically, constitute an equivalence class. The configuration in the representation scheme is a description of that equivalence class. Much of the history of computational vision research can be seen as a continuing investigation into this aspect of descriptive adequacy.

D8. *Detail.* An adequate visual representation provides descriptions at a variety of physical scales, both in the image and in the scene.

D9. *Stability.* In a stable representation scheme a minor change in the world causes, at most, a minor change in the representation. The detail and stability criteria are decoupled in the sense that it is possible to satisfy one and not the other.

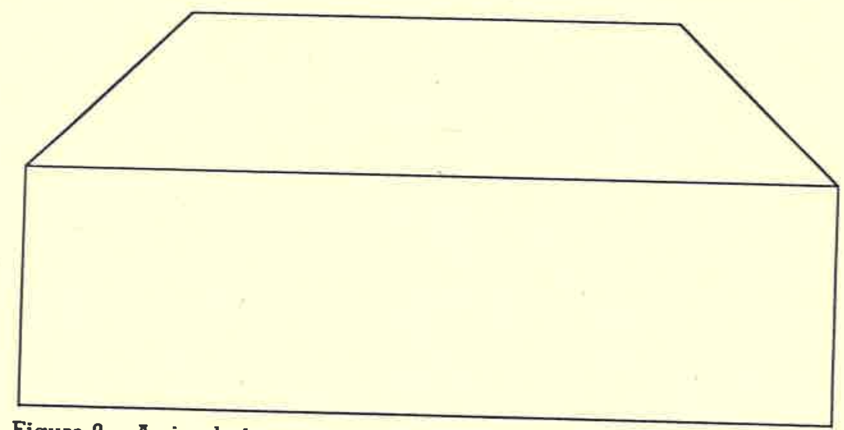

Figure 2. A simple image.

D10. *Invariance.* It is often desirable for the representation to be essentially invariant under transformations of the world just as the deep structure of a sentence is essentially invariant under the passive transformation. The paraphrase technique is useful for evaluating invariance.

D11. *Correct.* It should not go without saying that the representation should be correct. The *relation of representation* is the relation between a situation and the configuration describing it. By insisting that the relation of representation be a total function from situations onto configurations we ensure, for example, that the representation of any unambiguous situation is a *canonical* configuration not allowing or requiring any arbitrary choices. Under *correctness* we include anomaly and ambiguity. An anomalous situation should have no coherent configuration while an ambiguous situation should have two or more. Anomaly and ambiguity are useful for evaluating correctness and, further, may provide diagnostic advice on how to enhance the adequacy of the representation.

This framework allows us to look retrospectively at all computational vision research and judge the extent to which descriptive adequacy is achieved.

As a familiar illustration, consider the Huffman-Clowes labeling scheme shown in Figure 3. It satisfies, to a greater or lesser extent, criteria D1 (Capacity), D2 (Primitives), D3 (Composition), D5 (Sub-

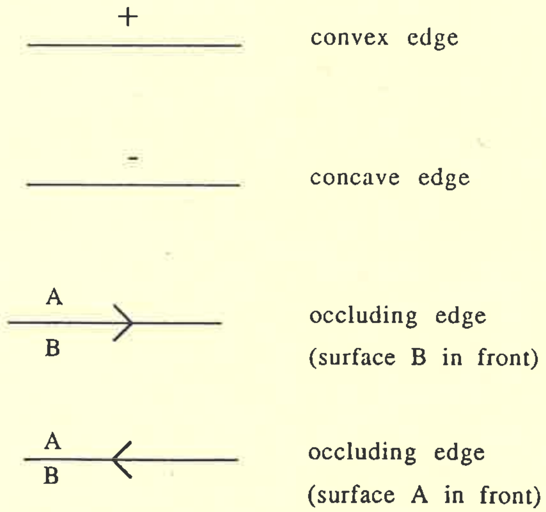

Figure 3. The Huffman-Clowes labels for an edge.

worlds), D6 (Depiction), D7 (Equivalence Classes), and D11 (Correctness). For an image of a single line, it generates the four interpretations of Figure 3, which is an extensional representation of an equivalence class of scenes. Each of the four interpretations corresponds to an intensional representation in that the degree of convexity/concavity of the edge is not specified. A description of the image in Figure 2, in terms of that scheme, is shown in Figure 4: it stands for an infinite set of legal scenes that produce that image. Unfortunately, the labeling of Figure 5 is also allowed by that

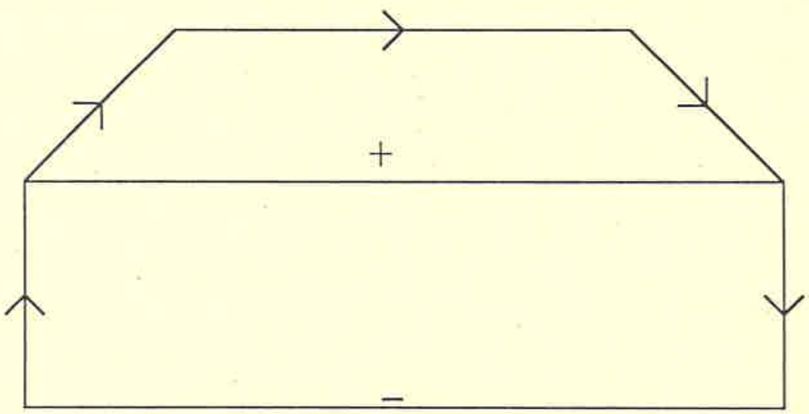

Figure 4. A correct interpretation of Figure 2.

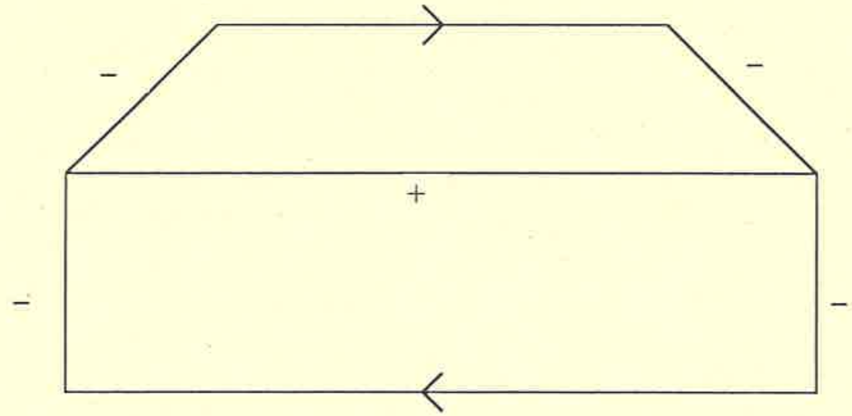

Figure 5. An anomalous interpretation of Figure 2.

scheme but does not correspond to any legal scene. In order to eliminate such spurious interpretations we were forced to provide richer descriptions of this simple visual world that include information on possible surface orientations and positions (Mackworth, 1977).

The design of Mapsee2, a system that interprets hand drawn sketch maps, was explicitly motivated by adequacy criteria (Havens & Mackworth, 1983). Criteria D1–D6 and D11 were, on the whole, satisfied for this domain through the use of schema based knowledge representation. Essentially in that system a partially instantiated schema instance corresponds to the evolving description of an equivalence class.

Recent work in shape description (Mackworth & Mokhtarian, 1984; Mokhtarian & Mackworth, 1986) provides a representation (curvature scale space) that satisfies D1–D3 and D8 (Detail), D9 (Stability), D10 (Invariance), and D11 (Correctness).

This framework allows us to look retrospectively at research into the use of knowledge representation in vision systems but more importantly, it encourages the use of explicit adequacy criteria as a guiding research strategy.

PROCEDURAL ADEQUACY

Procedural adequacy criteria assess aspects of the use and acquisition of the knowledge embodied in a representation scheme. Five procedural adequacy criteria are outlined here.

P1. *Soundness*. The use of the knowledge is sound if the system embodying it produces only interpretations allowed by the knowledge representation.

P2. *Completeness*. The use of the knowledge is complete if the system embodying it produces all the interpretations allowed by the knowledge representation.
Soundness and completeness are adopted, by analogy, from theorem proving where an inference strategy is sound and complete if it proves only and all the theorems logically implied by the axioms.

P3. *Flexibility*. Another procedural adequacy consideration is the flexibility of use of the available information. An ideal image-based system would regard all of its potential information sources as inputs or outputs, depending on the availability of information (Mackworth, 1983a). The system would allow control flow from image to scene (image analysis) or from scene to image (image synthesis), bidirectionally or multidirectionally if more than two domains or information sources were available. Or, indeed, information sources may start as partially specified by a symbolic description and the system would refine that description as it used the other information sources and the constraints embedded in the domain knowledge representation. Under this view, the dichotomous classification of potential information sources/sinks as inputs or outputs becomes obsolete. Until we achieve the ideal, we also discuss under procedural adequacy the control facilities provided in the knowledge representation language and the ease of reprogramming the system from, say, analysis to synthesis (Stanton, 1972).

P4. *Acquisition*. A key aspect of procedural adequacy is the knowledge acquisition process. Is the knowledge representation suitable for knowledge acquisition? Does acquisition occur through evolution, learning, teaching, or explicit programming? Are there appropriate, efficient, and flexible algorithms for knowledge acquisition?

P5. *Efficiency*. Under efficiency criteria we can include standard computational complexity arguments. Using such arguments we should first establish a lower bound on the time (and/or space) complexity of the task or problem itself, that is, a lower bound on the resources that *any* correct algorithm must consume. This is the inherent complexity of the task. Then any proposed algorithm may be analyzed for its worst or average case performance. The worst case performance of an *optimal* algorithm will equal the complexity of the problem.

A vision program that interpreted an image by exhaustive analysis-by-synthesis enumerating the members of the infinite set of all possible images until one matched the input would not satisfy the efficiency criterion of procedural adequacy. A variety of ways of controlling that search and reducing the resources consumed are available such as the use of image features to index into the knowledge representation.

However, many of the interesting problems in AI are NP-complete—that is the problems themselves apparently inherently require exponential resources (as a function of the size of the input). For example, the Huffman-Clowes labeling problem itself has recently been shown to be such a problem (Kirousis & Papadimitriou, 1985).

Also under efficiency we should discuss the use of parallelism. Many new models and technologies of computation challenge the standard von Neumann model of computation. Many of these display coarse or fine-grained parallelism often on a massive scale; these affect radically the complexity measures and criteria used.

THE INTERACTION BETWEEN DESCRIPTIVE AND PROCEDURAL ADEQUACY

Apart from the claim that a good meta-framework for knowledge representation allows us to design and build better knowledge representations, what are the other practical implications of the point of view advocated here? The most interesting implications arise from the interactions between descriptive and procedural adequacy.

Given two knowledge representation schemes equal from the perspective of descriptive adequacy we should, of course, choose the one that more nearly meets the criteria of procedural adequacy. Or, more constructively, design a system to more nearly meet those criteria.

The twin sets of criteria are, of course, often apparently in conflict. The approach more usually taken in AI has been to favor procedural efficiency if not full procedural adequacy (which includes soundness, completeness, acquisition, and flexibility of use of knowledge) by hand coding special purpose procedures into application programs. The current trend is away from that approach as the full dimensions of the adequacy issue become apparent.

However, there are fundamental theoretical obstacles to achieving simultaneously full descriptive and procedural adequacy that are as important to theories of knowledge representation as comparable *laws of impotence*, such as Heisenberg's Uncertainty Principle and Einstein's Laws of Relativity, are to physics.

For example, Levesque and Brachman (1985) consider an aspect of descriptive adequacy ("expressiveness") and an aspect of procedural adequacy ("tractability") to be involved in a fundamental tradeoff. Full first order logic is only semi-decidable. Certain subsets of first order logic are decidable but, putatively, only in exponential time while more restricted subsets are decidable in polynomial time.

Horn clause form is a restriction of first order logic that requires each sentence to have one of the two forms:

$$P \leftarrow Q \wedge R \wedge \ldots$$
$$\text{or} \quad W$$

where P, Q, R, W are predicate symbols (which may take terms as arguments). Prolog programs are sets of sentences that take this form. This limitation on expressive power (sentences such as $P \vee Q$ are not allowed) combined with assumptions such as the unique names assumption (Reiter, 1978), and negation as failure (Clark, 1978) brings some advantages in procedural adequacy. In particular, although the logic is still undecidable, the resolution strategy used is relatively efficient and the synthesis/analysis flexibility of the system is useful; the same knowledge can be used in several "directions", although the user may optimize the knowledge base for use in a certain procedural direction.

Given that many interesting tasks (such as the labeling problem mentioned earlier) seem to be inherently exponential the strategy of weakening descriptive adequacy to achieve efficient (polynomial) algorithms is not available. A complementary strategy is to search for efficient *approximation* algorithms for such tasks. An approximation algorithm might, for example, provide a necessary but not always sufficient test for, say, the acceptability of a picture. For the labeling problem, the Waltz (1972) arc consistency "filtering" algorithm is such a test which runs in linear time (Mackworth & Freuder, 1985).

Generalizations of this approach known as network consistency algorithms can be used for many NP-complete tasks and there is a general network consistency approach that provides a spectrum of levels of consistency (applying tighter and tighter tests of consistency) at increasing computational cost. For many (but, provably, not for all) problems the cheaper, lower levels of consistency are sufficient. For example, if the constraint graph is a tree then simple arc consistency (running in linear time) is both a sufficient and a necessary test for the existence of a solution (Mackworth and Freuder, 1985).

Mapsee3 explicitly constructs a constraint graph among schema instances. It uses a hierarchical arc consistency algorithm, exploiting the tree-structured specialization hierarchy, to search efficiently for an interpretation of the sketch map (Mackworth, Mulder, & Havens, 1985).

Parallelism may offer attractive solutions to some problems but others are not amenable to that approach. In particular, the relaxed form of constraint satisfaction known as arc consistency is apparently inherently sequential. Kasif (1986) has shown that arc consistency is log-space complete for P, the class of problems solvable on a single Turing machine in polynomial time. The implication of this is that it is unlikely that arc consistency can be solved in polylogarithmic time with a polynomial number of processors.

CONCLUSION

In summary we have developed explicit criteria for evaluating the descriptive and procedural adequacy of visual knowledge representations. These criteria may be used to evaluate representation schemes and to design better ones but, as pointed out, there are theoretical obstacles to satisfying fully all the criteria simultaneously.

REFERENCES

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Clark, K. L. (1979). Negation as Failure. In H. Gallaire & J. Minker (Eds.), (pp. 293–322) *Logic and data bases*. New York: Plenum.

Clowes, M. B. (1972). On seeing things. *Artificial Intelligence, 2*, 79–112.

Havens, W. S., & Mackworth, A. K. (1983). Representing knowledge of the visual world. *IEEE Computer, 16*, 90–96.

Hayes, P. J. (1974). Some problems and non-problems in representation theory. AISB Summer Conference, University of Sussex, pp. 63–79.

Huffman, D. A. (1971). Impossible objects as nonsense sentences. *Machine Intelligence, 6*, (p. 295–323) Meltzer & D. Michie (Eds.), American Elsevier, NY.

Kasif, S. (1986). On the parallel complexity of some constraint satisfaction problems. Proceedings of the Fifth National Conference on Artificial Intelligence, Philadelphia, pp. 349–353.

Kirousis, L. M., & Papadimitriou, C. H. (1985). The complexity of recognizing polyhedral scenes. *26th Annual Symposium on Foundations of Computer Science*, IEEE Computer Society, October 21–23, pp. 175–185.

Levesque, H. J., & Brachman, R. J. (1985). A fundamental tradeoff in knowledge representation and reasoning (revised version). In R. J. Brachman & H. J. Levesque (Eds.), *Readings in knowledge representation*, (pp. 41–70). Morgan Kaufmann Publishers, Inc., Los Altos, CA.

McCarthy, J., & Hayes, P. (1969). Some philosophical problems from the standpoint of artificial intelligence. *Machine Intelligence 4*, (pp. 463–502). B. Meltzer and D. Michie (eds.), American Elsevier, NY.

Mackworth, A. K. (1977). How to see a simple world: An exegesis of some computer programs for scene analysis. *Machine Intelligence, 8*, E. W. Elcock and D. Michie (eds.), pp. 510–540. New York: John Wiley.

Mackworth, A. K. (1983a). Recovering the meaning of diagrams and sketches. *Proceedings Graphics Interface '83*, Edmonton, Alberta, Canada, pp. 313–317.

Mackworth, A. K. (1983b). Constraints, descriptions and domain mappings in computational vision. In O. J. Braddick & A. C. Sleigh (Eds.), *Physical and biological processing of images* (pp. 33–40). Berlin: Springer-Verlag.

Mackworth, A. K., & Freuder, E. C. (1985). The complexity of some polynomial network consistency algorithms for constraint satisfaction problems. *Artificial Intelligence, 25*(1), 65–74.

Mackworth, A. K., & Mokhtarian, F. (1984). Scale-based descriptions of planar curves. *Proceedings of the Canadian Society for Computational Studies of Intelligence*, London, Ont., 114–119.

Mackworth, A. K., Mulder, J., & Havens, W. S. (1985). Hierarchical arc consistency: Exploiting structured domains in constraint satisfaction problems. *Computational Intelligence, 1*(3), 118–126.

Marr, D. (1982). *Vision* San Francisco: W. H. Freeman.

Mokhtarian, F., & Mackworth, A. K. (1986). Scale-based description and recognition of planar curves and two-dimensional shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 8*, (1) 34–43.

Reiter, R. (1978). On closed world data bases. In H. Gallaire & J. Minker (Eds.), *Logic and data bases* (pp. 55–76). New York: Plenum.

Stanton, R. B. (1972). The interpretation of graphics and graphic languages. In F. Nake & A. Rosenfeld (Eds.), *Graphic languages* (pp. 144–159). Amsterdam: North-Holland Publishing Co.

Waltz, D. L. (1972). Understanding line drawings of scenes with shadows. In P. H. Winson (Ed.), *The psychology of computer vision* (pp. 19–91). New York: McGraw-Hill.

Winograd, T. (1975). Frame representations and the declarative/procedural controversy. In D. G. Bobrow & A. Collins (Eds.), *Representation and understanding: Studies in cognitive science* (pp. 185–210). New York: Academic Press.