SCHEMATA-BASED UNDERSTANDING

OF

HAND-DRAWN SKETCH MAPS

William S. Havens

Computer Sciences Department
University of Wisconsin
Madison, Wisconsin
53706

Alan K. Mackworth

Computer Science Department
University of British Columbia
Vancouver, British Columbia
V6T-1W5

Abstract

This paper describes current research in applying schemata-based recognition methods to the understanding of hand-drawn sketch maps. In this system, schemata are employed as representations for models of the cartographic objects and systems of objects possible in sketch maps. The resulting hierarchical network is then searched using a combination of both data-driven and model-driven methods. Low-level models are invoked by primary cues computed directly from the input image. Once invoked, schema models apply object-specific procedural methods to complete their recognition. Completed schema instances are then used as abstract cues to invoke other models higher in the schema hierarchy. A multiprocessing control regime is utilized to permit a number of schemata to apply their recognition procedures concurrently.

1. Introduction

In order to cope with the enormous complexity of visual information, computer vision systems must employ extensive model-specific knowledge of the visual world. A major problem in model-driven vision systems is the invocation of appropriate models to interpret a given image. Typically, data-driven methods are employed to generate low-level image cues to select likely models as hypotheses. It has been pointed out that this method is ineffective. Low-level cues are highly ambiguous matching to many inappropriate high-level models.

As a solution to this problem, we are currently integrating model-driven and data-drive recognition in schemata representations by employing a recursive hierarchy of cues and models. Schema models are invoked both by primary cues computed directly from the image and by abstract cues created recursively as the result of recognition. The successful recognition of a schema instance at one level in the hierarchy yields a context-sensitive cue to invoke schema models at higher levels.

Sketch maps have been chosen for this research for the following reasons:

1) We believe that the conventional semantics of of cartography accurately reflects geographic features in real aerial and satellite imagery.
2) The use of vector graphic input data greatly reduces the amount of low-level processing required while still capturing the essential difficulties of geographic image analysis. The research is therefore able to focus on issues of cue generation and model invocation.
3) The enhanced abilities of this approach can be easily compared to a previous sketch map system, MAPSEE, [Mackworth, 1977a] employing a constraint network representation and a network consistency search method [Mackworth, 1975]. By testing both systems on the same input maps, we should obtain a quantitative measure of the expected improvement of schemata over constraint network methods.

2. Model-Driven Recognition

Computer vision can be characterized as the task of mapping a two-dimensional sensory image into an abstract symbolic description of the three-dimensional scene represented by that image [Clowes, 1971]. This process necessarily involves the interpretation of sensory signals that are voluminous in their quantity and simultaneously highly ambiguous in their possible meanings. In order to cope with this complexity, com-

In Proc. Can. Soc. Comp. Studies of Intelligence Conf., 1980, Victoria B.C.

uter vision systems must employ both model-riven and data-driven recognition methods. odel-driven recognition utilizes knowledge of he objects and their abstract relationships ossible in the visual world. Conversely, data-riven methods exploit spectral knowledge about he signal source and physical knowledge about he processes of image formation and surface re-overy. Indeed, the representation and coordi-ated application of knowledge is the central roblem in computer vision [Reddy, 1978].

We are exploring a recognition paradigm or computer vision that integrates top-down, odel-driven recognition with bottom-up, data-riven methods in hierarchical schemata-based nowledge representations [Havens, 1978a]. A ajor problem in model-driven vision systems is ne invocation of appropriate models to inter-ret a given image. In most current systems, ata-driven methods are employed to generate ow-level image cues to select likely models as ypotheses. Cues can be regions of statistical-y homogeneous properties or edges inferred rom characteristic changes in image intensity. t has been pointed out that this methodology is neffective [Barrow & Tenenbaum, 1975]. Region d edge-finding algorithms have no knowledge of e real objects to which they may belong. As result, low-level cues are highly ambiguous tching too many inappropriate high-level dels.

As a solution to this problem, we argue at high-level object models must be invoked appropriate high-level cues. The discovery such abstract cues is, of course, recursively e recognition problem, thereby necessitating e use of a recursive hierarchy of cues and dels. Schema models must be invoked both by imary cues computed directly from the image d by abstract cues created recursively as the sult of recognition. The successful recog-tion of a schema instance at one level yields context-sensitive cue to invoke schema models the next higher level.

To realize this recognition paradigm, we e employing a multiprocessing programming language methodology that supports the concurrent execution of top-down and bottom-up search processes in hierarchical knowledge representations [Havens, 1978b].

3. Schemata Representations

Recent research has focused on the application of schemata [Bartlett, 1932] as a representation of knowledge [Minsky, 1975] [Bobrow & Winograd, 1977] [Rumelhart & Ortony, 1976]. Schemata have been used or proposed in a number of computer vision systems [Freuder, 1976] [Hanson & Riseman, 1978b] [Brady, 1978]. Schemata are object centered representations which represent complex concepts as specific compositions of simpler schemata thereby forming hierarchical knowledge structures. By exploiting composition, a finite number of schema stereotypes can be used to represent an arbitrary number of object instances. Schemata may contain both active and passive knowledge. Passive knowledge represents descriptive models of stereotypical objects. Active knowledge is represented as procedures attached to schema models to guide the recognition process for instances of those schema stereotypes [Winograd, 1975].

The recognition process in schemata-based systems can be characterized as a search of the schema hierarchy to find a best match of the information present in the input image to the knowledge represented in the knowledge-base. Havens [1976] has shown that this search can be neither a strict top-down nor bottom-up search. Instead, recognition must be an integration of top-down, hypothesis-driven search and bottom-up, data-driven methods [Rumelhart & Ortony, 1976]. Schemata represent models providing expectation and guidance for top-down search. At the same time, features discovered in the image provide cues for the bottom-up selection of particular schemata as likely hypotheses.

We are investigating the integration of model-driven and data-driven recognition by employing both a model hierarchy and a cue hierarchy within a schemata knowledge representation. The interactions between model-driven and data-driven processes in computer vision are poorly understood [Brady, 1978]. This research is concerned with characterizing that interaction.

A preliminary schemata knowledge representation for sketch maps is illustrated in Figure 3. The nodes in this tree represent schema stereotype models of various cartographic objects and systems possible in sketch map scenes. The arcs represent composition with nodes higher in the hierarchy being composed of connected nodes lower in the hierarchy. The interpretation of the arcs, however, depends on whether a top-down or bottom-up search method is being applied. Using top-down search, the arcs are possible subgoal paths. To recognize a Road-System, for instance, this schema can selectively call the Town, Road and Bridge schemata as subgoals. Using bottom-up search, on the other hand, the arcs represent cue paths to select possible supergoals. As an example, if the Bridge schema has satisfied its expectations for a bridge instance in the input image, it must invoke plausible higher schemata as supergoals. In this case, both Road-System and River-System are very likely to be found in a sketch map scene containing a bridge.

For this sketch map system, the image is represented as plot vectors taken directly from a vector graphics tablet. Conceptually, the data consists of connected image points called Links and blank space called Patches. We have employed a simple recursive quadrant-splitting region finder to yield a conservative first segmentation. For this domain, regions are thought to be poor cues. Instead, a line finder which attempts to connect plot vectors into chains is used to provide primary cues. This algorithm is again chosen to be conservative, forming chains only where the distance between links is small and there is no ambiguity as to chain direction.

4. Cycles of Perception

Unfortunately, to completely segment a complex image requires the use of model-specific information about the scenes interpretation, yet that information is only fully available after the segmentation has been performed. In order to avoid this "chicken and egg problem" [Mackworth, 1977b] [Havens, 1976], an integration of low and high-level processing must be achieved. Mackworth [1978] has advocated a "cycle of perception" theory for computer vision to avoid this problem (see Figure 1). Kanade [1977] defines a similar cyclic model. An initial conservative segmentation is used to generate primary cues that invoke appropriate object models. Once invoked, these models can guide a context-sensitive resegmentation of the image, thereby providing new more powerful cues to repeat the cycle.

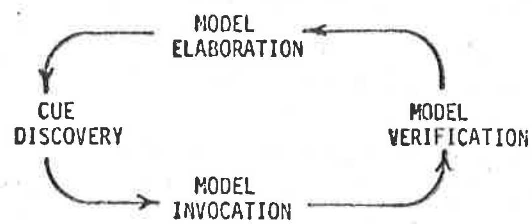


Figure 1

We argue that this cycle of perception can as well be characterized as a recursive process. When all the expectations of a particular model have been satisfied, the instantiated model becomes an abstract cue to recursively invoke appropriate models higher in the knowledge hierarchy (see Figure 2). Instead of relying only on primitive context-free cues, the recognition of schema instances at intermediate levels in the hierarchy can provide context-sensitive cues for the next level.

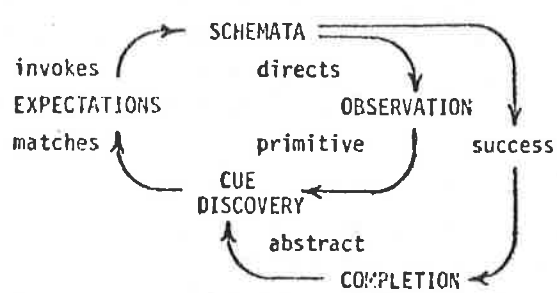


Figure 2

Programming Methodology

The development of programming methodology or such tasks as computer vision is an active rea of research [Bobrow & Raphael, 1974]. Re- ent work has focused on the development of chemata-based programming languages such as KRL Bobrow & Winograd, 1977] and MAYA [Havens, 978b]. Such languages define data structures for representing and accessing schemata and for constructing schemata networks. A method of copying stereotype schemata to provide specific schema instances is also essential. Since sche- mata may contain both descriptive and procedural knowledge, a mechanism must also be included for allowing attachment of procedures to data within schemata [Winograd, 1975].

The procedures associated with each schema are considered model-specific methods for guiding the search process for that schema. Procedural methods can be used in both top-down and bottom- up search of the schema hierarchy. Both require multiprocessing capabilities. Top-down, subgoal search can be realized by using generators [Suss- man & McDermott, 1972] as independent processes that can be recalled on failure to repeatedly at- tempt alternative solutions to their subgoals.

Bottom-up search requires that a number of models be allowed to be active hypotheses simul- taneously. Therefore, each procedural method as- sociated with an active schema must be realized as an independent process. Each such process is allowed to guide the recognition process for its schema stereotype. The coordination of multiple competing processes in goal-directed systems is poorly understood [Brady, 1978]. KRL defines a hierarchy of scheduler processes but leaves the specification of these schedulers to the program- mer.

To the contrary, MAYA defines four control primitives for implementing bottom-up, data- driven recognition in schemata networks. The first primitive, PROCESS, creates a new process associated with some schema and begins its execu- tion. This process may attempt to satisfy its schema's model by employing subgoal search or by invoking low-level iconic processes to generate primary cues. If the search is unproductive, the process can suspend itself, using SUSPEND, to simple n-tuple patterns representing the unful- filled expectations of the schema model. When such information is discovered later by a lower process, the process can be restarted, using RESUME, by a successful pattern match to its pat- tern. A number of schemata can, therefore, con- duct their recognition in pseudo-parallel being activated by the discovery of cues or information matching their model's expectations, applying their methods, suspending themselves when infor- mation is not available, and being resumed later by the discovery of additional matching cues of information. See Figure 4.

This iterative cycle continues for each active schema until some schema succeeds in satisfying its model's expectations. If the schema is intermediate in the schema hierarchy, then the completed schema instance is an ab- stract cue. The control primitive COMPLETE al- lows this schema to perform two essential control operations. A pattern match determines which higher schemata processes are waiting for the in- formation provided by this completed instance. The completed process is suspended and the matched higher-level processes are resumed, in turn, to continue their own methods.

6. Conclusion

This research is concerned with extending the use of model-driven recognition methods in computer vision. By employing a recursive hier- archy of cues and models, represented as sche- mata, the acknowledged difficulties of invoking models by low-level cues are avoided. By using a schema-based multiprocessing programming en- vironment, a number of models can simultaneously be active hypotheses applying their object- specific methods concurrently. Finally, by test- ing these techniques on the idealized domain of cartographic sketch maps, both qualitative and quantitative measures of their performance can be obtained.

References

Barrow, H.G. & Tenenbaum, J.M. (1975) Representation and Use of Knowledge in Vision, Tech. Note 108, Artificial Intelligence Center, SRI International, Menlo Park, Ca.

Bartlett, F.C. (1932), Remembering, Cambridge University Press, England.

Bobrow, D.G. & Raphael, B. (1974) New Programming Languages for Artificial Intelligence Research, Comp. Surveys 6, #3, Sept. 1974, pp. 153-174.

Bobrow, D.G. and Winograd, T. (1977). An Overview of KRL: A Knowledge Representation Language, Cognitive Science, Vol. 1, #1, January 1977.

Brady, J.M. & Wielinga, B.J. (1978) Reading the Writing on the Wall, in Hanson & Riseman [1978a], pp. 283-302.

Clowes, M.B. (1971) On Seeing Things, Artificial Intelligence 2, pp. 79-116.

Freuder, E. (1976) A Computer System for Visual Recognition Using Active Knowledge, Ph.D. Thesis, AI-TR-345, MIT AI Lab, Cambridge, Mass.

Hanson, A.R. & Riseman, E.M. (1978a) (eds.) Computer Vision Systems, Academic Press, New York.

Hanson, A.R. & Riseman, E.M. (1978b) VISIONS: A Computer System for Interpreting Scenes, in Hanson & Riseman [1978a].

Havens, W.S. (1976) Can Frames Solve the Chicken and Egg Problem?, Proc. 1st National Conference, CSCSI, Vancouver, Canada, August 1976.

Havens, W.S. (1978a) A Procedural Model of Recognition for Machine Perception, TR-78-3, Ph.D. Thesis, Department of Computer Science, University of British Columbia, Vancouver, Canada.

Havens, W.S. (1978b) MAYA Language Reference Manual, TM-13, Computer Science Department, University of British Columbia, Vancouver, Canada.

Kanade, T. (1977) Region Segmentation: Signal vs. Semantics, Proc. 3rd IJCPR.

Mackworth, A.K. (1975) Consistency in Networks of Relations, Artificial Intelligence 8, #1, pp. 99-118.

Mackworth, A.K. (1977a) On Reading Sketch Maps, Proc. 5IJCI, MIT, Cambridge, Mass, August 1977, pp. 598-606.

Mackworth, A.K. (1977b) How to See a Simple World, in Machine Intelligence 8, E.W. Elcock and D. Michie (eds.), Halsted Press, New York.

Mackworth, A.K. (1978) Vision Research Strategy: Black Magic, Metaphors, Mechanisms, Mini Worlds, and Maps, in Hanson and Riseman [1978].

Minsky, M. (1975) A Framework for Representing Knowledge, in The Psychology of Computer Vision, P. Winston (ed.), McGraw-Hill, New York.

Reddy, R. (1978) Pragmatic Aspects of Machine Vision, in Hanson and Riseman [1978a], pp. 89-98.

Rumelhart, D.E. & Ortony, A. (1976) The Representation of Knowledge in Memory, TR-55, CHIP, Dept. of Psychology, Univ. of Ca. at San Diego, La Jolla, Ca.

Sussman, G.J. & McDermott, D. (1972) Why Conniving Is Better Than Planning, AIM-255A, A.I. Lab, MIT, Cambridge, Mass.

Winograd, T. (1975) Frame Representations and the Procedural Declarative Controversy, in Representation and Understanding, D.G. Bobrow & A. Collins (eds)., Academic Press, New York, pp. 185-210.

MAPSEE2 COMPOSITION HIERARCHY

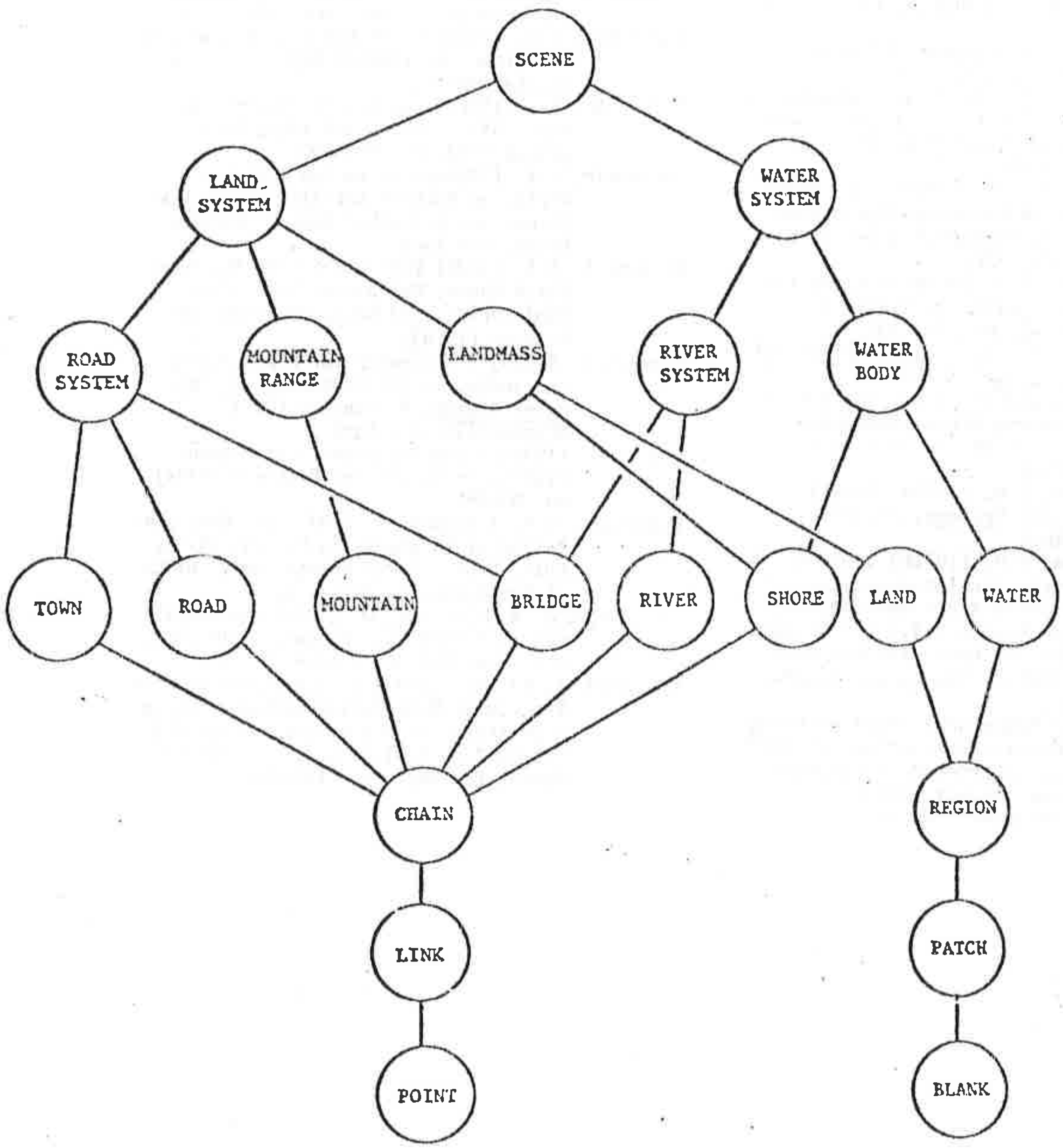
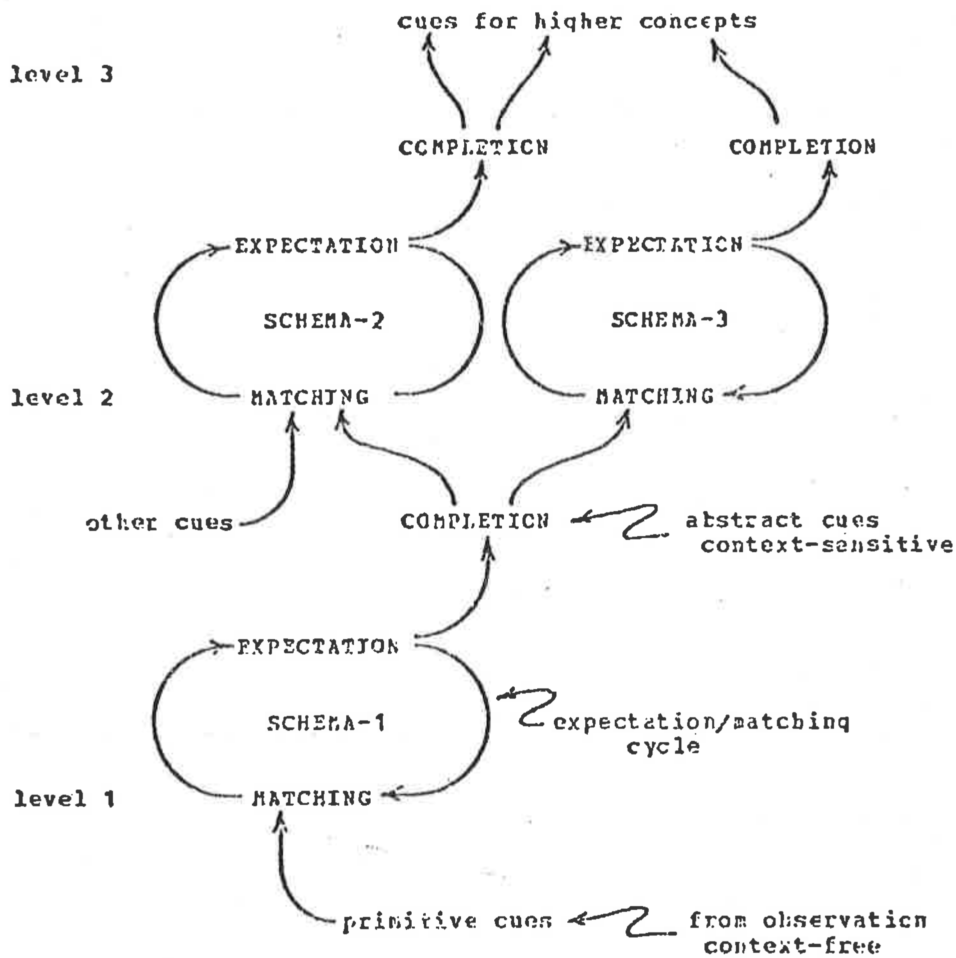


Figure 3



cues for higher concepts
level 3
COMPLETION
COMPLETION
EXPECTATION
SCHEMA-2
MATCHING
EXPECTATION
SCHEMA-3
MATCHING
level 2
other cues
COMPLETION
abstract cues
context-sensitive
EXPECTATION
SCHEMA-1
expectation/matching
cycle
level 1
MATCHING
primitive cues
from observation
context-free

Figure 4