STRUCTURING DOMAIN KNOWLEDGE FOR VISUAL PERCEPTION

Alan K. Mackworth

Dept. of Computer Science
Univ. of British Columbia
Vancouver, B.C.
Canada V6T 1W5

William S. Havens*

Dept. of Computer Sciences
Univ. of Wisconsin at Madison
1210 West Dayton Street
Madison, Wisconsin 53706

I. INTRODUCTION

Early visual processing is conjectured to be almost entirely domain-independent and data-driven; however, high-level vision must exploit domain and object specific constraints and methods, intermingling data-driven and model-driven recognition. We are concerned with knowledge representation and its use for high-level vision. This communication is a brief report on the theory underlying Mapsee2, a second program for interpreting sketch maps.

IT. WHAT'S WRONG WITH MAPSEE1?

This paper is best read as a sequel to an earlier vision paradigm embodied in the program Mapseel [3] . That program interprets sketch maps representing scenes containing roads, rivers, bridges, towns, mountains, islands, mainland, lakes and oceans. The paradigm suggests extracting cues from a lower level segmentation which are used to invoke, ambiguously, sets of object models. These models are then made to satisfy internal and external consistency constraints using a generalized network consistency constraint satisfaction algorithm. Here we shall be concerned not with praising Mapseel but with burying it. In our view, Mapseel died of representational inadequacy. It exhibits seven symptoms of that disease:

1. Its cue/model structure has but a single level.

2. Each model's domain of interpretation is a finite list of labels with no structure within or between the interpretations.

3. Many of the real world scene domain constraints are represented poorly if at all.

4. Its recognition process is essentially data-driven despite the grafting on of the "cycle of perception" and interpretation-driven re-segmentation.

5. The network of models is completely constructed before any model constrains another.

6. *The* model language is impoverished. In particular, local control of recognition is not possible.

7. Alternative scene interpretations are not explicitly available to guide the program.

The first three symptoms can be ascribed to descriptive inadequacy while the last four are symptomatic of procedural inadequacy.

III. THE KNOWLEDGE REPRESENTATION

In order to rectify these inadequacies we have a schema-based knowledge representation. To have a baseline to evaluate the success or failure of the project and to build on the advances of Mapseel, we have used the same scene domain.

At a theoretical level, Mapseel's contributions were to demonstrate the use of network consistency constraint satisfaction and the exploitation of data-driven, conservative segmentation techniques. In Mapsee2, we have concentrated on the interaction of <u>decomposition</u> and <u>specialization</u> in schema representations, exploiting <u>generalized network consistency</u> techniques and the homogeneous use of <u>data-driven</u> and <u>model-driven</u> interpretation strategies.

Our scene domain can be structured almost hierarchically, based on the composition or part-of relation as shown in Figure 1. A downward pointinq arc indicates that an instance of the object above is composed of zero or more instances of the object below. The decomposition is not a strict tree in that the same Town (instance), for example, may be part of both a Road-system and a Geo-system.

This work was supported in part by NSF Grant MCS-8004882 and NSERC Grant A9281.

* Now at Department of Computer Science, University of British Columia.

IV. SPECIALIZATION LABELLING

When an object is hypothesized to exist in the scene, an undifferentiated instance of its generic schema is created. As new components are recognized as part of the instance, its description becomes progressively specialized. At the same time, its possible semantic relationships with other schemata become further constrained.

Initially each incomplete schema instance must represent all possible final interpretations for that object. The schema is ambiguously labelled. In contrast, Mapseel employed explicit exhaustive label sets associated with each object. Unfortunately these labels must be low-level interpretations of the primitive lines and regions which appear in the scene. Attempting to encode high-level abstract relationships in these explicit label sets leads to an unavoidable combinatorial explosion of possible final interpretations.

In order to avoid this problem, we have developed an intensional labelling method for schemata representations called Specialization Labelling. The set of possible final interpretations (labels) • for a schema are represented as a tree. Each node of the tree represents an implicit number of final interpretations. Each descendant node represents a further specialization of the possible final labels implied by its parent. The root node of the tree is the undifferentiated schema instance. At any time in the recognition of the schema instance, the set of all possible final interpretations for the schema is represented by the union of all the labels implied by the current fringe of the tree.

Figure 2 shows the complete specialization tree for the Geo-system schema. Initially a Geo-system is fully undifferentiated, its labelling represented only by the schema instance itself. The recognition of the Geo-system involves constructing this tree as new components are added to its description (generally increases ambiguity of labelling) and pruning the tree as constraints are applied amonq related schemata (decreases ambiguity of labelling).

Specialization labelling avoids the problem of representing exponentially large numbers of labels for abstract high-level objects. As well, it provides a representation for manipulating semantically related labels as a single group.

V. THE RECOGNITION PROCESS

The recognition process is a search of the composition hierarchy for a complete and mutually consistent set of schemata instances. The search is guided by the procedural methods embedded in each schema. As each schema instance satisfies its own internal criteria for completed recognition, it is used as a cue for the hypothesis of higher schemata in the composition hierarchv. For example, when a Town instance has been found it in used as a cue for invoking the procedural methods of both the Road-System and Geo-System schemata above it in the hierarchy. In this way, appropriate high-level schemata can be hypothesized by the discovery of appropriate high-level cues, thus realizing a recursive cue/mode1 hierarchy.

The invoked higher schemata must find or make instances of themselves that can incorporate the lower instance or return failure if none can. Spatial adjacency is the usual constraint used to find an existing instance. Mutual consistency is then enforced between the higher and lower instances. Model-consistency modifies the description of the higher instance[1] to conform to the lower. The modification includes possible further differentiation of the schema instance down its specialization tree or failure. (Model consistency is a generalization of Waltz-like pruning of the corner label lists.) Object-consistency modifies the lower instance to conform to the upper. (This is a generalization of Mapseel-like pruning of the component object chain and region label lists). This consistency process may then be propagated to other schema instances related to these under the control of the relevant schema methods.

For example, suppose a Shoreline SI has been recognized as an interpretation of a chain closinq on itself. SI requires two Geo-system instances tc be part of, G-SI, the inside one, and G-S2, the outside one. Suppose G-SI already exists and is already known to be a Landmass, because it contains a Road-system. Model consistency of G-SI with respect to SI requires that G-SI be further specialized to Island. Object consistency of SI w.r.t. G-SI specializes SI to be Island-shore (as opposed to Lakeshore). Suppose the outside Geo-system for SI, G-S2, has to be created as a new undifferentiated Geo-system. Model-consistency of G-S2 w.r.t. the Island-shore SI specializes G-S2 to be a Waterbody. Object consistency of SI w.r.t. G-S2 succeeds. All of these specialization constraints may propagate to other related schema instances which may in turn continue to propagate them.

VI. CONCLUSION

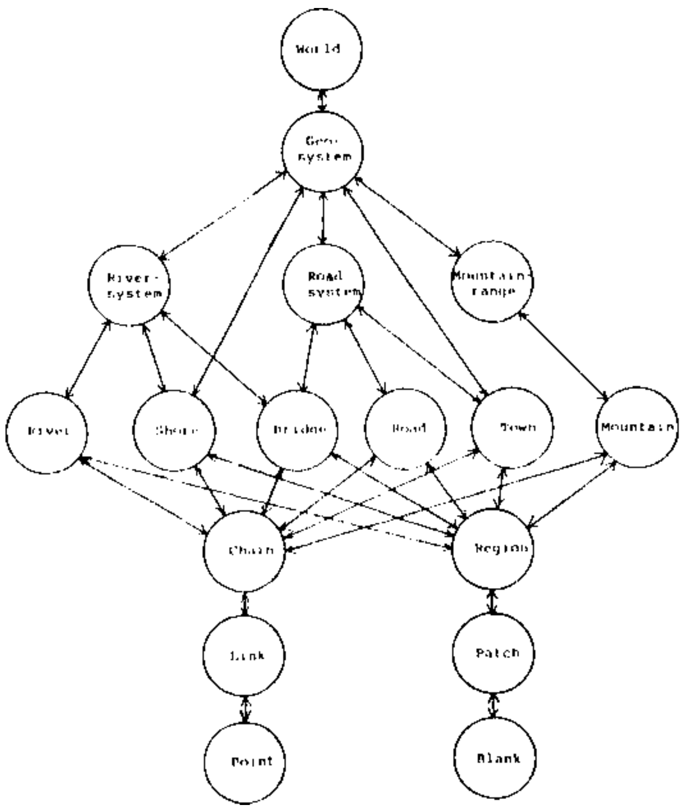
This model of perception is embodied in the program Mapsee 2. Mapsee2 is written in MAYA, a multiprocess LISP dialect supporting schemata data and control structures [l,2]. Our experience with it so far has been limited so this communication should be viewed merely as a brief progress report with tentative summary conclusions. In the meantime it should be clear that each of the inadequacies in Mapseel has been rectified in Mapsee 2. Our model is intended to provide a coherent framework for the desiqn and implementation of high-level, schema-based perceptual systems.

REFERENCES

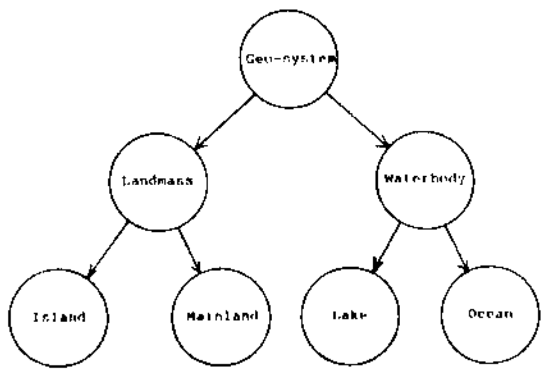
[I] Havens, W.S., "A Procedural Model of Recognition for Machine Perception", TR-78-3, Ph.D. Thesis, Department of Computer Science, University of British Columbia, Vancouver, Canada, (1978)

(21 Havens, W.S. and Mackworth, A.K., "Schemata-based Understanding of Hand-drawn Sketch Maps". In Proc. <u>3rd CSCSI Nat</u>. <u>Conf</u>., Victoria, B.C., May 1980, pp. 172-178.

[3] Mackworth, A.K., "On Reading Sketch Maos". <u>Proc</u>. <u>IJCAI 5</u>, MIT, Cambridge, Mass., August 1977, pp. 598-606.



Fiqure 1. Mapsee 2 Composition Hierarchy



Fiqure 2. Geo-system Specialization Tree