

USING MULTIPLE INFORMATION SOURCES IN A COMPUTATIONAL VISION SYSTEM

Jay Glicksman

Department of Computer Science
University of British Columbia
Vancouver, British Columbia*

ABSTRACT

One method of resolving the ambiguity inherent in interpreting images is to add different sources of information. The multiple information source paradigm emphasizes the ability to utilize knowledge gained from one source that may not be present in another.

A schema-based system has been developed that can take advantage of multiple sources of information. Schemata are combined into a semantic network via the relations decomposition, specialization, instance of, and neighbour. Control depends on the structure of the evolving network and a cycle of perception. Schemata cooperate by message passing so that attention can be directed where it will be most advantageous.

This system has been implemented to interpret aerial photographs of small urban scenes. Geographic features are identified using up to three information sources: the intensity image, a sketch map, and user guidance. The product is a robust system where the accuracy of the results reflects the quality and amount of data provided. Images of several geographic locales have been analyzed, and positive results are reported.

I INTRODUCTION

An information source is a set of input data. It can either be what is initially input to the system or an intermediate result that becomes available to later stages of processing. Standard input information sources (IS's) in Computational Vision are digitized images and line drawings. Other types of information can include multispectral images, range data, digital terrain models, sketch maps, and geographic data bases.

II THE CONJECTURE

The goal of image understanding is to determine the mapping from what is actually sensed, the image, to the scene of interest. Unfortunately, the problem is vastly underconstrained in general. Images are ambiguous and can be the result of an infinite number of scenes.

The hypothesis is that,

Multiple input information sources are useful in resolving ambiguities.

This is because similar aspects of the scene will be encoded in different ways by the different information sources.

III PREVIOUS RESEARCH

The usefulness of combining a second source of information with a problem solver was recognized in the early days of Artificial Intelligence research. For example, Gelertner's work on proving geometry theorems was aided by the information provided in a diagram which represents the problem (Gelertner, 1963).

In the field of Computational Vision, several researchers have argued for bringing "diverse sources of knowledge together" (Hanson and Riseman, 1978, p. 316; see also Flinchbaugh and Chandrasekaran, 1981, p. 391 and Witkin, 1981, p. 44). A classic use of combining several intensity images is found in pattern recognition where classification can be done by multidimensional cluster analysis. Using different spectral bands to find distinguished features is effective because some properties stand out in certain bands more than in others.

Work has been done at SRI to combine a map data base with a digitized image (Tenenbaum et al., 1978). The map data base contains three-dimensional locations and shapes of landmarks and monitoring stations. SRI researchers have produced a working system that demonstrates the feasibility of using this technique in several applications, such as measuring the volume of water in a reservoir and counting railway boxcars.

Disparate sources of information can be combined in systems that use evidential reasoning if the model knowledge has been suitably encoded (Wesley and Hanson, 1982). This has been done in the VISIONS system using a dependency graph and a model of belief that manipulates both support and plausibility. It has the ability to effectively handle information that is imprecise and uncertain.

IV A SYSTEM FOR COMBINING INFORMATION SOURCES

The MISSEE (Multiple Information Source SEE) system is a schema-based program that can take

*Author's current address:
Computer Science Laboratory
Texas Instruments
Dallas, Texas 75266

advantage of multiple information sources to interpret aerial photographs of small urban areas (Glicksman, 1982a). Figure 1 shows an overview of the components of MISSEE. The core of the system is a semantic network of schemata which both controls the interpretation process and builds the resulting representation. The schemata are created and utilized by the schemata manipulation subsystem (called MAIDS--MISSEE Aids). The three input information sources that feed into MISSEE are a digitized image, a sketch map, and user guidance.

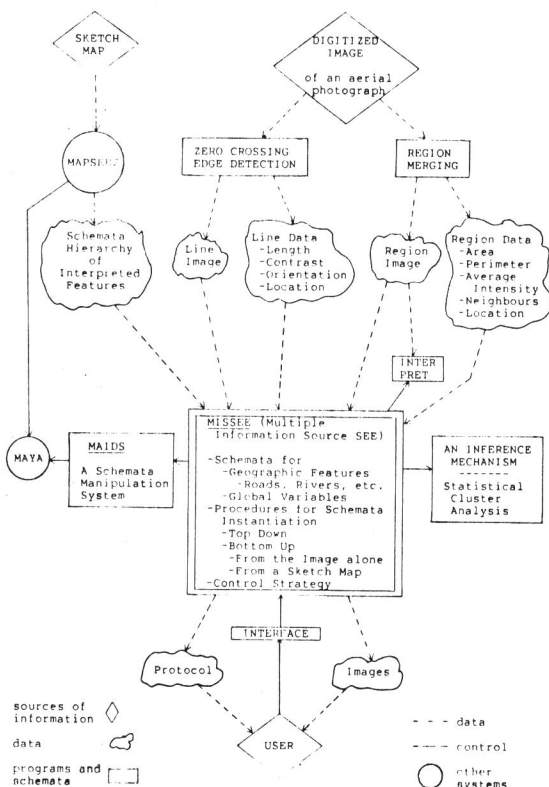


Figure 1. The MISSEE System

MAIDS (Glicksman, 1982b) is a schemata manipulation system built on top of Maya (Havens, 1978). Schemata are made up of four distinct types of attributes. The VALUE type is used to store the facts that define the represented object. It is an extension of Maya's attribute type to include FRL's slot facets and more (Roberts and Goldstein, 1977).

The schemata are contained in a semantic network with relationships indicated by LINKS. Any number of binary relations can be formed, and the system maintains the pointers and their inverses and allows for several forms of automatic traversal of the network. LINKS can be used with VALUED types to generate property inheritance.

Another distinguished part of a schema is its procedural attachments. Procedures are used to instantiate schemata (i.e. fill in their slots) and control the interpretation. This would include procedures to invoke the schema as a model (top-down) or to account for data (bottom-up). The

procedures would also typically send out messages which would cause other schemata to become "current" and thus evaluated.

The basic control mechanism in MISSEE is a cycle of perception (Neisser, 1976). In the three-node cycle, "objects" in the image cause "schema invocation and instantiation" which in turn directs "exploration" in the image where "objects" are again sampled. The cycle of perception is similar to other feedback paradigms such as hypothesize and test. It is particularly well-suited to a schema-based system because it identifies the natural places where schemata can move into and out of the focus of attention.

The cycle of perception promotes cooperation among the schemata. As objects are instantiated they communicate to other schemata (by sending messages) to give advice or to build the semantic network. This in turn initiates another loop around the cycle as a new schema is instantiated. The schemata work as individuals (via attached procedures) and together (via messages) to bring about the interpretation of particular information sources.

V THE THREE INFORMATION SOURCES

The major source of information input to the system is a digitized image of an aerial photograph. This information source is exploited by describing it in terms of edges and regions.

Edges are found using a Marr-Hildreth edge detector (Marr and Hildreth, 1980). Region merging is accomplished by a method described by Freuder called affinity (Freuder, 1976). The ambiguous results these algorithms produce require model-dependent information to generate informed interpretations.

Sketch maps are analyzed by Mapsee2 (Havens and Mackworth, 1980). Input is a series of moves/draws which trace the curves of the sketch. Output is a decomposition hierarchy of schemata instances of geographic objects ranging from points and lines to roads, rivers, road systems, and geosystems.

Sketch maps can guide the interpretation of the digitized image by reducing the number of possible models that must be matched against image features. Additionally, spatial information from the sketch map reduces the amount of search that must take place to find the appropriate image features.

The user's means of providing information and influencing the interpretation is via a priority queue used in the cycle of perception. While the user can modify the interpretation process so that results exactly suit his/her purposes, in a robust system such as MISSEE, it is not necessary for such interaction to take place.

VI RESULTS

MISSEE has been tested with 6 images of small urban areas. One measure of success for an image understanding system is the number of image

features (either edges or regions in this case) that have been correctly associated with geographic entities. When the only information source available is the digitized image, there are two types of errors that can arise: either a feature in the image is given an incorrect interpretation or a scene object did not lead to the right features to cause instantiation (i.e. it was omitted). In the test examples, there were 61 incorrect and 25 omitted objects out of 360. This leads to an error rate of 24%.

Further results were found when the information from an analyzed sketch map is made available to guide interpretation of the digitized image. Since curves (chains) in the sketch map may be ambiguous, their interpretations can be either intended (i.e. what the person who drew them intended them to be) or unintended. Furthermore, each sketch map instance can lead to either a corresponding instantiation in the image or it can remain uninstantiated. Thus there are four possibilities, two of which are correct and two incorrect. It is correct when 1) an intended interpretation is instantiated (54 cases), or 2) when an unintended interpretation is not instantiated (37). The two erroneous situations arise when 1) the intended model could not be instantiated in the image (3 cases), or 2) an unintended one could (6).

As expected, the results using a sketch map are better than when one is not available. The error rate has decreased from 24% to 9%. Similar results are found for the number of regions that have to be searched during the instantiation process. When sketch maps are used, 22% of the regions have to be examined compared to 41% when only the images are used.

VII DISCUSSION

MISSEE is a schema-based vision system that has been implemented to test the usefulness of the multiple information source conjecture. By using the information that is available to interpret images, it has proved effective in terms of adequacy and robustness. The design of MISSEE led to the development of several tools which are useful in reconciling different types of information to solve many of the problems in model-based vision. These tools plus the schema-based representation might also be fruitfully applied to non-perceptual domains that try to combine information sources.

An open question is the expandability of this system. Since entities are all related by the semantic network, it may be difficult to splice in new objects if they disturb the old relationships. Furthermore, some objects (e.g. clouds) do not neatly fit into the geographic hierarchies that are currently being used. A meta-level, network restructuring component could help alleviate these problems.

Schema-based systems are capable of sufficiently describing the attributes of geographic entities for their interpretation, as well as directing the interpretation process itself. Exploiting

information sources in a cooperative fashion gives results qualitatively better than those obtainable from interpreting the information sources separately.

REFERENCES

- Flinchbaugh, B.E. and B. Chandrasekaran, "A Theory of Spatio-Temporal Aggregation for Vision." Artificial Intelligence 17:1-3 (1981) 387-407.
- Freuder, E.C. "A Computer System for Visual Recognition Using Active Knowledge," Technical Report AI-TR-345, AI Lab, MIT, June 1976.
- Gelertner, H. "Realization of a Geometry-Theorem Proving Machine" In E. Feigenbaum & J. Feldman (eds.) Computers and Thought. New York: McGraw-Hill, 1963, 134-152.
- Glicksman, J. "A Cooperative Scheme for Image Understanding Using Multiple Sources of Information," Technical Report TN 82-13, Dept. of Computer Science, U.B.C., November 1982.
- Glicksman, J. "A Schemata-Based System for Utilizing Cooperating Knowledge Sources in Computer Vision" In Proc. Fourth Conf. CSCSI. Saskatoon, Canada, 1982, 33-40.
- Hanson, A.R. and E.M. Riseman "VISIONS: A Computer System for Interpreting Scenes" In A.R. Hanson and E.M. Riseman (eds.), Computer Vision Systems. New York: Academic Press, 1978, 303-334.
- Havens, W.S. "A Procedural Model of Recognition for Machine Perception," Technical Report TR-78-3, Dept. of Computer Science, U.B.C., March 1978.
- Havens, W.S. and A.K. Mackworth "Schemata-Based Understanding of Hand-Drawn Sketch Maps" In Proc. Third Conf. CSCSI. Victoria, Canada, 1980, 172-178.
- Marr, D. and E.C. Hildreth, "Theory of Edge Detection." Proc. Royal Society of London B:207 (1980) 187-217.
- Neisser, U. Cognition and Reality: Principles and Implications of Cognitive Psychology. San Francisco: W.H. Freeman, 1976.
- Roberts, R.B. and I.P. Goldstein "The FRL Manual," Technical Report AIM 409, AI Lab, MIT, Sept. 1977.
- Tenenbaum, J.M., M.A. Fischler, and H.C. Wolf "A Scene-Analysis Approach to Remote Sensing," Technical Report TN 173, AI Center, SRI, Oct. 1978.
- Wesley, L.P. and A.R. Hanson "The Use of an Evidential-Based Model for Representing Knowledge and Reasoning About Images in the VISIONS System" In Proc. Workshop on Computer Vision: Representation and Control. Rindge, New Hampshire, 1982, 14-25.
- Witkin, A.P., "Recovering Surface Shape and Orientation from Texture." Artificial Intelligence 17:1-3 (1981) 17-45.