A Goal-Directed Intermediate Level Executive for Image Interpretation

Charles A. Kohl Allen R. Hanson Edward M. Riseman

Department of Computer and Information Science
University of Massachusetts
Amherst, Massachusetts 01003

ABSTRACT

GOLDIE is a system that has been developed to provide the top-down control of the low and intermediate level processes that create and modify the intermediate-level descriptions of image data used in image interpretation. The basic control structure of GOLDIE is the *schema,* a declarative specification of *control* strategies invoked in response to a request *(goal)* for a particular form of intermediate-level data. Using this control paradigm, high-level interpretation processes gain the ability to create or refine the intermediate-level data according to pre-defined goals and/or current hypotheses during interpretation. A variety of schemas for tasks such as region segmentation, line extraction, collinear line grouping, and line-based segmentation are currently implemented within the GOLDIE system.

1. THE INTERMEDIATE LEVEL OF INTERPRETATION

Due to the complexity of the visual interpretation task and the inherent unreliability and ambiguity of image data, completely bottom-up (i.e. data-directed) approaches to interpretation can not be expected to be generally effective. The local nature of the processes that extract descriptions of image events typically leads to intermediate level descriptions of the image data in which individual semantic objects from the scene are either broken into a number of pieces (fragmentation), or are merged with other objects in the scene (overmerging) (4,11). The use of expectations, or context, provided by the systems's knowledge-base and/or partial scene interpretation can be used to compensate for these factors to produce data descriptions that more closely match the semantic content of the scene. Such a methodology requires that control of processing be bi-directional, shifting between data-directed and goal-directed methods in an opportunistic manner [1,5,9,11]. This implies that most of the processes utilized by a vision system during the construction of an interpretation must be sensitive to high-level goals and constraints.

This concept of bottom-up and top-down control may be viewed in terms of the levels of abstraction used in the image interpretation process. Within the VISIONS Image Understanding System (8,11), three such levels may be identified. Low-level processes are those which operate on pixel data to produce a set of intermediate-level symbolic *tokens* representing significant events contained in the raw image data (e.g. regions, lines, or surfaces). Intermediate-level grouping processes manipulate this initial set of tokens to produce new tokens, and high-level processes provide the semantic interpretation of the image based on the full set of image tokens.

GOLDIE (fifial Directed Intermediate-Level Exectutive) [6,7] is a system within the VISIONS environment that provides the mechanisms for the top-down control of the low and intermediate-level processes that create or modify tokens. The specification of a request for a particular type of image token is expressed as a *goal,* a data structure that defines the generic class of processing required (e.g. region segmentation, line extraction, grouping, etc.). *Constraints,* stored as attributes within the goal data structure, express the desired characteristics of the tokens to be produced, and may be represented through either semantic or image-based criteria. Semantic constraints are defined in terms of semantic labels (e.g. *"segment a specific portion of the image to separate Tree and Sky"),* while the image-based constraints are expressed in terms of measurable image features (e.g. *"produce regions that exhibit homogeneous texture measures").*

The basic control structure of GOLDIE is the *schema,* a declarative specification of *control* strategies that may be used by the system to satisfy a specific goal. Schemas provide a flexible and extensible control structure that is currently used within VISIONS to direct the high-level semantic interpretation processes [4,10,12]. GOLDIE extends this concept of schema-directed control to the intermediate and low levels of processing. The power of this system lies not in the particular set of specific low and intermediate-level processes that have been integrated into the current system, but rather in the various representations for knowledge, data, and control that implement goal-directed processing.

2. GOAL-DIRECTED CONTROL

GOLDIE supports both goal-directed and data-directed control of processing. Data-directed processing is controlled by sets of evaluation rules which associate *hypotheses* (e.g. "candidate for resegmentation", "candidate for merge", etc.) with tokens; the hypotheses are used to direct the subsequent activities of the system. Current goal constraints are used to determine the specific set of rules used to compute the value of the hypothesis. For example, if a particular set of goal constraints indicated that uniformly textured regions were of primary importance, the rules used to establish the acceptability of region tokens would include measures of the homogeneity of both hue and short line density. Tokens representing large regions with high variance in either of these features would probably be considered unacceptable.

Various types of intermediate-level hypotheses may be established in this manner. For example, the *initialization* schema, which provides a data-directed region segmentation for either the entire image (at system startup) or for an area defined by a (set of) region token(s), uses the rule mechanism to hypothesize whether individual region tokens should be merged or reaegmented. An instance of this schema first uses the *region segmentation* schema to select the image feature(s), segmentation algorithm, and parameters which will be used to produce an initial "best guess" for a set of region tokens. Using the set

*ThIs work was supported in part by the Air Force Office of Scientific Reiearch Grant AFOSR-86-0021 and the National Science Foundation Grant DCR.8318776

of rules selected according to the goal constraints, the initialization schema instance then evaluates this set of tokens and continues to process those which have unacceptable values for

Figure 1: Outdoor Scene

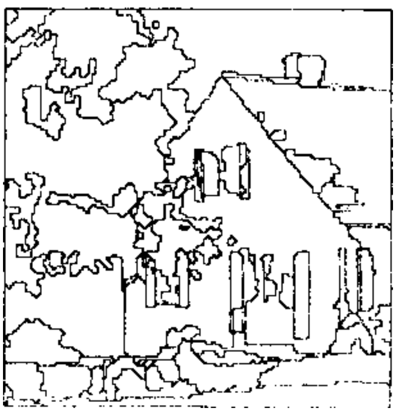

Figure 2: Initialisation Schema Segmentation

region-merging-hypothesis or region-resegmentation-hypothe8iB As each of these tokens is processed, the schema is able to perform resegmentation or merging using algorithms and criteria *appropriate to the characteristics of the particular region.* This form of data-directed processing demonstrates intermediate level control that uses knowledge of region characteristics, as well as knowledge about the performance of the segmentation algorithms under varying conditions of the data; we call this *non-semantic* (or *image-domain)* knowledge Figure 2 shows the segmentation produced by this schema for the image from Figure 1.

Although this form of intermediate-level control provides image tokens that appear to correlate reasonably well with image content, this fact alone does not guarantee the suitability of the tokens for interpretation. *The ultimate significance or quality of a particular token is dependent only upon its utility with respect to a set of interpretation goals and cannot be measured independent of the interpretation process.* Therefore GOLDIE also supports goal-directed control in which a goal contains an explicit specification of the token characteristics which must be met by any potential set of results. This *evaluation-constraint* is expressed as a function value pair; the function is applied to each set of tokens produced by the contracting schema in-

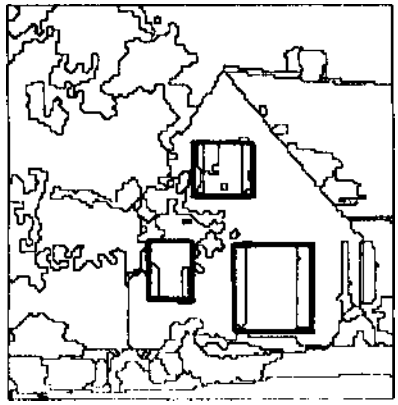

Figure 3: Rectangular Shutter Regions

stance, and only if the return value of the function exceeds the value specified is the set of tokens judged to be potentially acceptable

For example, if there has been an inference of the presence of a pair of shutters on the side of a house, an interpretation process might request a resegmentation of the area defined by that pair in an attempt to define tokens that could be unambiguously labeled as either window or shutter The goal specification for this request

```
                Region-Segmentation
         '((region-token   Region-000042)
(evaluation constraint  ((shape-evaluation 'rectangular) 1.0))
  (object-labels  (shutter))(region-characteristic  smooth))
```

includes an evaluation-constraint that indicates a preference for region tokens that are rectangular Figure 3 shows the result of this goal specification over the three areas bounded by pairs of initially hypothesized shutters in the original segmentation (Figure 2) The dark lines in this figure represent region boundaries which were constructed by GOLDIE in response to the definition of an "area of interest"[1] by the interpretation process Each of these newly defined regions was then resegmented according to the specified goal. Note that although image noise, aliasing, and window reflections prevent a "perfect"[1] segmentation of these areas, the resegmentation process has provided a set of tokens which are appropriate for interpretation of windows and shutters. The tokens produced through the satisfaction of this goal can be labeled and precisely located, thereby making the information generated by this particular interpretation process available to any other high-level processes that may be concerned with this data.

Conversely, if the intent of the interpretation process is to identify trees in the image, the interpretation system could specify a merge goal having an evaluation-constraint that indicates a preference for large green textured areas (thereby producing the region in Figure 4).

3. ARCHITECTURE OF GOLDIE

The GOLDIE system, represented by the modules within the large dashed rectangle of Figure 5, may be described in terms of four major functional components: the process controller, the data structures for intermediate-level tokens and hypotheses (ISTM), the representation of explicit intermediate-level knowledge (ILTM), and the representation of the control state of the system (Goal Blackboard and Schema Instantiations).

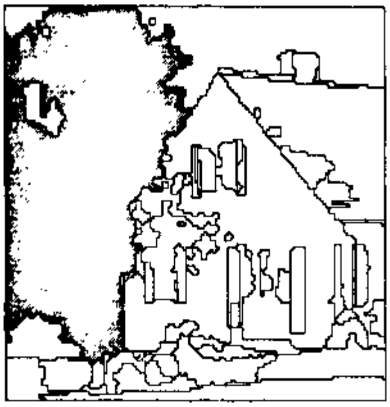

Figure 4: Region Produced with Goal for Tree-like Regions

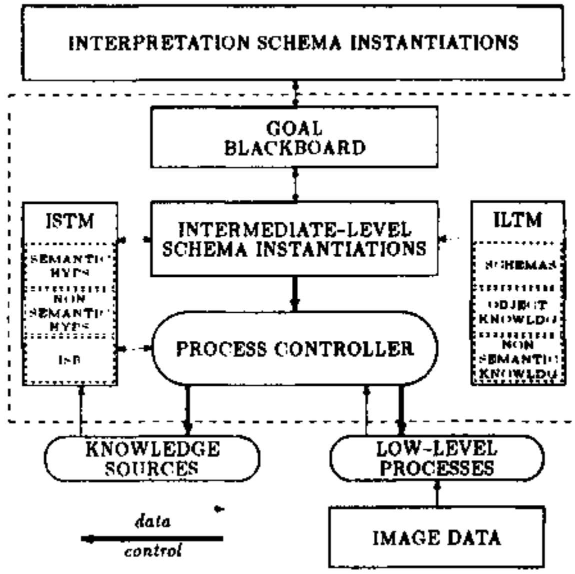

Figure 5: Overview of the GOLDIE System

The set of low and intermediate-level processes that arc utilized by GOLDIE to create, manipulate, or evaluate image tokens are managed by the *process controller.* The controller maintains a consistent protocol for the activation of both the low-level processing tasks that operate on pixels (e g threshold segmentation, one or two dimensional histogram-based segmentation, feature calculation, line extraction, etc) and the intermediate-level knowledge sources that operate on tokens (e.g. collinear line grouping, token evaluation, or region merging) This interface permits the schema representation of these processes to be expressed in a manner independent of the process implementation Note that the processes themselves (both low-level image-based procedures and intermediate-level knowledge sources) are not considered to be an integral part of GOLDIE; rather they represent the set of tools available to the system

The *data structures* of GOLDIE that are particular to a specific interpretation are the sets of tokens and hypotheses about tokens that are stored in ISTM (Intermediate-level Short Term Memory). The tokens are defined through the ISR (Intermediate Symbolic Representation), a combination database and semantic network in which tokens are defined as graph nodes with associated bitmaps representing the spatial extent of the token in the image. Attributes of the tokens, such as intensity or size, are computed on demand by the ISR and then stored in the database for efficient retrieval by any other process that may require the same data. Specific relations between tokens, which have been requested by some high or intermediate-level process (e.g. the boundary contrast between two adjacent region tokens), are represented as arcs that have values indicating the nature of the relation. The sets of semantic (e.g. tree-object) non-semantic (e.g region-merge) hypotheses are stored in a similar database structure which is accessible from both the intermediate and high levels of processing

ILTM (the Intermediate-level portion of the VISIONS Long Term Memory) encodes the image-independent intermediate-level knowledge structures of GOLDIE which include schemas, object knowledge, and non-semantic image knowledge. Schemas are represented declaratively as a set of strategies by which a particular goal, such as region segmentation or line extraction, may be satisfied. The execution of the strategies will typically involve a sequence of operations including the posting of subgoals, the execution of image processing tasks through the process controller, and the construction of hypotheses based on data represented in the ISTM. As an example, the *region-feature* schema, which selects an image feature for a region segmentation, includes strategies for the selection of features based on frequency distributions of image data, on semantic hypotheses, and on textural characteristics. Message-passing constructs allow the various executing strategies to communicate and synchronize their activities within the context of a particular schema instance.

The object knowledge of ILTM encodes the intermediate-level characterization (e.g. textured or smooth) of various semantic objects as well as process specific information that can be useful in the specification of the processes that are to produce tokens corresponding to these objects (e.g. image features or segmentation algorithms that have been shown to be useful in the discrimination of the object from other objects). Non-semantic knowledge includes a set of useful characterizations for tokens (e.g. textured-region, smooth-region, high-contrast-line, etc.) as well as the set of rules, image features, segmentation algorithms, and segmentation parameters appropriate to each of these token classes.

The goal blackboard and the set of intermediate-level schema instantiations are represented in a semantic network that encodes the *control state* of the system [2,3,12]. The goal blackboard is a section of the network in which goal nodes, representing requests for intermediate-level processing, are created by the schema instance that requires the results of that processing. The *control process* of the GOLDIE system continually monitors this goal blackboard for the existence of new goal nodes. As a new goal is observed, the control process creates a new instance of the corresponding schema. The node for this schema instance is linked to the goal node by an arc that represents a contract by the instance to attempt to satisfy the goal.

A typical scenario for the interaction between GOLDIE and the interpretation system is a situation where a high-level process had formed two conflicting hypotheses regarding the semantic label to be assigned to a particular region token (e.g. tree and sky for the cross-hatched region on the left of Figure 6). Since this is a region in which sky is visible through the leaves of the tree, both hypotheses are reasonable. However, an interpretation process which was to unambiguously label the image tokens could post a goal for region-segmentation.

Region-Segmentation
'((*region-token* . Region-000017)(*evaluation-constraint* . nil)
(*segmentation-sensitivity* . high))(*object-labels* . (tree sky))
(*objective* . object-hypothesis-conflict))

In response to the posting of this goal, the system would activate an instance of the region-segmentation schema This schema instance would then select image features and algorithms appropriate for the discrimination of the two objects and then initiate a segmentation process which would be expected to split the original region into a set of new regions, each corresponding to only one of the two objects. Tokens corresponding to these new regions would then be returned to the interpretation schema instance for evaluation (right side of Figure 6). If the interpretation schema instance were satisfied with the results, the goal and region-segmentation schema instance would be deleted. Otherwise, the interpretation schema instance would send a continuation request to the segmentation instance so that additional region tokens could be produced and evaluated. In this case, the region-segmentation schema instance would continue processing until an acceptable segmentation were found or until all of its strategies were exhausted

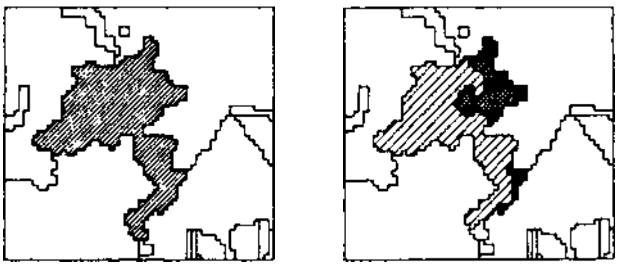

Figure 6: Tree/Sky Region

4. CONCLUSION

A variety of schemas for tasks such as region segmentation, line extraction, collinear line grouping, and line segmentation are currently implemented within the GOLDIE system at the University of Massachusetts, and research is proceeding on the interface between this system and the high-level interpretation schemas being developed by other researchers in the VISIONS group (2). As currently implemented, the GOLDIE system provides an evolving mechanism for the control of the processes that create or modify the intermediate-level data representation for the interpretation process.

There are four major aspects of GOLDIE that contribute to the utility of the system. Foremost is the fact that the system is goal-driven. This paradigm provides a coherent mechanism for top-down control in which low or intermediate-level processes may be tuned to produce tokens that meet the expectations of higher level processes. Secondly, the system makes use of a unified data representation that allows processes at any level of the interpretation hierarchy to access the token and hypothesis data which represents the current state of interpretation processing. Thirdly, the system contains an explicit representation of knowledge about low-level processes and the image domain which provides a flexibility that permits extension to additional processes or image domains. Finally, the GOLDIE system is capable of operating within a hypothesize and -test protocol, exploring a variety of potential solutions to high-level requests rather than operating in a strictly deterministic manner.

REFERENCES

[I] Rodney A. Brooks and Thomas O. Binford. Interpretive vision and restriction graphs In *Proceedings of The First Annual National Conference on Artificial Intelligence,* pages 21-27, August 1980.

|2| B. Draper, R. Collins, J Brolio, J. Griffith, A Hanson, and E. Rise man. Tools and experiments in knowledge-based interpretation of road scenes In *Proceedings: DARPA Image Understanding Workshop,* pages 178-193, Morgan Kaufman, February 1987.

|3| B. Draper, A Hanson, and E Riseman. *A Software Environment for High Level Vision.* Technical Report, Computer and Information Science Department, University of Massachusetts at Amherst, 1986 In preparation.

[4] Allen R. Hanson and Edward M. Riseman VISIONS image understanding system - 1986. In Christopher M Brown, editor, *Advances m Computer Vision,* Erlbaum Assoc, 1987 To appear

|5| Vincent Shang-Shouq Hwang, Larry S Davis, and Takashi Matsuyama *Hypothesis Integration m Image Understanding Systems* Technical Report TR-1137, Center for Automation Research, University of Maryland, College Park, Maryland 20742, June 1985

(6| Charles Kohl, Allen Hanson, and Edward Riseman Goal-directed control of low-level processes for image interpretation In *Proceedings. DARPA Image Understanding Workshop,* pages 538-551, Morgan Kaufman, February 1987.

|7] Charles A. Kohl. *Goal-Directed Control for Computer Vision* PhD thesis, Computer and Information Science Department, University of Massachusetts, Amherst, Massachusetts 01003, 1987. In Preparation.

|8] R R. Kohler and A. R. Hanson. The VISIONS image operating system. In *Proceedings Sixth International Conference on Pattern Recognition,* pages 71-74, Munich, October 1982.

|9| David Marr. *Vision.* W. H. Freeman, San Francisco, CA, 1982.

(10| Cesare C. Parma, Allen R. Hanson, and Edward M. Riseman. Experiments in schema-driven interpretation of a natural scene. In R. Haralick and J. C. Simon, editors, *NATO Advanced Study Institute on Digital Image Processing,* Bonas, France, 1980.

(11) Edward M. Riseman and Allen R. Hanson. *A Methodology for the Development of General Knowledge-Based Vision Systems.* Technical Report 86-27, Computer and Information Science University of Massachusetts at Amherst, Amherst Massachusetts 01003, 1986.

[12) Terry E. Weymouth. *Using Object Descriptions in a Schema Network for Machine Vision.* PhD thesis, University of Massachusetts, Amherst, Massachusetts 01003, 1986.