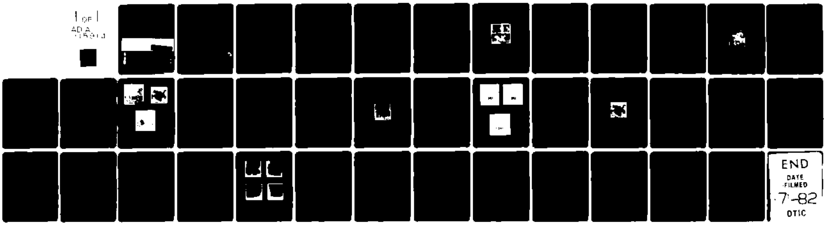
AD-A115 914 MASSACHUSETTS UNIV AMHERST DEPT OF COMPUTER AND INF--ETC F/G 9/4
STRATEGIES FOR KNOWLEDGE-BASED IMAGE INTERPRETATION.(U)
MAY 82 C MCCORMICK N00014-75-C-0459
UNCLASSIFIED COINS-TR-82-10 NL



(12)

AD A115914

Computer and Information Science

(12)

STRATEGIES FOR KNOWLEDGE-BASED*
IMAGE INTERPRETATION

Clif McCormick

COINS Technical Report 82-10

May 1982

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

Abstract

Strategies suggested by general observations can be used to control processing in a knowledge-based image interpretation system. Several strategies are discussed and experiments are presented to illustrate their use.

*This research was supported in part by the National Science Foundation under Grant MCS79-18209 and the Office of Naval Research under Grant N00014-75-C-0459. The author acknowledges support from a University of Massachusetts Fellowship.

I. INTRODUCTION

The goal of the VISIONS project is to develop a system that can interpret static color images of outdoor scenes. [HAN78a,b] The interpretation task consists of labeling the various objects in an image and describing the relationships among them. This task is difficult, given the complexity and variety inherent in the domain. The set of objects and possible relations is large, lighting varies, exact camera models are often not available, shadows and occlusion obscure the shapes of objects, and seasonal changes introduce spectral and textural variety. A great deal of knowledge must be brought to bear in understanding images of outdoor scenes.

A large part of this knowledge concerns the set of objects that can and do appear in these images and the possible and probable relations among them. In order to understand the images, detailed information about the distinguishing characteristics of each object class must also be available. This paper presents preliminary results showing how various strategies utilizing four types of simple features--size, shape, color, and location--can be used to recognize objects and form the basis of a simple interpretation system.

The size of an object can aid in its recognition. However, in images absolute sizes are rarely available and furthermore members of an object class often appear in a range of sizes. Thus size is most important in relative terms: objects can be recognized using the sizes of reference objects that have been located in the image.

Characteristics of an object's (2-D) shape provide recognition cues. Curvature, compactness, height-to-width ratio, rectangularity-- these are a few shape features that may help to distinguish objects. For example, certain man-made objects (windows, doors, shutters) exhibit high rectangularity, little curvature, and a vertical orientation (greater height than width). However, representing and recovering complex shape characteristics is very difficult. [YOR81] [KEN80]

Color or spectral features are especially useful in identification, particularly for "natural" objects such as grass, sky, and foliage whose color tends to be more predictable than that of man-made objects such as cars and houses. Spectral features include the red, green, and blue components of an image element's intensity, color transforms, and simple texture measures. Sets of these features can be used to characterize different objects. Sky, for example, tends to have a high blue component value but a low saturation value. Foliage, on the other hand, tends to be quite saturated. In this case, saturation is used in recognizing foliage because it not only CHARACTERIZES an aspect of all foliage but also DISCRIMINATES foliage from other objects.

Location plays a part in object recognition. The location of certain objects can often be predicted: sky often appears at the top of an image; grass, road, or ground often appear at the bottom. As was the case with size, these object location features can be used not only to identify possible object classes but also to eliminate other object classes from consideration. Location can also be characterized in relative terms, providing identification information via expected spatial relations among objects. In general, the information characterized by the four types of features-- size, shape, color, location-- is important not only in absolute but also in relative terms (in the form of relations). Objects are often identified using other objects as references. This observation implies that object recognition can be carried out in at least two ways: simply by listing the expected feature values of an object class and searching for a match (a local approach) OR within a context using some kind of strategy that operates on the feature values (a global approach). Matching alone does not seem to be sufficient for most recognition tasks. Thus it is clear that the process of object identification should consist of a variety of strategies operating on the types of feature information outlined above.

The experiments presented below demonstrate the utility of various strategies operating on feature information in developing an image interpretation. The strategies are simple; each can be "fooled" in certain cases. Used together, however, they provide a fairly robust foundation for a first pass interpretation system.

II. EXPERIMENTS

The four 128 x 128 images of house scenes used in the experiments are shown in Figure 1. Three of the images are different views of the same house. The information gathered during the interpretation of one of these three images could be used to guide the interpretation of either or both of the other images, assuming similar or identical lighting conditions. Such an approach would be especially useful in motion processing. In the experiments presented below, however, interpretation strategies have been applied independently to each of the images.

The domain of house scenes is fairly complex, yet it is manageable in that the set of commonly occurring object types is not too large (less than 20), and there are a variety of structural and relational constraints that can be exploited in object recognition. For example, with many houses, windows are constrained to be located between two shutters. This type of constraint generates predictions about the existence and location of certain objects based upon a partial interpretation and can be incorporated into strategies for both hypothesis formation and hypothesis verification

Certain assumptions have been made in the experiments. The system assumes a camera position that is approximately level so that the horizon is expected to be near the center of the image;

Figure 1. The four 128×128 images used in the experiments. Label them 1-4 starting in the upper left and proceeding clockwise. Image 1 is used in most of the other figures.

this assumption allows the system to predict the extents of sky and ground regions. The second assumption is that the spectral attributes of objects are fairly typical; for example, grass is green rather than brown. This assumption allows reasonable hypotheses of object identities to be developed using expected spectral attributes of objects. Finally, the system assumes that a good segmentation has been provided for establishing a correspondence between regions and object surfaces.

II.1 IMAGE-INDEPENDENT SPECTRAL ATTRIBUTE MATCHING

Spectral attributes can be used to characterize certain "natural" objects--bush, grass, sky, tree--whose features are fairly predictable. There are also certain classes of man-made objects whose color and texture are predictable, such as roads, sidewalks, fire hydrants, and stop signs. The simplest use of color and texture attributes consists of matching the expected feature values of an object class with those of image regions to form hypotheses of object identity. The technique of object to region matching of attributes that is presented below has been used previously and is described only briefly here. (See [WIL81].)

Given a set of features and set of training images of outdoor scenes, the mean, standard deviation, maximum value, and minimum value of each feature were computed using hand-selected regions known to represent the "natural" objects mentioned above. These statistics were used to form prototype templates of the ranges of feature values for each object class. The matching process consists of forming a confidence by comparing the feature values of a region to the feature values of each of the templates. The confidence value obtained symbolizes a hypothesis that a certain region represents a certain object or object part. Maximum confidence is assigned to a region whose mean feature value is within one standard deviation of the expected mean for an object. The confidence decreases linearly to zero at the minimum and maximum values.

The results of spectral attribute matching in the four images are summarized in Table 1. Tree, grass, and sky regions are identified fairly accurately. Bush regions were most often misclassified as tree, accounting for six of the eight bush regions incorrectly labeled and generating six false alarms for tree. It is not unreasonable for a system to make errors between different classes of foliage when the classification is based purely on local features. Grouping tree and bush under a category of "foliage" produces better results, with 23 of 26 target regions being correctly identified. The portions of the image that are correctly labeled are shown in white in Figure 2.

Because this matching strategy only deals with a restricted subset of the objects commonly occurring in outdoor scenes, regions representing objects not in the subset are always labeled

Table I

Attribute Matching Results (large regions from 4 images)

	Actual	Correctly Identified	False Alarms
Bush	12	4	1
Grass	6	5	3
Sky	5	5	0
Tree	14	12	7
Foliage	26	23	–
TOTAL (using foliage)	37	33	

Figure 2. Portions of Image 1 correctly labeled by spectral attribute matching alone are shown in white.

incorrectly. In many cases the confidence assigned to this erroneous labeling is sufficiently low (compared to expected values) that the labeling can be rejected; in other cases the confidence is relatively high and a labeling error results. For example, in the images of Figure 1, the white house walls "acquired" many of the spectral characteristics of sky and hence are often interpreted as sky. In cases such as this it is unreasonable to expect the system to distinguish between high match value non-target regions whose color and texture attributes are similar to those of the target object prototypes and actual target regions. While it is possible that better results could be achieved by formulating the target vs. non-target problem as a classical statistical hypothesis testing problem, it is conjectured that many of the erroneous labels may be eliminated by the application of labeling constraints derived from the relationships between objects and the structural properties of objects appearing in the scene. Experiments described later are a first attempt to show how this may be accomplished.

Spectral attribute matching is computationally inexpensive if the object training data has been analyzed previously. If one ignores the errors involving confusion of foliage categories and the problems of high match value non-target regions, then the approach has yielded excellent results. It might be made still more powerful in several ways. Collecting object attributes across a larger set of images might strengthen the predictive abilities of the prototype templates. On the other hand, further data collection might pollute the statistics already computed. In this case it might be necessary to add new object sub-classes such as "tree-in-winter" and "tree-in-spring". Adding new features and readjusting the importance of each feature used in matching might improve the prototype templates' characterizations of object classes and thus yield a better labeling performance.

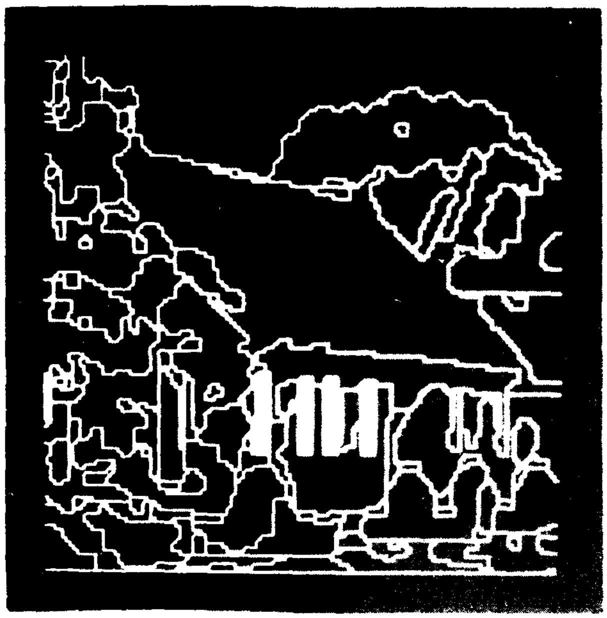
Spectral attribute matching, as it is currently implemented, can often provide an accurate initial set of hypotheses upon which to base the rest of the interpretation.

II.2 IMAGE-DEPENDENT ATTRIBUTE MATCHING VIA OBJECT EXEMPLARS

The process of matching spectral attributes described in the previous section involves a comparison of feature values of regions to image-independent feature values of object prototypes. Another approach that might prove more robust and context-sensitive is the use of a partial interpretation of the image. Assuming a region in an image has been identified as a particular object using some interpretation strategy, the feature values of that region can serve as an image-specific object template. These feature values can be used in finding similar regions that most likely represent instances of the same object class, using the same matching process described in Section II.1

Consider the example in Figure 3. Suppose shutter regions have been identified using shape characteristics Knowledge about the structure of houses suggests that regions of significant size that surround the shutter regions will represent house wall or windows Figure 3a shows the identified shutter regions and Figure 3b the neighboring regions hypothesized to represent house wall or window Here region neighbors are strictly adjacent; this requirement could be relaxed so that nearby regions that are not strictly adjacent would be included. A house wall template region was selected from among these regions by searching for the first region that was larger than a minimum size and had greater than a minimum value on a color transform feature. The "Q" value of the YIQ television color transform was used because house regions had consistent values on this feature across several images. Other features such as intensity and simple texture measures were not as useful in this respect

The house wall template region was used in a matching process in attempting to identify other house regions. Utilizing the level camera assumption, the knowledge that house wall regions will appear in a horizontal band of the image can be used to constrain the processing. Matching was restricted to those regions that overlapped a horizontal band defined by the upper and lower extents of the template region. This simple spatial constraint limits the matching and reduces the number of false alarms that would otherwise occur.



3a

3b

3c

Figure 3. (a) Identified shutter regions. (b) House wall surround regions. (c) House template in white with matching regions.

The strategy labels some regions incorrectly. In Figure 3c the light-colored region is the selected template and the slightly darker regions are those that matched. Note the errors in the sky region and tree highlight regions. The houses in the images are white; they tend to exhibit characteristics of the incident illumination. Highlights are smooth surface reflections and hence also exhibit characteristics of incident illumination. The strategy also fails to idenitfy all house wall regions, particularly those regions that represent shadowed house wall. In this case, internal contrast tends to be lower, affecting the texture measures used, and the spectral components are distributed over lower ranges, resulting in a poor match between these feature values and those of the selected exemplar. Both of these kinds of errors are reasonable given the overall goals of the approach: the formation of label hypotheses based on a loose notion of feature similarity.

As is the case with many of these simplified strategies, there are many plausible ways for achieving improvement in performance. The process might be made more powerful by incorporating stricter spatial constraints based on world knowledge. For example, matching might be restricted to those regions strictly adjacent to the template region or to those regions whose centroids lie within the horizontal band defined by the template region's upper and lower extents. Also, the features used in matching can be tailored to the object type being identified. Finding these characteristic features involves studying the consistencies of appearance of an object across many images. Only the features which tend to be invariant for an object class would be used in matching, thereby reducing the cost and hopefully producing better results.

While image-specific region templating avoids some of the problems faced in using an image-independent attribute matcher (e.g. lighting variations), the choice of a template remains crucial and is dependent upon the power and variety of the other interpretation strategies. For example, the house wall templating strategy described above depends directly on a strategy for locating shutters. Within the general structure of VISIONS, strategies are applied and interpreted in an environment of cooperation and competition among the various hypotheses developed [HAN78b] [PAR80] [WIL77]; labeling conflicts arising from the partial evidence available to each strategy are resolved in the context of more global information. Thus, although the region templating strategy is dependent upon correct identification of some of the image, it still serves as a powerful mechanism for extending a partial interpretation.

II.3 SKY/GROUND FILTERING

The techniques described so far have relied on color features alone in attempting to label the regions in an image. A strategy that incorporates the expected locations of two objects--sky and ground--can be used to eliminate or "filter out" erroneous hypotheses or to reduce conflicts between labels generated by separate processes.

In order to implement this strategy a sky template region and a grass template region must be selected. The sky template is chosen based on size, color, and location near the top of the image. The grass template is chosen based on color and location near the bottom of the image. The spatial extents of these regions are used to mark the probable lower limit of sky and the upper limit of ground. Figure 4 shows the sky line and ground line selected.

These two lines provide a rough approximation to the location of the horizon in the image. This information is used to filter the results of spectral attribute matching. For example, a region hypothesized to represent grass that appears above the sky line would have to be relabeled. This relabeling is accomplished by setting the confidence value for grass to the lowest possible value of -99.99. By doing this the next highest confidence value becomes the highest, and the region has a new object label.

The filtering process is helpful but, like region templating, is dependent upon careful selection of the sky and grass template regions. The selection of a low sky line or a high ground line does not provide much information but neither does it cause

Figure 4. Sky and ground lines.

accurately labeled regions to be relabeled incorrectly. On the other hand, a high sky line or a low ground line imposes strict constraints on the region labels and can cause the filtering process to eliminate correct labelings.

Better strategies for template selection might eliminate this problem. Having a model of the camera would provide the actual location of the horizon and furnish more accurate information about the actual extents of sky and ground. Finally, the groundplane can sometimes be approximately located by searching for the bottom edges of vertically oriented surfaces.

II.4 RECTANGLE FINDING

Rectangularity is a shape feature that characterizes many man-made objects. Doors, windows, and shutters that appear in a house image are usually rectangular or nearly so. Even rectangular objects that have been foreshortened by the camera angle can be identified by locating regions of high rectangularity in the image. These regions can be identified by applying a function that checks each region's deviation from rectangularity and saves those regions that survive a threshold. The deviation is a percentage calculated as follows:

$$\text{deviation} = 100.0 * \frac{(\text{area of enclosing rectangle} - \text{actual area})}{\text{area of enclosing rectangle}}$$

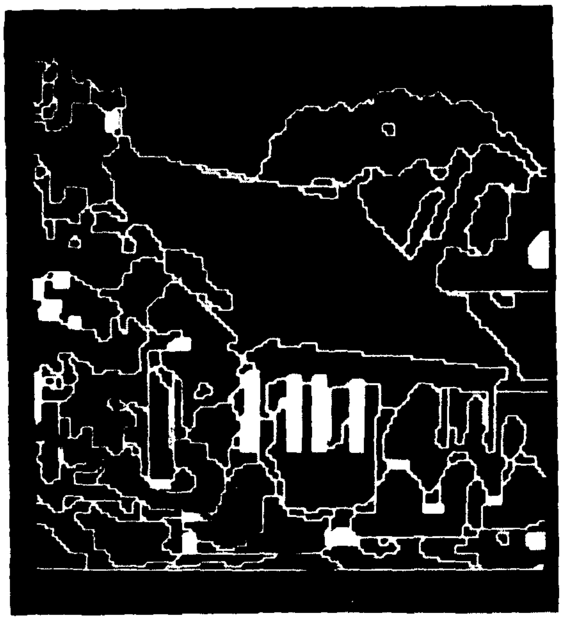


Figure 5. Regions with deviation from rectangularity of ≤ 25%.

Figure 6. Identified shutter regions.

Figure 7. Regions with height-to-width ratios > 5.

Figure 5 shows those image regions that survived a threshold of 25%.

Adding height-to-width ratio and size constraints to the rectangle finding strategy results in a shutter identification procedure. Figure 6 shows those image regions that were labeled as shutters. These regions have a height-to-width ratio greater than 5, a deviation from rectangularity of no more than 25%, and an area of at least 20 pixels (assumes a certain scale). Figure 7 shows image regions that were selected based on the height-to-width ratio constraint alone.

The parameters for the shutter identification procedure were set so as to give good results in the images under consideration. The size constraint helps to eliminate small regions that really have no significance in the interpretation. However, in images where a house is located far from the camera, the shutters will appear small and the procedure will fail to label them correctly. Also the strategy is likely to confuse doors, windows, and shutters since these objects have similar shapes and sizes.

Tailoring the height-to-width ratio to the object being searched for might eliminate some of the confusion. Shutters and doors often exhibit high contrast with respect to house walls; perhaps this information could also be employed. Finally, some shutters have not been recognized because the segmentation

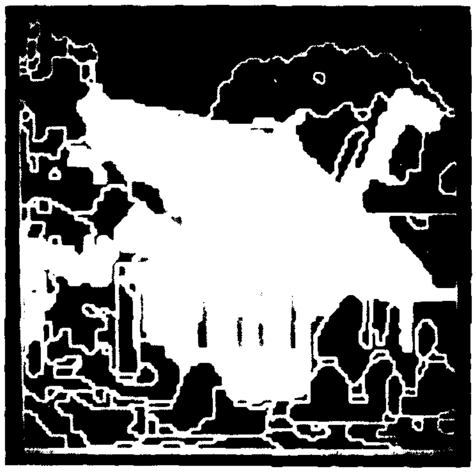


Figure 8. House wall regions from shutter surround.

processes have divided them into two or more regions that are less rectangular. Improving the results of the segmentation processes, possibly through a merging process, would likely yield better performances.

II.5 INFERENCING USING SPATIAL RELATIONS

Within the domain of house scenes, shutters, windows, and doors can serve as landmarks for locating a house. A house is a structure; its subparts are objects that exhibit certain typical spatial relations (e.g. windows fall between shutters). (See [WEY81].) The location of an identified object (the landmark) together with some spatial relation allows the inference of the location of another object [GAR78]. For example, shutters are usually surrounded (the spatial relation) by house wall. House wall can be identified by finding a shutter region (the landmark) and then labeling those regions that surround the shutter. This is the same idea that was used to identify "housepart" regions in the region templating example described earlier. Information about the structure of objects and the relations between objects and object subparts is currently built into various strategies. Work is in progress to develop a consistently structured database that will store and provide this type of information to the strategies that need it.

Figure 3 shows the results of finding shutters and then labeling neighboring regions as "house-part." Since the strategy is based solely on spatial relations, shadowed and unshadowed regions alike are labeled as "house-part", even though they differ greatly in their spectral attributes. This behavior can result in incorrect labelings when parts of the house are occluded. For example, a tree in front of the house might have parts located in proximity to the shutters and be labeled as "house-part." Also, the segmentation processes often produce small, thin horizontal or vertical regions that surround the shutter regions. These are the regions that are located by the strategy, while other larger, more significant regions are missed.

Expanding the neighbor idea to include "nearby" regions as well as those that are strictly adjacent might produce better results. Also, the merging and high contrast ideas mentioned in the previous section are applicable in this case, too. Finally, much work remains to be done in capturing the spatial relations that commonly occur between objects in natural scenes and structuring them for use by the interpretation strategies.

II.6 INFERENCING USING SIZE RELATIONS

The sizes of objects tend to vary a great deal, even within a single object class. This variability makes it difficult to characterize an object class based solely on size in absolute terms. Instead an object is often described or recognized in terms of its size relative to the size of some other object. For example, the actual height of a person is often less important than the relationship between the person's height and the heights of other people or objects in the environment. This observation suggests that object recognition can be based in part on relative size relations.

Given the ability to identify some object with reasonable accuracy, that object's size can be used to predict sizes for other objects that are located nearby in the scene. The relation of the region size to the object size can also provide some information about distance, elevation, and the perspective transformation.

Several tools were developed to investigate the use of size relations in image interpretation. An object size database was built; it contains the expected size ranges for the heights and widths of commonly occurring objects. A perspective module relates the camera model and image regions to real world surface characteristics such as orientation, range, elevation, height, and width. A strategy that uses both these tools was developed. The strategy consisted of labeling some region based on other features such as color and shape and then accessing the object size database to find the expected dimensions for the object label assigned to the region. These dimensions were passed to the perspective module which calculated the range and elevation of the object.

The strategy did not work well. One problem was the basic inability to label any region with great accuracy. Another was the variability of expected dimensions stored as ranges of values in the object size database; it was unclear whether to use the minimum value, the maximum value, the mean value, or something else. Also, the perspective module requires a camera model and these details were only available for one image. The perspective module has never been extensively tested, so the validity of the values it returned were usually in question. For these reasons, the strategy was not included in the interpretation process.

As further evidence of the difficulties involved in using object size information in interpretation, the sizes of house and shutter were compared in the four images. Two different measures of house size were used: the area of the rectangle that bounded those regions labeled as "house-part" and the summed areas of those same regions. The area of the shutter was simply the area of the shutter region. The ratios of house to shutter are presented in Table II. The variability exhibited precludes the reliable use of size relations in object recognition in this context. Furthermore, problems with segmentation errors and occlusion make the recovery of accurate size information very difficult.

The processes that develop the image segmentation and the strategies for object recognition must be improved before object size relations can be effectively exploited in image

Table II

Ratios of House Area to Shutter Area

	Image 1	Image 2	Image 3	Image 4
Object Extents	324:1	270:1	327:1	816:1
Region Areas	60:1	82:1	66:1	80:1

Expected Area Ratio = 139:1

(based on stored values for expected heights and widths) interpretation. Some method of determining the camera parameters would be helpful. Even when strategies for using size relations have been developed they most likely will be used only as a means of verifying hypotheses formulated by other strategies.

III COMBINING THE STRATEGIES: INTERPRETATION

The strategies outlined above rely on color, size, shape, and location features to identify objects in a scene. Combining these strategies with a simple blackboard like hypothesis space [ERM80] and a scheme for conflict resolution based on strategy reliability yields a fairly powerful image interpretation system. Processing is serial, control is hardwired, and all thresholds and parameters are set automatically.

The interpretation process proceeds as follows. The segmentation routines produce a set of labels that divides the image into regions. After initializing the hypothesis space and a few parameters, the system extracts features for every region, storing the calculated values in arrays that can be accessed by other procedures. (The values are also stored by region in the hypothesis space. Each process/strategy invocation adds new hypotheses to the space.) Next, spectral attribute matching is performed and the resulting hypotheses filtered after locating the approximate bounds of sky and ground. Object exemplars are chosen based on the preceding results and used to carry out region templating. Next a simple foliage finder locates regions likely to represent foliage by thresholding saturation values. The system then tries to locate shutters based on rectangularity, height-to-width ratio, and significant size. If shutters are found, the surrounding regions are hypothesized to represent house wall (or windows). One of the surrounding regions is chosen as an exemplar of house wall and other wall regions located using region templating. The roof is identified using expectations about rawblue and saturation values and size. Finally, regions are grouped by object type and conflicts resolved based on the reliability of the processes that generated the hypotheses involved.

The results of applying this system to the four images are shown in Figures 9-12. Labels have been compressed into foliage, house-part, grass, sky, road. In general, the system performs well. Sky, grass, and foliage regions are labeled accurately. Most of the house has been recognized. There are many small mistakes: house shadow is labeled as foliage, some tree highlight and sky regions are labeled as house, and so on. Some regions are not labeled at all.

What can be done to improve the results? Many suggestions for improving the individual strategies have been outlined in the previous sections. Other strategies need to be developed, especially in the areas of space and size relations. As these new

Figure 9

Figure 10

Figure 11

Figure 12

Figures 9 - 12. Interpretation results for Images 1-4. Labels in order of decreasing brightness: sky, house, foliage, grass, road, unknown. (Some labels may be difficult to distinguish due to flaws in reproduction.)

strategies are included control will become more important. The system must move from a fixed control flow to a flexible control architecture that can decide where to focus the system's attention and which strategies to apply [WIL82a] [WEY82] Finally, many more experiments must be designed and run in different domains and on different images. The results of these experiments will provide the best suggestions for designing new strategies and improving those already in use.

IV CONCLUSIONS

Experience with the simple system described above and its performance on several images provides some insights into the process of image interpretation. The most obvious of these is that any image interpretation system must incorporate a great deal of knowledge. This knowledge base must include information about the entities and relations that can and do occur in static images of outdoor scenes, structured so that it can be efficiently accessed and updated by the system. The complexity of this information and the structure inherent in the world of outdoor scenes suggest a representation composed of different levels of abstraction, ranging from simple edge elements "up" to more abstract schemas (structures that embody or aggregate knowledge about scenes and their constituent objects and relations). Future research will help to indicate the point at which such knowledge should move from the declarative (e.g. object descriptions) to the procedural (e.g.

processes that identify objects). Efforts to develop a robust, consistent representation are currently underway [WES82b].

The experiments described in this paper have also demonstrated the utility of four types of features--size, shape, color, and location--in object identification. Features of these types can be used in a knowledge base to describe objects and in procedures that implement generally-applicable strategies for recognizing objects in scenes. Further research will be aimed at developing finer-grained strategies and features to be used in identifying a larger set of objects.

Finally, the workings of the simple interpretation system have shown that features and relations become most important after having been incorporated within a variety of identification and verification strategies; descriptions alone do not constitute an interpretation system. Variety is the key word. Since all of the strategies are error-prone, redundancy is required to achieve any sort of sucess; strategies must compete, cooperate, and interact.

The strategies presented are simple and strengthened by several assumptions and yet each strategy seems fairly powerful and robust. Future work in different domains will test the validity of this claim.

Strategies are control mechanisms. They correspond roughly to the coordinated application of knowledge sources in a Hearsay architecture [ERM80], to meta-rules [DAV79] or control rules [AIK80] in a production system, to the processes attached to frames [MIN75]. While some commitments have been made to incorporating both bottom-up and top-down processing and parallel techniques for employing alternative models, much work remains to be done in choosing or developing an architecture of control that is powerful enough to guide the interpretation process and handle such problems as focus of attention, inferencing, and conflict resolution. [HAY77]

ACKNOWLEDGMENTS

Edward Riseman and Allen Hanson were instrumental in the writing of this paper, I thank them for their efforts. All of the VISIONS research group members contributed time and thought to the results reported here. Tom Williams originally programmed the feature extraction and attribute matching routines. Frank Glazer wrote the software for the perspective module. Ralf Kohler and Charlie Kohl made the strategies easier to implement by providing a powerful, supportive system and environment. Lenny Wesley and Terry Weymouth provided valuable criticism and ideas. My thanks to all.

References

[AIK80] J. S. Aikins, "Representation of Control Knowledge in Expert Systems," Proc. of 1st American Assoc. for Artificial Intelligence, 121-123, August 1980.

[BAJ74] R. Bajcsy and L. I. Lieberman, "Computer Description of Real Outdoor Scenes," Proc. of 2nd International Joint Conference on Pattern Recognition, Copenhagen, August 1974.

[BAJ78] R. Bajcsy and A. K. Joshi, "A Partially Ordered World Model and Natural Outdoor Scenes," in Computer Vision Systems, A. Hanson and E. Riseman, Eds., Academic Press, New York, 1978.

[BAL78] D. H. Ballard, C. M. Brown and J. A. Feldman, "An Approach to Knowledge-Directed Image Analysis," in Computer Vision Systems, A. Hanson and E. Riseman, Eds., Academic Press, New York, 1978.

[BAR72] H. G. Barrow, A. P. Ambler, and R. M. Burstall, "Some Techniques for Recognizing Structure in Pictures," in Frontiers of Pattern Recognition, S. Watanabe, Ed., Academic Press, New York, 1972.

[BAR75] H. G. Barrow and J. M. Tenenbaum, "Representation and Use of Knowledge in Vision," Stanford Research Institute Artificial Intelligence Center, TR 108, July 1975.

[BAR78] H. G. Barrow and J. M. Tenenbaum, "Recovering Intrinsic Scene Characteristics from Images," in Computer Vision Systems, A. Hanson and E. Riseman, Eds., Academic Press, New York, 1978.

[BIN80] T. O. Binford, R. A. Brooks, and D. G. Lowe, "Image Understanding Via Geometric Models," Proc. 5th International Conference on Pattern Recognition, 364-369, 1980.

[BOL76] R. C. Bolles, "Verification Vision within a Programmable Assembly System," Stanford AI Laboratory Memo AIM-295, December 1976.

[BRO78] R. A. Brooks, R. Greiner, and T. O. Binford, "A Model-Based Vision System," Proceedings Image Understanding Workshop, 36-44, May 1978.

[BRO79] R. A. Brooks, R. Greiner, and T. O. Binford, "The ACRONYM Model-Based Vision System," Proc. 6th Int. Joint Conf. on Artificial Intelligence, 105-113, August 1979.

[BRO81] R A Brooks, "Model Based Three Dimensional Interpretations of Two Dimensional Images," Proc 7th Int Joint Conf on Artificial Intelligence, 619-624, August 1981

[DAV79] R Davis, "Meta Rules: Reasoning about Control," MIT AI Memo, 1979

[DUD73] R O Duda and P E Hart, Pattern Classification and Scene Analysis, Wiley Interscience, 1973

[ERM80] L D Erman, F Hayes-Roth, V R Lesser, and D R Reddy, "The Hearsay-II Speech Understanding System: Integrating Knowledge to Resolve Uncertainty," ACM Computing Surveys, Vol 12, No 2, June 1980

[GAR74] T D Garvey and J M Tenenbaum, "On the Automatic Generation of Programs for Locating Objects in Office Scenes," Proc 2nd Int Conf on Pattern Recognition, 162-168, August 1974

[GAR76] T D Garvey, "Perceptual Strategies for Purposive Vision," SRI Artificial Intelligence Center Technical Note 117, Stanford Research Institute, Menlo Park, California, February 1976

[HAN78a] A R Hanson and E M Riseman, "Segmentation of Natural Scenes," in Computer Vision Systems, A Hanson and E Riseman, Eds , Academic Press, New York, 1978

[HAN78b] A R Hanson and E M Riseman, "VISIONS: A Computer System for Interpreting Scenes," in Computer Vision Systems, A Hanson and E Riseman, Eds , Academic Press, New York, 1978

[HAR80] R L Harr, "The Representation and Manipulation of Position Information Using Spatial Relations," Univ of Maryland Computer Vision Lab TR-923, August 1980

[HAY77] F Hayes-Roth and V R Lesser, "Focus of Attention in the Hearsay-II Speech Understanding System," Proc 5th Int Joint Conf on Artificial Intelligence, 27-35, 1977

[KAN78] T Kanade, "Task Independent Aspects of Image Understanding," Proceedings, Image Understanding Workshop, 45-50, May 1978

[KEN80] J R Kender, "Shape from Texture," Ph D thesis, Computer Science Dept , Carnegie Mellon University, 1980

[KES80a] W Kestner, "Considerations about Knowledge-Based Interpretation," Proc. 5th Int. Conf. on Pattern Recognition, 330-332, 1980.

[KES80b] W. Kestner, M Bohner, R Sharf, and M. Stiess, "Object Guided Segmentation of Aerial Images," Proc. 5th Int. Conf. on Pattern Recognition, 529-531, 1980.

[LEV81] M D. Levine and S I. Shaheen, "A Modular Computer Vision System for Picture Segmentation and Interpretation," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol PAMI-3, No 5, September 1981.

[MIN75] M Minsky, "A Framework for Representing Knowledge," in The Psychology of Computer Vision, P. Winston, Ed., McGraw-Hill, New York.

[NAG79] M. Nagao, T Matsuyama, and H Mori, "Structured Analysis of Complex Aerial Photographs," Proc. 6th Int. Joint Conf. on Artificial Intelligence, 610-616, 1979.

[NEW74] A Newell and D R Reddy, "Image Understanding: Some Notes," Minutes of the ARPA Image Understanding Workshop, Science Applications Inc., September 1974.

[OHT80] Y. Ohta, "A Region-Oriented Image Analysis System by Computer," Ph D. Thesis, Dept. of Information Science, Kyoto Univ., March 1980.

[PAR80] C.C. Parma, A R Hanson, and E M Riseman, "Experiments in Schema-Driven Interpretation of a Natural Scene," COINS Technical Report No 80-10, University of Massachusetts, Amherst, 1980.

[PRI78] K Price, "Change Detection and Analysis of Multispectral Images," Ph D Thesis, Dept of Computer Science, Carnegie-Mellon University, 1978.

[RUS79] D.M Russell, "Where do I look now? Modeling and Inferring Object Locations by Constraints," Proc. Pattern Recognition and Image Processing '79, 175-183, 1979.

[SEL81] P G Selfridge, and K R Sloan, Jr., "Reasoning about Images Using Meta-Knowledge in Aerial Image Understanding," Proc. 7th Int. Joint Conf. on Artificial Intelligence, 755-757, August 1981.

[SHA79] S I. Shaheen, "Image Segmentation and Interpretation Using a Knowledge Database," Ph D thesis, Dept of Electrical Engineering, McGill Univ., Montreal.

[SHI78] Y. Shirai, "Recognition of Real-World Objects Using Edge Cues," in Computer Vision Systems, A. Hanson and E. Riseman, Eds., Academic Press, New York, 1978.

[SHP79] L. Shapiro and R. Haralick, "Structural Descriptions and Inexact Matching," Virginia Polytechnic Institute Technical Report CS 79011 R, November 1979.

[TEN73] J. M. Tenenbaum, "On Locating Objects by Their Distinguishing Features in Multi-Sensory Images," Computer Graphics and Image Processing, Vol. 2, Nos. 3 and 4, December 1973.

[TEN77] J. M. Tenenbaum and H. G. Barrow, "Experiments in Interpretation-Guided Segmentation," Artificial Intelligence, Vol. 8, 241-274, 1977.

[WEY81] T. E. Weymouth, "Experiments in Knowledge-Driven Interpretation of Natural Scenes," Proc. 7th Int. Joint Conf. on Artificial Intelligence, 628-630, August 1981.

[WIL81] T. D. Williams, "Computer Interpretation of a Dynamic Image from a Moving Vehicle," COINS Technical Report No. 81-22, University of Massachusetts, Amherst, May 1981.

[YAK73] Y. Yakimovsky and J. A. Feldman, "A Semantics-Based Decision Theoretic Region Analysis," Proc. 3rd Int. Joint Conf. on Artificial Intelligence, 580-588, 1973.

[YOR81] B. W. York, "Shape Representation in Computer Vision," COINS Technical Report No. 81-13, University of Massachusetts, Amherst, May 1981.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER COINS TR 82-10	2. GOVT ACCESSION NO. AD-A115 914	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) STRATEGIES FOR KNOWLEDGE-BASED IMAGE INTERPRETATION		5. TYPE OF REPORT & PERIOD COVERED INTERIM
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Clif McCormick		8. CONTRACT OR GRANT NUMBER(s) ONR N00014-75-C-0459
9. PERFORMING ORGANIZATION NAME AND ADDRESS Computer and Information Science University of Massachusetts Amherst, Massachusetts 01003		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Arlington, Virginia 22217		12. REPORT DATE May, 1982
		13. NUMBER OF PAGES 37
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Distribution of this document is unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

computer vision
image interpretation
scene analysis
image understanding

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Strategies suggested by general observations can be used to control processing in a knowledge-based image interpretation system. Several strategies are discussed and experiments are presented to illustrate their use.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

DAT

FILM

7–