A REGION-ANALYSIS SUBSYSTEM FOR INTERACTIVE SCENE ANALYSIS

J. . Tenenbaum and S. Weyl
Artificial Intelligence Center
Stanford Research Institute
Menlo Parky California 94025

Abstract

An Interactive Scene Interpretation System (ISIS) is being developed by Stanford Research Institute's Artificial Intelligence Center as a tool for constructing and experimenting with man-•achine and automatic scene analysis methods tailored for particular image domains. A region analysis subsystem was developed recently based on the work of Brice and Fennema, and Yaklmovsky. Using this subsystem a series of experiments was conducted to determine good criteria for initially partitioning a scene into atomic regions and merging these regions into a final partition of the scene along object boundaries. Semantic (problem-dependent) knowledge is essential for complete, correct partitions of complex real-world scenes. An interactive approach to semantic scene segmentation was developed and demonstrated on both landscape and indoor scenes• This approach provides a reasonable methodology for segmenting scenes that cannot be processed completely automatically at present and Is a promising basis for a future fully automatic system.

I. Introduction

Most computer analyses of real-world scenes first attempt to partition the Image Into coherent regions corresponding to known objects (1, 2, 3, 4). Regions provide a convenient basis for semantic analysis by reducing both the amount of detail and the ambiguities of Interpretation found at the picture-element level.

An Interactive Scene Interpretation System (ISIS) is being developed by Stanford Research Institute's Artificial Intelligence Center as a tool for constructing and experimenting with man-machine and automatic scene-analysis methods tailored for particular image domains (5, 6). A subsystem was developed to explore systematically region-analysis techniques applied to several image domains. We summarize our most informative findings in this paper. A more complete presentation of our results can be found in (6).

II. Basic model

Following Brlce and Fennema (7), we divide region analysis Into two stages: first partition and region growing. The purpose of the first partition stage is to obtain a conservative Initial segmentation of the image where each region consists of picture elements belonging to only one object. In the region growing stage, adjacent regions with similar characteristics are merged into a single region to simplify further the organization. The desired result Is a set of regions corresponding to distinct objects in the scene.

Region analysis is computationally expensive, making It desirable to obtain a first partition that is as coarse as possible without extending the starting regions beyond object boundaries. The common practice is to sample the picture to reduce resolution and then Immediately combine adjacent samples with identical characteristics. When data are finely quantized or multidimensional (i.e., when color and range data are available) demanding strict equality on all attributes can lead to an unnecessarily large number of regions, many of which are caused by quantization noise. It may be helpful, therefore, to classify each sample into a small number of categories and then to treat picture elements assigned to the same category as identical (1, 3). We have conducted experiments with several sampling and quantization schemes for obtaining first partitions of landscape and office scenes, and a few methods have proved empirically to be adequate. We report our findings in Section III.

Following Yaklmovsky (3), our region-growing algorithm proceeds by serially selecting the pair of adjacent regions in the current organization that are globally "most alike," and merging these into a single region. The order in which regions are merged Is determined by a function that compares the similarity of a given pair of adjacent regions with the similarities of other pairs of regions that remain to be merged. A variety of criteria for region similarity have been used, including brightness contrast (7, 8) and average color contrast (3). We present experimental results with various similarity functions in Section IV.

The experiments described in this report are illustrated by the landscape and Indoor scenes in Figure 1. The basic region analysis process is shown in Figures 2 through 5. The first partition in Figure 2 is based on the brightness values associated with the landscape scene shown in Figure la, sampled to a 40 x 40 resolution. This partition Includes 806 Initial regions, half as many regions as individual picture elements. Region growing proceeds sequentially by merging adjacent regions across the weakest remaining boundary in the present partition. The boundary strength Is based on average absolute brightness and color difference. The results of merging arc illustrated, respectively, with 600, 450, and 250 regions remaining (see Figures 3 through 5).

III. First-Partition Experiments

The objective of the first partition stage is to group adjacent picture samples having similar attributes to obtain the fewest initial regions without a false merge occurring. In the example above, the picture was sampled to a 40 x 40 resolution; then adjacent samples with identical brightness were combined to form homogeneous atomic regions (Figure 2). We consider here some attempts to improve the quality of the first partition by using different sampling methods and different criteria for Judging the similarity of adjacent samples.

•The research reported herein was supported by the National Aeronautics and Space Administration under contract NASW-2086.

Sampling Experiments

We experimented with modal, mean, and straight sampling to determine which method is the most advisable for first partition. All experiments were performed on gray scale images using a 40 X 40 rectangular sampling grid. Note that in scenes with periodic texture a random sampling strategy would have been necessary to avoid aliasing effects.

In modal sampling (Figure 6), the gray level of each grid point is taken to be the most frequently occurring value of gray level for nearby points. The number of initial regions obtained in this manner is significantly reduced (by about one-third) because many small "noise" points (which occurred in the treetop and the ground) disappeared. Some fine detail and contrast boundaries were lost, however, in the modal smoothing. Mean sampling of the gray level in a small neighborhood around each grid point is a poor technique because it tends to smooth discontinuities (see Figure 7).

From experiments with the Images presented here and several others, we concluded that first partitions based on a straight sampling of the gray-scale Image taken through a neutral density filter contained most of the information necessary to arrive at a conservative partition of real landscape and indoor scenes.

Color Quantization

The sampling experiments of the last section were based entirely on brightness information. The other extreme is to base the first partition exclusively on color information, with brightness normalized out. One way to accomplish this is to transform the original image into a two-dimensional color space based on the relative content of red, green, and blue at each point [using the model described in (5)]. This color space can be quantized into uniform intervals, with image points falling into the same color quantum grouped together into regions. We found that the partitions obtained in this manner were consistently worse than partitions based on sampled brightness. Major leaks occurred between semantlcally distinct regions in both landscape and indoor scenes. Moreover, because of textural irregularities, the total number of regions in landscape scenes was greater than the number produced by partitions based on sampled brightness.

We tried to improve first partitions based on color quantization by selecting quantization intervals corresponding to characteristic colors of prominent objects. Our attempts were unsuccessful in both landscape scenes, because of the overlapping hue distributions of the principal objects, and in indoor scenes, because interior surfaces are generally color coordinated and hues tend to cluster in a narrow range.

Extraction of Distinguished Objects

Objects can often be extracted from the image prior to general partitioning by selecting contiguous collections of image points with distinguishing properties. For example, the sky was the only region in our landscape scenes that is brighter than 30 (on a scale of 31). Extraction is especially effective when multisensory data are available, increasing the probability that particular objects will be distinguished along some dimension. By extracting distinguished objects sequentially (or, in general, hierarchically) we can take advantage of context reductions achieved by earlier predicates. For example, although the picture on the wall In Figure la is a complex pattern, we were able to extract it by simple conditions on height, hue, and saturation once the wall samples were accounted for

Table 1 presents a set of sequentially applied criteria developed Interactively using ISIS for Stanford Research Institute office scenes. Points not classified by the criteria given in Table 1 were partitioned on the basis of their sampled brightness. Figure 8 provides a comparison of the partition obtained using these criteria with the partition obtained by sampled brightness alone. The simplification of the partition is clear from the comparison.

In landscape scenes objects are usually distinguished by texture and shape rather than by features at the image point level. Consequently, object extraction techniques are more difficult to apply and we have not, as yet, been able to use them successfully.

IV Merge Priority Experiments

The sequence of merges performed during region growing is ordered by nonsemantic measures of region similarity. The purpose of these experiments was to determine conservative similarity measures, which will defer questionable merges in the hope that the decision will be clarified or even rendered unnecessary in the context resulting from the execution of more reliable merges-

We compared the quality of several different measures of region similarity by performing a first partition based on sampled brightness, and then by applying a global, best-first merge order based on each measure of similarity until only 250 regions remained in the scene. The results of the merge sequences were compared on the basis of how well they honored the correct organization on the scene.

Figures 5 and 9 through 13 present the results of merging, down to 250 regions with six different measures of region similarity. Each figure legend gives the formula used for computing the similarity of adjacent regions in each case. In these formulas, br . is the brightness seen through a neutral density filter, of the i^{tn} image point on the boundary of region a, and brbi is the brightness of adjacent image points in region b; r_{ai}, g_{ai}, and b_{ai} are the brightnesses of the i^{th} boundary element from region a seen through the red, green, and blue filters respectively, and r_{bi}, gb, and b_{bi} are the corresponding brightnesses from region b; $T_{\&t}$ £$_{ft}$, and B_{fl} are the average brightnesses over region a seen through the red, green, and blue filters, and *$_{b}$, K_{D}» and B_{b} are the corresponding averages over region b.

In Figure 11, the similarity of adjacent regions was determined by averaging over sample points along the common boundary the maximum color contrast between any two picture elements drawn from the original, full-resolution, sampling neighborhoods on opposite sides of the boundary. This method was expected to overcome errors resulting from inadequate sampling; however, textured regions, such as the ground and the treetop, failed to coalesce before

before distinct smooth regions grew together. Our best results were obtained using the measure of similarity applied in Figure 13. In this case the similarity of two regions was computed conservatively using the maximum of both the boundary color contrast and the region color contrast as defined in the legends of Figures 5 and 12.

V. Semantic Region Growth

We saw in the merge priority function experiments of the previous section that, regardless of the nonsemantlc similarity criterion used, an erroneous merge is proposed well before a final partition is obtained. Semantics must be used either to refine the boundary strength criterion so that it proposes fewer erroneous merges *(3)* or to block proposed merges that are incorrect. Stepping through merges proposed by our best nonsemantic similarity criterion, we observed that serious false merges seldom occurred until the regions involved had grown sufficiently large to permit semantic interpretations based on region properties. This suggested that merging errors could be avoided on semantic grounds simply by refusing to merge regions with different interpretations. We tested this idea interactively by modifying the region growing algorithm to check semantic compatability before performing a proposed merge. Merging is allowed only if both regions carry the same interpretation or if at least one of the regions is not yet interpreted. Newly merged regions inherit the interpretation of their parents (or parent, if only one region is interpreted). When two uninterpreted regions are merged, if the size of the resultant regions exceeds a threshold, the program requests the experimenter to supply manually a correct interpretation.

This interactive region-growing algorithm partitioned both a landscape and an indoor scene with only minor errors (caused primarily by inadequate spatial sampling). In both experiments, the size threshold for manual interpretation was set empirically at seven samples. The final partition depicted in Figure 14 was based on the first partition in Figure 8b. Initially, manual Interpretations were provided for the 20 (out of 253) first partition regions that exceeded threshold size. About 20 additional interpretations were provided during the subsequent analysis when uninterpreted regions attained threshold size by merging. Approximately the same number of interpretations had to be supplied initially and during region growing in the landscape scene. When our semantic region-growing algorithm was applied to the first partition of Figure 2, we obtained the results shown in Figure 15. In both cases there were seven distinct interpretations.

VI. Conclusion and Future Plans

Several experiments in automatic and interactive scene analysis using ISIS were performed. Our positive results with the interactive, semantic region grower suggest two directions for future work

First, this method can be used as the basis for a practical approach to cooperative (man-machine) segmentation of scenes that are too complex to process completely automatically, or too detailed to segment rapidly by hand. With relatively little effort a user could crudely outline and label major regions. These outlines would provide most of the required region interpretations and also serve as a good initial partition from which detailed boundaries can be grown rapidly.

Second, the semantic region growing algorithm provides a promising basis for a future automatic system in which region interpretations are deduced from local attributes and contextual constraints imposed by previously interpreted regions. The automatic system will have to deal with regions that are ambiguous at a given stage of partitioning. Merges involving such regions will be deferred until the ambiguity has been resolved as a result of other, more reliable merges. The use of semantics for blocking merges, rather than for altering the order in which merges are proposed [cf. (3)J, should simplify training, since region labeling criteria and contextual constraints can be introduced or refined in direct response to specific merging errors as these errors are observed.

References

1. R. Bajscy and L. Lieberman, "Computer Description of Real Outdoor Scenes," Proceedings 2nd International Joint Conference on Pattern Recognition, p. 174, Copenhagen, Denmark (August 1974).

2. F. P. Preparata and S. R. Ray, "An Approach to Artificial Non-Symbolic Cognition," Information Science, Vol. 4, pp. 65-86 (1972).

3. Y. Yakimovsky and J. A. Feldman, "A Semantics Based Decision Theoretic Region Analysis," Proceedings 3rd International Joint Conference on Artificial Intelligence, p. 580 (August 1973)

4. N. J. Nilsson et al., "Artificial Intelligence –Research and Applications," Progress Report to ARPA Covering the Period 9 October 1972 to 8 March 1974, Stanford Research Institute, Menlo Park, California (April 1974).

5. J. M. Tenenbaum et al., "An Interactive Facility for Scene Analysis Research," Technical Note 87, SRI Project 1187, Artificial Intelligence Center, Stanford Research Institute, Menlo Park, California (January 1974).

6. J. M. Tenenbaum et al., "Research on Interactive Scene Analysis," Final Report, SRI Project 8721, Artificial Intelligence Center, Stanford Research Institute, Menlo Park, California (March 1975).

7. C. R. Brice and C L. Fennema, "Scene Analysis Using Regions," Artificial Intelligence, Vol. 1, No. 3, pp. 205-226 (Fall, 1970).

8. H. Barrow and R. Popplestone, "Relational Descriptions in Picture Processing," In Machine Intelligence, Vol. 6, pp. 377-396 (Edinburgh University Press, Edinburgh, Scotland, 1971). (Editors B. Meltzer and D. Michie).

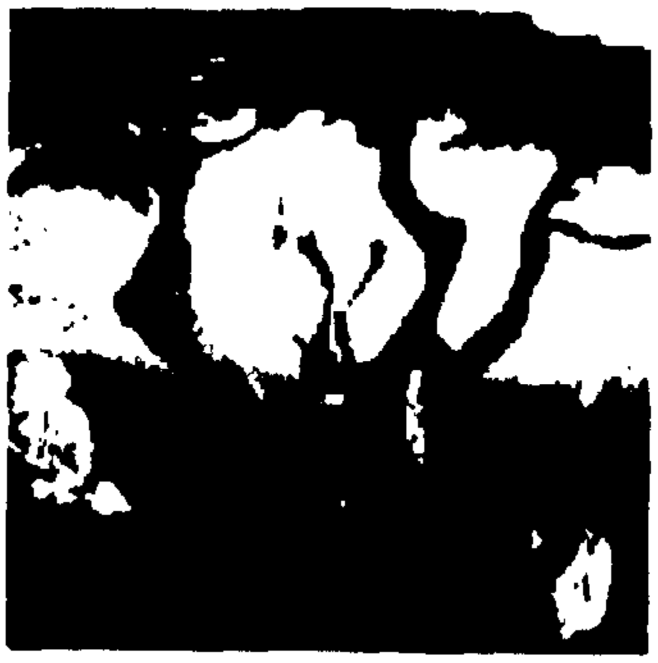

Figure 1a. Landscape Scene (Monterey, Calif.)

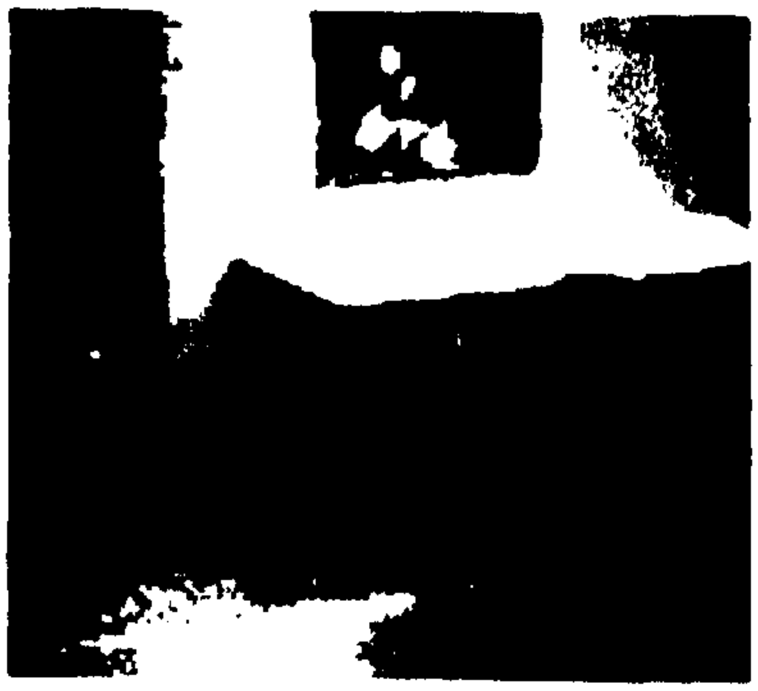

Figure 1b. SRI Office Scene

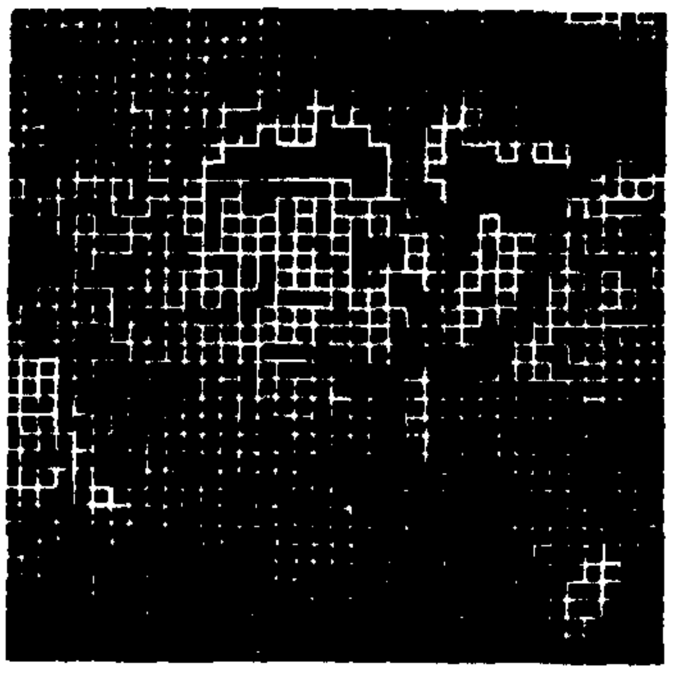

Figure 2. First Partition of Landscape Scene (806 Regions)

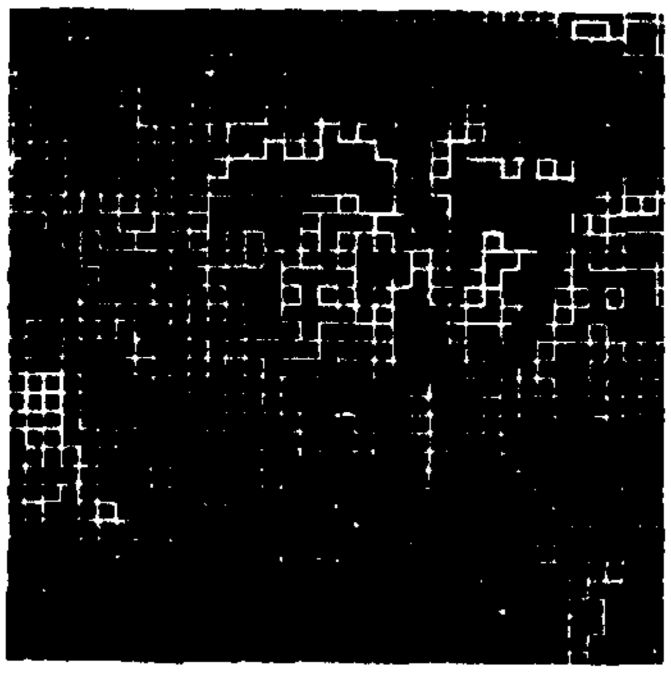

Figure 3. Partitioned Landscape Scene After 206 Merges (600 Regions Remaining)

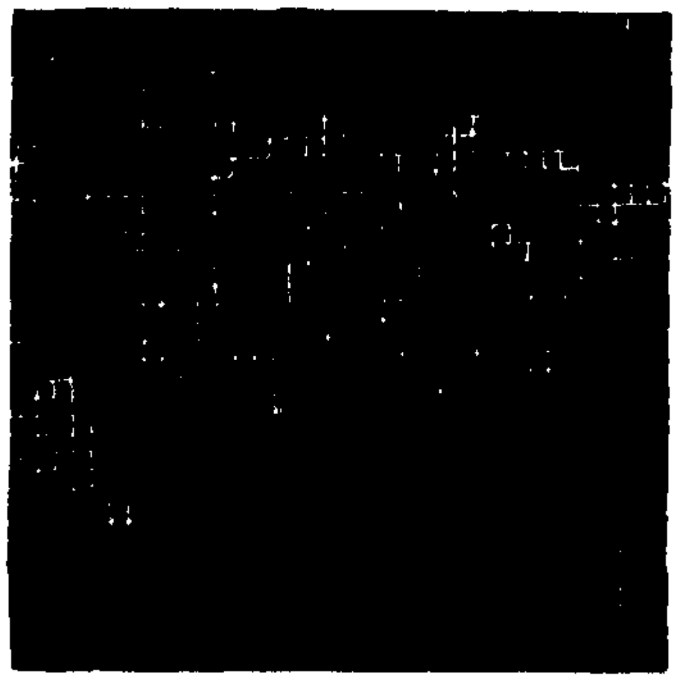

Figure 4. Partitioned Landscape Scene After 150 Additional Merges (450 Regions)

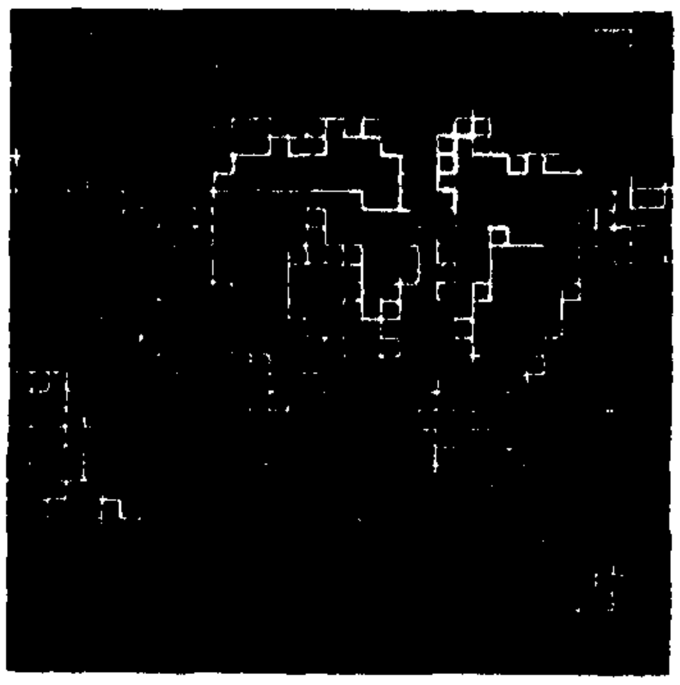

Figure 5. Partitioned Landscape Scene After 200 Additional Merges (250 Regions). "Boundary Color Contrast" Similarity Criterion =

$$(\sum_{i=1}^{N} | r_{ai} - r_{bi}| + |g_{ai} - g_{bi}| + |b_{ai} - b_{bi}|)/N$$

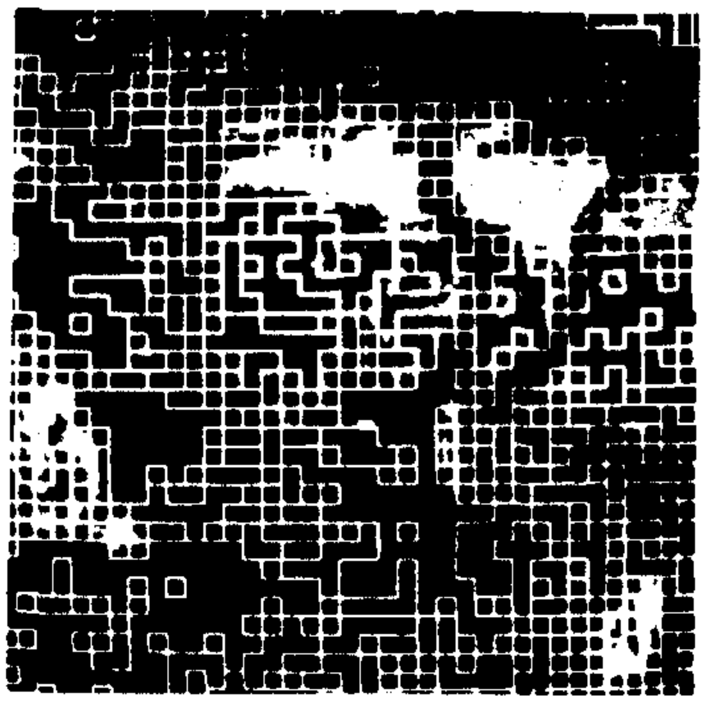

Figure 6. First Partition of Landscape Scene Produced from Modal Samples (See Figure 5)

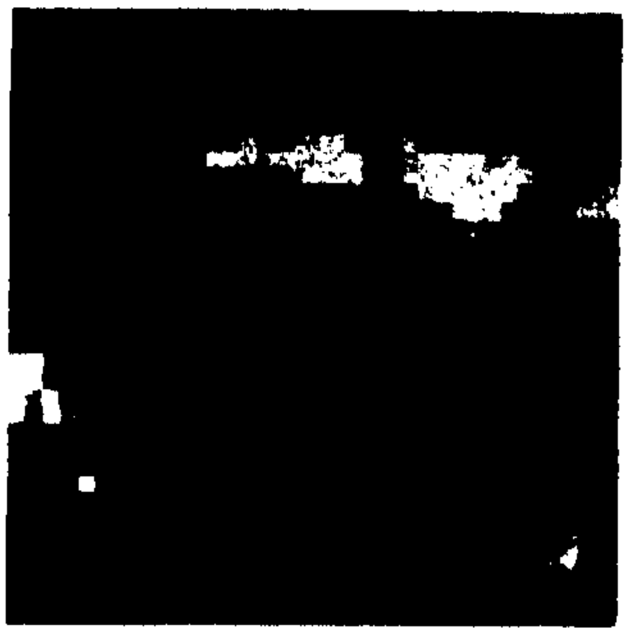

Figure 7. Sampled Landscape Scene with Each Sample Displayed at the Average Brightness Over a 3 x 3 Neighborhood

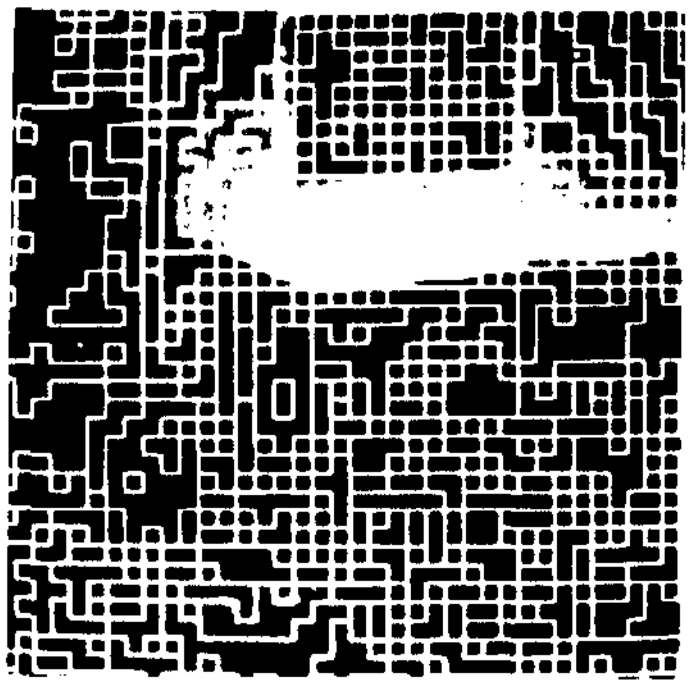

Figure 8a. First Partition of SRI Office Scene (Figure 1b) Based on Sampled Brightness (583 Regions)

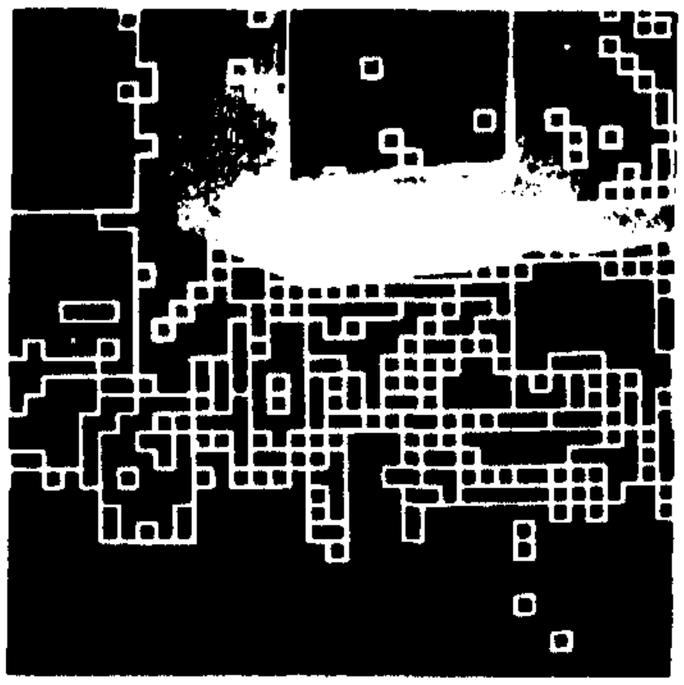

Figure 8b. First Partition based on sequential classification of image points in SRI Office Scene (235 Regions)

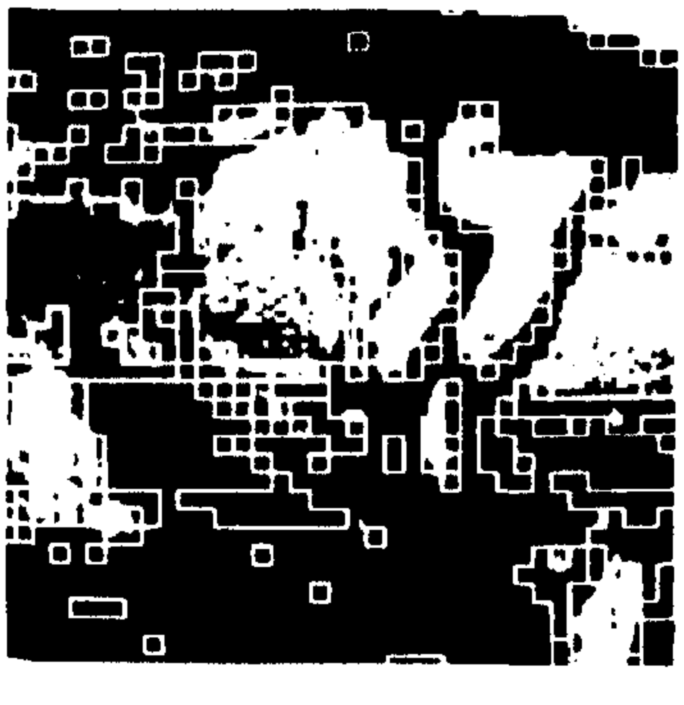

Figure 9. Landscape Scene Partitioned by "Brightness Contrast." Similarity Criterion=

$$\frac{\sum_{i=1}^{N} |br_{ai} - br_{bi}|}{N}$$

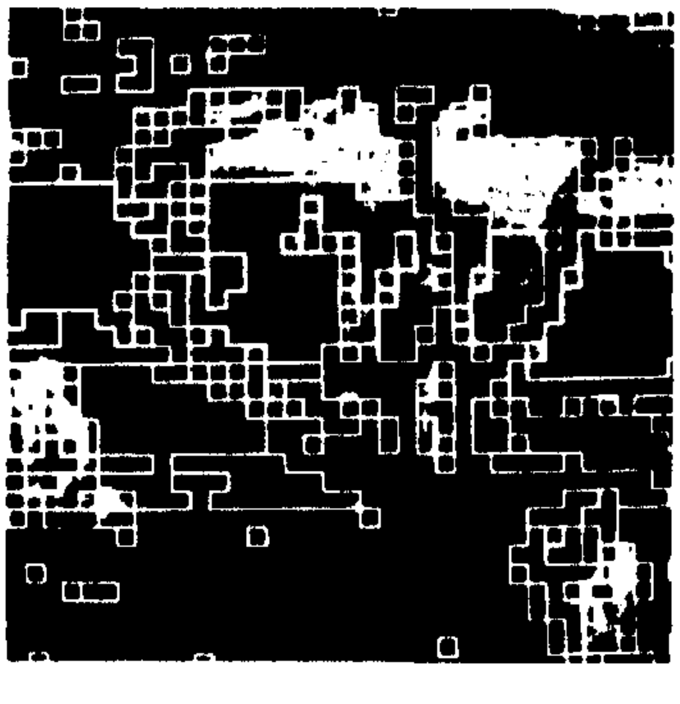

Figure 10. Landscape Scene Partitioned by "Boundary Color Contrast." Similarity Criterion-

$$\frac{\sum_{i=1}^{N} (r_{ai} - r_{bi})^2 + (g_{ai} - g_{bi})^2 + (b_{ai} - b_{bi})^2}{N}$$

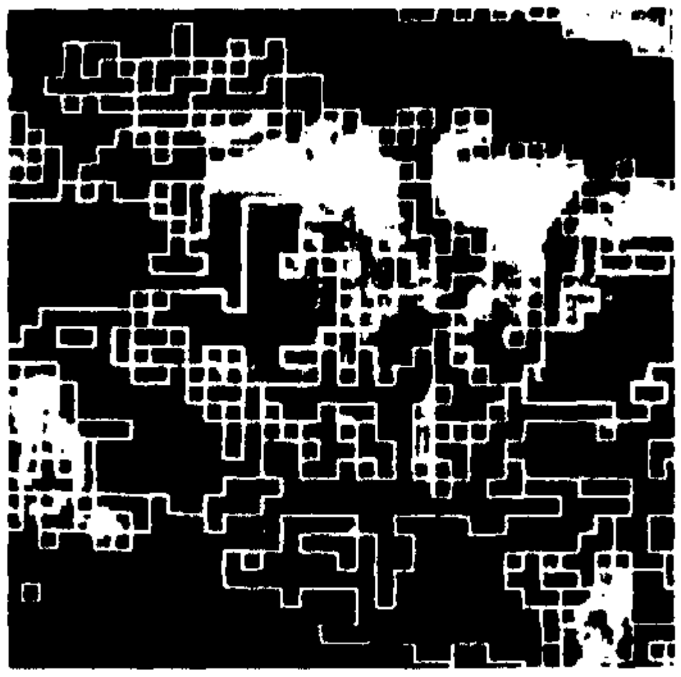

Figure 11. Landscape Scene Partitioned by Maximum Boundary Color Contrast Computed at Full Spatial Resolution (see text)

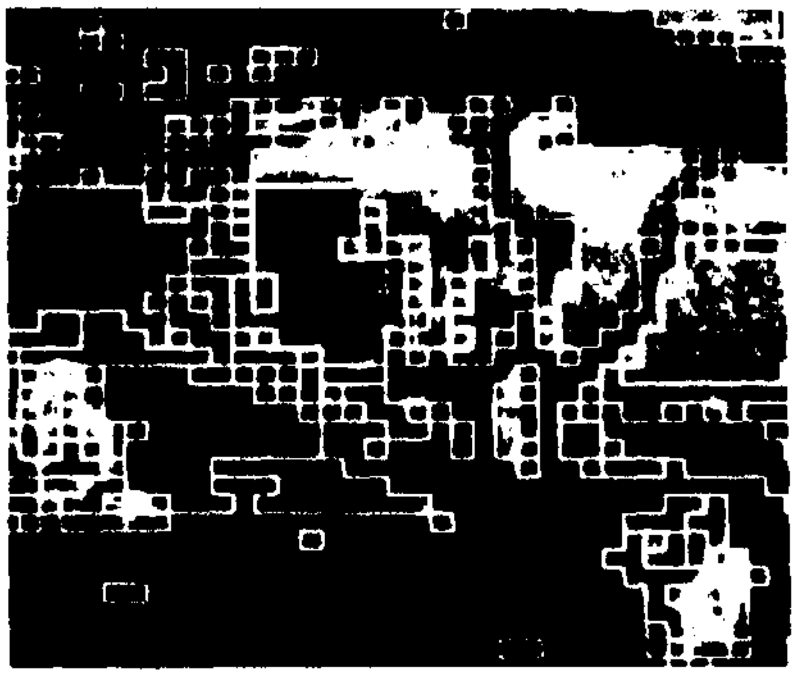

Figure 12. Landscape Scene Partitioned by "Region Color Contrast." Similarity Criterion = $|\bar{r}_{ai} - \bar{r}_{bi}| + |\bar{g}_{ai} - \bar{g}_{bi}| + |\bar{b}_{ai} - \bar{b}_{bi}|$

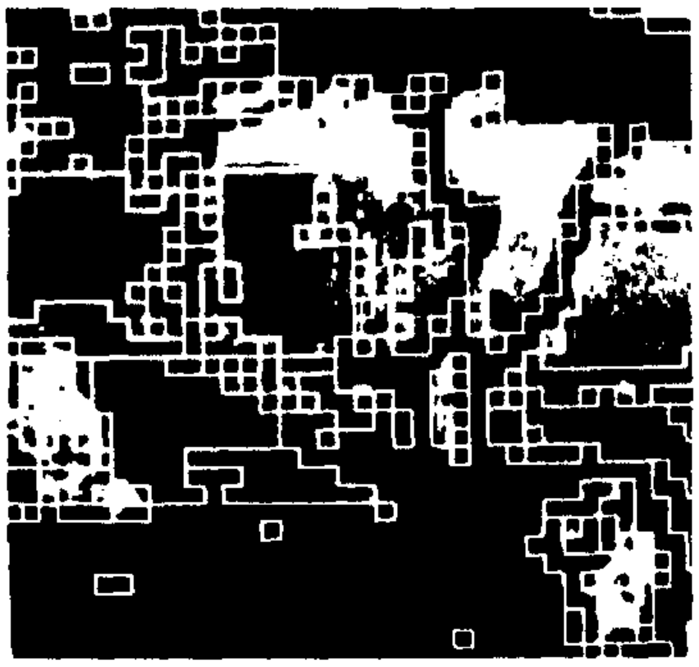

Figure 13. Landscape Scene Partitioned by Maximum of Boundary and Region Color Contrast Functions.

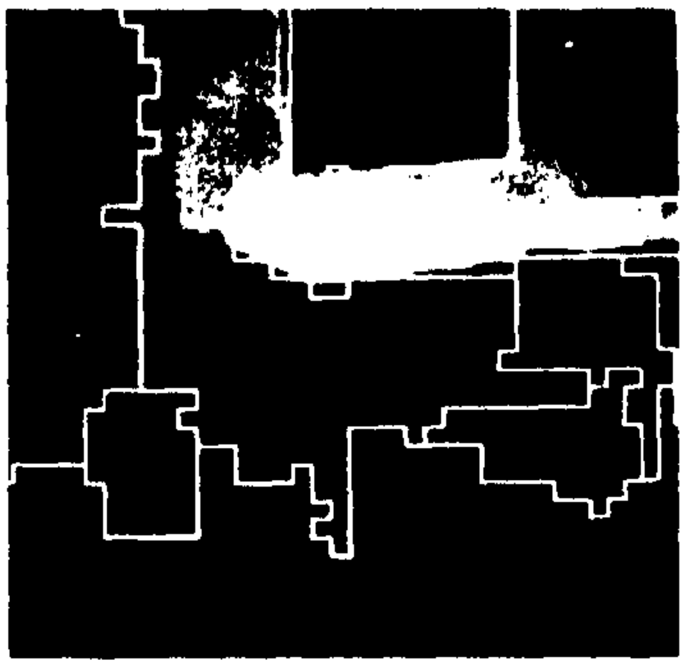

Figure 14. Final Semantic Partitioning of SRI Office Scene

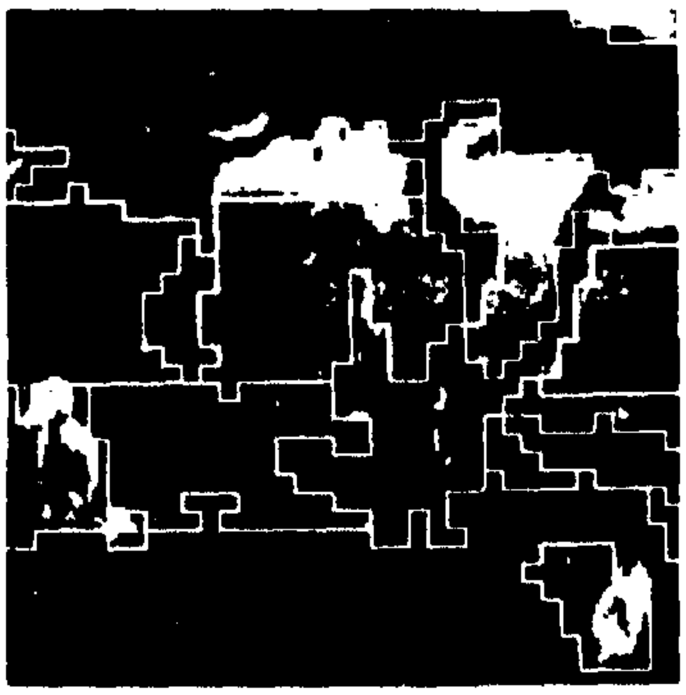

Figure 15. Final Semantic Partitioning of Landscape Scene

TABLE 1

Sequential Classification Criteria for Figure 8b

(1) Extract floor samples by height (0.1 foot)

(2) Extract chairseat samples by characteristic height and horizontal orientation

(3) Extract tabletop samples by characteristic height and horizontal orientation

(4) Extract picture samples in two passes:

(a) By characteristic height and saturation greater than maximum saturation for wall

(b) By characteristic height and hue outside the hue range of wall

(5) Extract chairback samples by characteristic height, vertical orientation, and saturation

(6) Partition remaining samples by brightness