Best Available
Copy
for all Pictures

AD-A012 857

ANALYSIS OF NATURAL SCENES

Ronald B. Ohlander

Carnegie-Mellon University

Prepared for:

Air Force Office of Scientific Research
Defense Advanced Research Projects Agency

April 1975

DISTRIBUTED BY:

NTIS
National Technical Information Service
U. S. DEPARTMENT OF COMMERCE

AFOSR - TR - 75 - 1090   218070

ADA012857

ANALYSIS OF NATURAL SCENES

Ronald B. Ohlander

D D C
[illegible]
AUG 4 1975
[illegible]
B

DEPARTMENT
of
COMPUTER SCIENCE

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH (AFSC)
NOTICE OF TRANSMITTAL TO DDC
This technical report has been reviewed and is
approved for public release IAW AFR 190-12 (7b).
Distribution is unlimited.
D. W. TAYLOR
Technical Information Officer

Carnegie-Mellon University

Reproduced by
NATIONAL TECHNICAL
INFORMATION SERVICE
US Department of Commerce
Springfield, VA. 22151

Approved for public release;
distribution unlimited.

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFOSR - TR - 75 - 1090	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER AD-A012 857
4. TITLE (and Subtitle) ANALYSIS OF NATURAL SCENES		5. TYPE OF REPORT & PERIOD COVERED Interim
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Ronald B. Ohlander		8. CONTRACT OR GRANT NUMBER(s) F44620-73-C-0074
9. PERFORMING ORGANIZATION NAME AND ADDRESS Carnegie-Mellon University Computer Science Dept. Pittsburgh, PA 15213		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61101D AO-2466
11. CONTROLLING OFFICE NAME AND ADDRESS Defense Advanced Research Projects Agency 1400 Wilson Blvd Arlington, VA 22209		12. REPORT DATE April 1975
		13. NUMBER OF PAGES 210
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) Air Force Office of Scientific Research (NM) 1400 Wilson Blvd Arlington, VA 22209		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

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

18. SUPPLEMENTARY NOTES

Reproduced from best available copy.

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

PRICES SUBJECT TO CHANGE

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

This report describes work performed on two aspects of the scene analysis process. These are segmentation, and the treatment of occlusions, shadows, and highlights. The eventual goal of the research is the formulation of knowledge sources which play an important role in a model for a general vision system. The model is based on the hypothesize-and-test paradigm and consists of a number of independent knowledge sources which cooperate through a global data base. The sources of knowledge modify the data base to effect eventual scene understanding (see chapter 2).

continued

20. abstract -continued

We propose a general segmentation algorithm which makes effective use of existing techniques to parse natural scenes. The principal operator employed is thresholding. Cutoff values for the thresholding operation are determined from histograms of multiple sensory parameters. Various discontinuities are often present in the histograms, and indicate an area possessing uniformity in some feature (e.g., hue or intensity). The thresholding operation is utilized in a recursive descent control structure to isolate and refine segments of the picture. At each level of recursion the histograms are derived for the largest unprocessed region remaining in the image.

Before recursive analysis begins, an estimate of the heavily textured (busy) areas is obtained. This estimate provides direction for the analytic process. If thresholding occurs over an area which is heavily textured, the output is handled in one of two ways. If there is evidence that there also exist non-busy regions with the same properities, the resultant point clusters are refined to eliminate the busy contribution. If evidence indicates the extracted points belong exclusively to a heavily textured area, then they are discarded.

The main goal of the recursive algorithm is to continually isolate segments of the image which can be refined. This will sometimes happen by direct application of the thresholding operation. It can also occur because regions are isolated when an extracted segment is removed from further consideration. In cases where the procedure halts with a substantial portion of the image unprocessed, methods have to be employed to force isolation of portions of the picture. This allows refinment of a segment of the picture without overwhelming interference. This forced isolation is accomplished, within our system, by extracting lightly and heavily textured regions of the scene. These are then refined by using the basic algorithm on the isolated subpicture. The completion of this phase of the processing will often result in additional isolated regions. They will be treated in the usual way. The forced isolation of the image permits analysis to be carried to a much further degree than would have been possible with the basic algorithm.

The other area of research, treatment of occlusions, shadows, and highlights, is attacked by performing a case analysis to determine types of these conditions. The resulting classifications permit us to identify invariants for the different phenomena. Proximity, similarity, and continuity are the invariants. One region cannot occlude or shadow a second region unless they are immediate neighbors. If an occlusion relationship exists two regions must also be dissimilar. If they bear a shadow or highlight relationship they must be similar. Continuity is an invariant that exists within types. It is a property that indicates the actual extent of occluded or shadowed areas through local clues in the picture. These local clues can be exploited to provide guidelines for "restoration" of a region by boundary reformation and extension of other regional attributes. Through this process we can "normalize" a given region, which is occluded or shadowed in different ways, to the same primitive model.

Results showing the strong and weak points of both aspects of the research are presented. They should be considered preliminary in the sense that continual refinement of the algorithms is expected.

ANALYSIS OF NATURAL SCENES

Ronald B. Ohlander

Department of Computer Science
Carnegie-Mellon University
Pittsburgh, Pennsylvania 15213
April, 1975

Submitted to Carnegie-Mellon University in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

This work was supported by the Advanced Research Projects Agency of the office of the Secretary of Defense (F44620-73-C-0074) and is monitored by the Air Force Office of Scientific Research. This document has been approved for public release and sale; its distribution is unlimited.

ABSTRACT

This report describes work performed on two aspects of the scene analysis process. These are segmentation, and the treatment of occlusions, shadows, and highlights. The eventual goal of the research is the formulation of knowledge sources which play an important role in a model for a general vision system. The model is based on the hypothesize-and-test paradigm and consists of a number of independent knowledge sources which cooperate through a global data base. The sources of knowledge modify the data base to effect eventual scene understanding (see chapter 2).

We propose a general segmentation algorithm which makes effective use of existing techniques to parse natural scenes. The principal operator employed is thresholding. Cutoff values for the thresholding operation are determined from histograms of multiple sensory parameters. Various discontinuities are often present in the histograms, and indicate an area possessing uniformity in some feature (e.g., hue or intensity). The thresholding operation is utilized in a recursive descent control structure to isolate and refine segments of the picture. At each level of recursion the histograms are derived for the largest unprocessed region remaining in the image.

Before recursive analysis begins, an estimate of the heavily textured (busy) areas is obtained. This estimate provides direction for the analytic process. If thresholding occurs over an area which is heavily textured, the output is handled in one of two ways. If there is evidence that there also exist non-busy regions with the same properities, the resultant point clusters are refined to eliminate the busy contribution. If evidence indicates the extracted points belong exclusively to a heavily textured area, then they are discarded.

The main goal of the recursive algorithm is to continually isolate segments of the image which can be refined. This will sometimes happen by direct application of the thresholding operation. It can also occur because regions are isolated when an extracted segment is removed from further consideration. In cases where the procedure halts with a substantial portion of the image unprocessed, methods have to be employed to force isolation of portions of the picture. This allows refinment of a segment of the picture without overwhelming interference. This forced isolation is accomplished, within our system, by extracting lightly and heavily textured regions of the scene. These are then refined by using the basic algorithm on the isolated subpicture. The completion of this phase of the processing will often result in additional isolated regions. They will be treated in the usual way. The forced isolation of the image permits analysis to be carried to a much further degree than would have been possible with the basic algorithm.

The other area of research, treatment of occlusions, shadows, and highlights, is attacked by performing a case analysis to determine types of these conditions. The resulting classifications permit us to identify invariants for the different phenomena. Proximity, similarity, and continuity are the invariants. One region cannot occlude or shadow a second region unless they are immediate neighbors. If an occlusion relationship exists two regions must also be dissimilar. If they bear a shadow or highlight relationship they must be similar. Continuity is an invariant that exists within

types. It is a property that indicates the actual extent of occluded or shadowed areas through local clues in the picture. These local clues can be exploited to provide guidelines for "restoration" of a region by boundary reformation and extension of other regional attributes. Through this process we can "normalize" a given region, which is occluded or shadowed in different ways, to the same primitive model.

Results showing the strong and weak points of both aspects of the research are presented. They should be considered preliminary in the sense that continual refinement of the algorithms is expected.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

I am deeply grateful to Professor Raj Reddy who has been a constant source of ideas and inspiration during my association with him. No one could ask for a better advisor or better friend. It is the simple truth when I say this dissertation could not have been written without him.

Special thanks go to Keith Price, my co-worker in the vision field, who has always been ready to provide sound advice and evaluate my own ideas. He has made so many contributions to the overall vision effort that I could not even begin to enumerate them

There are many additional people who have provided hours of assistance to this body of reasearch and who have helped me in my hour of need: Alan Frisch, Greg Lawson, Chris Moriondo, Paul Masulis, Eric Ostram, and Bob Watkins. I will especially remember John Kender, whose only reward for long and late hours assisting me in segmentation research, is my heartfelt gratitude.

My sincerest appreciation goes to my thesis committee: Lee Erman, Sam Fuller, and Frank Quick, who have had the kindness to provide a sound evaluation of this body of work in an unreasonably short time.

I dedicate this thesis to my wife, Gail, who has provided love, warmth, and devotion through the last four years of stress and long hours. This past year has meant practical widowhood for her.

Finally, I am indebted to the U.S. Navy for making my education possible.

1 INTRODUCTION

The work of the vision group in the Computer Science Department at Carnegie-Mellon University (CMU) has been primarly concerned with the construction of a computer vision system that will eventually approach human performance. We want to structure an information processing model which will be capable of analyzing visual stimuli from a variety of natural scenes in order to arrive at identification of objects within a given context. Of course such an ambitious project will involve many man-years of labor and contributions from many individuals. What we wish to discuss in this dissertation are our own particular contributions to that effort.

In such a massive undertaking as a general vision system, considerable thought must be given to the framework within which research can be carried out. We discuss a model around which various components of the system can be constructed. It is based upon the conception of cooperating independent knowledge sources which operate on a global data base. The model has already been successfully utilized in the HEARSAY I (Reddy et al., 1973a, 1973b) and HEARSAY II (Lesser et al., 1974; Erman and Lesser, 1975) speech understanding projects at CMU and has been shown to provide a good foundation for complex perceptual tasks. One of the ways in which it particularly suits our needs is by allowing substantial independent research on different knowledge modules without undue attention paid to communication issues. All one need understand is how to operate on the constructs of the underlying data structure of the global data base.

In addition to a framework or model within which to work, large tasks also require methodologies which provide a means of attacking very difficult problems. A great many facts must be accumulated and implemented in the form of procedures, production systems, or other mechanisms which operate on the data base to produce desired results. Some facts are always obtainable from past investigations. More often though, when one is working at the forefront of research, relevant information must be culled from huge amounts of experimental data and acted upon in different ways to see what results are obtainable. One of the ways in which we have tried to provide a more methodical approach to these issues is by providing interactive subsystems which allow a wide range of experimentation with a minimum of effort. Such subsystems, of course, require large initial expenditures in time and careful consideration of what primitives to provide.

Two principal topics were investigated within the organizational structure described above. They are segmentation, and the treatment of occlusions, shadows, and highlights. Segmentation is a principal issue in image analysis and has plagued researchers since vision research was begun. It is a necessary conjunct to scene analysis and there is good evidence to show that any general vision system will be limited, to some extent, by the ability of its low-level segmentation processes. Our approach to the problem is not novel as far as the operators used are concerned. In fact, there are a number of tools around which are well suited to specific partitioning operations. We contend that segmentation modules must be prepared to make use of any or all of them to achieve success with images of great complexity. We do break new ground, however, in the way we employ available techniques to effect reasonable partitions of natural scenes.

The primary mechanism used to effect isolation of regions possessing uniformity along some dimension is thresholding. The thresholding operation is performed for a specified parameter on a given image using cutoff limits obtained from histograms of various parameters for the picture. The cutoffs are determined from bounding minima of prominant peaks in the histograms. The parameters used include red, green, and blue sensory data which are provided by the digitization process; hue, saturation, and intensity information obtained from transformations performed upon the original data; and three additional television industry color components called "Y", "I", and "Q" which are obtained from a different set of transformations. Use of multiple sources of data often provides a peak along one dimension even though the same points indicate no dicontinuities for other parameters. This allows us to carry the thresholding process further than has been possible with only light intensity information.

The operations described above are utilized by a recursive algorithm which is continually applied to each of the resulting subregions of the picture. As a picture is subdivided, and smaller sections are considered, features that were concealed by large amounts of interfering data become prominant. This often results in new evidence for further thresholding applications. The process eventually halts when no significant portions of the picture remain which are not "uniform" in all dimensions.

One of the very difficult problems that we have had to handle is the treatment of texture. Very crude methods are employed to estimate regions possessing heavy texture. These estimates are then used to prohibit thresholding operations on heavily textured (busy) areas. In this way an attempt is made to isolate busy regions by elimination of surrounding homogeneous areas. If this is not successful, the original rough approximation is refined and accepted as the best available estimate of the textured portion. The thresholding operation is then applied to this data in an attempt to refine the busy region.

It is sometimes the case that scenes are not sufficiently rich in sensory variations to provide a number of peaks for any of the parameters which indicate cutoff limits for the thresholding operation. In these instances the aforementioned recursive algorithm halts early in the process, with few results to show for the effort. If we can find a reasonable breakdown for the remaining unanalyzed portions of the image we may be able to obtain small enough subimages to derive useful histograms. We extract such subimages by first isolating unprocessed homogeneous areas of the picture; the same is then done for busy areas. Histogram analysis and thresholding operations applied to these isolated subpictures can produce quite reasonable results. With these additional steps, analysis can be carried to a much further degree than would have been possible with the original algorithm. A modified procedure incorporating this additional isolation mechanism permits successful parsing of six scenes which range, in complexity, from a simple room to a panoramic view of the Pittsburgh skyline. The full power of the algorithm must be employed to achieve segmentation in this latter case. All in all, the results show the algorithm to be more powerful than any proposed in the past. Finer segmentation for more complex scenes is achieved than has heretofor been possible.

In contrast to the problem just discussed there has not been a great deal of research involved with the explicit investigation of the difficulties presented by occurrences of occlusions, shadows, and highlights in natural scenes. In our treatment of these phenomena, we have endeavored to accomplish two ends. We have attempted to formalize some of the knowledge concerning them and we have investigated various means of removing some of their ill effects. Our first goal is achieved by classifying occlusions, shadows, and highlights by types. In the case of occlusions this has resulted in six categories. These categories are determined along guidelines of decreasing continuity features. "Continuity" refers to those properties of a picture that indicate along which dimension an interrupted feature should be continued. The first type of occlusion requires full containment of one region within another; the last type is the completely hidden object.

This kind of categorization focuses attention on certain conditions that remain invariant within a given type. The most prevailing of these is proximity. One region cannot occlude or shadow a second region unless they border each other. A second invariant is similarity. Shadowed and highlighted areas must be similar, for some properties, to an adjoining region. On the other hand, regions which bear an occlusion relationship to one another must have a point of dissimilarity along some dimension (e.g., hue, range). The last invariant which we consider is continuity. For an occlusion to be present in a scene there must exist indicators of the actual extent of the obstructed object. The greater the degree of continuity, the greater the number of clues that point out proper reconstruction of borders and other region attributes. Exploiting these clues allows us to reformulate boundaries for some types of occlusion involving regular objects.

Continuity assumes less importance for shadows and highlights than it does for occlusions. In this case, similarity receives the major emphasis. As noted above, a shadowed region must resemble an adjoining region which depicts a normally lighted portion of the same surface. It is also true that there must exist some points of dissimilarity (e.g., intensity). Similarity of hue is found to be the most useful determiner of the presence of a shadow or highlight, while differences in intensity indicate which condition has occurred and to what degree.

Several other aspects investigated in this same period of research are not reported here because of lack of time. These include: identification procceses, representation issues, the use of high resolution picture inputs, analysis of human protocols to determine possible useful operators for image understanding, and details of the picture point accessing subsystem. These topics will be covered in forthcoming technical reports.

History of Past Research

In this section we present a brief history of scene analysis by computer, starting with the classic work of Roberts (1963). It is not intended that this be an exhaustive survey of the literature. We do hope that it will provide some insight for the interested reader into the basic trends in the research and into the major techniques that have been used. For a complete coverage of the field of image analysis see Rosenfeld (1969, 1969a, 1972, 1973).

Introduction

Research in picture processing has been going on for nearly twenty years, and work in scene analysis has been progressing for more than a decade (Rosenfeld, 1969, 1972, 1973). With only one or two exceptions, the contributions in this field have evaded the difficult problem of analysis of natural scenes. A good amount of the work has not been scene analysis at all, but an application of special techniques in highly restricted images to obtain limited information (Stevens, 1972; Strand, 1972; Sutton and Hall, 1972). In scene analysis the majority of investigation has centered around simple environments containing planar-faced objects. Motivation in this area has been provided by the robotics researchers (Feidman et al., 1969; Nilsson, 1969; Ejirl et al., 1971) and by those seeking techniques which could be generalized to more complex natural scenes (Roberts, 1963; Guzman, 1968; Winston, 1970; Waltz, 1972). In the more recent past there has been some investigation into real-world images (Bajcsy, 1972; Yakimovsky, 1973; Tenenbaum, 1973, 1974; Lieberman, 1974).

Roberts [1963]

In terms of scene analysis, computer vision starts with Roberts. His work spanned the entire field from camera input to interpretation of planar-surfaced objects. The bulk of his work concentrated on the aspects of representing and recognizing three-dimensional objects.

His program is conceptually divided into three main processes. An input process produces a line drawing from a photograph. The line structure is input to a 3-D construction module which produces a three-dimensional object which is compared to given models and classified. The final 3-D display program outputs a two-dimensional projection from any point of view.

The picture is input through a facsimile scanner and is quantized to a 256 X 256 raster with eight bits of intensity information. A new raster is thresholded from the output of a local differential operator which detects edges in the picture. The process continues with the applications of correlations of line fit at selected points in the differential picture which meet a specified threshold level. In this way a set of feature points is obtained. A small number of heuristics are now used to connect feature points and to eliminate multiple interconnections and spurs. Straight lines are fitted to the sequences of points by a sequential least-mean-square error-fitting routine. Line fitting and merging of lines is the last step in the procedure.

Input to the second part of the program consists of a planar line drawing generated from the first section or from the output of the 3-D display process. The lines should be a perspective projection of the surface boundaries of a set of three-dimensional planar objects. A three-dimensional description of the object(s) shown in the drawing, in terms of models and their transformations, is produced. The models used in Roberts' system are a cube, a wedge, and a hexagonal prism. The models can be translated, rotated, and extended in any dimension so that a model will match any structure which differs only in orientation or size. The cube, for instance, will match any parallelpiped. The models are not allowed to vary in perspective or skew.

The program attempts to find all polygons from the line drawings. Lists of convex polygons, concave polygons, and exterior boundaries are kept with associated points and lines noted. The next phase of the program cycles through the structure attempting to match topological features with one of the models. It does this in a series of four steps, applying the next step only if the previous one failed:

1) Locate a point which is completely surrounded by approved polygons.

2) Locate a line with approved polygons on either side.

3) Test each remain approved polygon one side of which is attached to a vertex.

4) Compare each three-line vertex with the models.

From one of the previous steps the program finds point-pairs between the picture and a model, and applies a similarity test to get the best transformation and a mean-square error of fit. If the error is less than some threshold the transformation is accepted. The object recognized is deleted from the scene and the process considers remaining objects.

Roberts' treatment of complex objects composed of conglomerates of instances of his simple models is interesting. Lines denoting the boundaries of juxtaposed simple objects are missing on input. If a simple object is recognized it is deleted from the scene and the three-dimensional representation is back-projected onto the scene to locate lines demarcating the recognized object and the adjoining structure. In this way new boundaries are discovered and additional simple objects are interpreted. A linkage of parts of the composite object is obtained and one depth assumed for all.

A support theorem assuming a ground plane is postulated to determine depth. For simplicity, all scenes are assumed to be upright with all objects touching the ground plane. If the camera model is known, absolute depth can be determined.

The 3-D display portion of the program can project a three-dimensional object from any orientation at any location. The display procedure is especially noteworthy in its handling of hidden line elimination.

Roberts' work was an important initial effort that set the tone for practically all block model systems that followed. As has been noted, the system is complete from preprocessing through recognition to display of the final output. Its weakness lies in the preprocessing section which must have almost perfect camera input. As we will see below, many of the researchers that follow Roberts seek to improve on this phase of the process.

Guzman [1968]

Guzman's work is distinguished by the fact that he proceeded to isolate planar-surfaced objects in complex scenes, utilizing a rather limited set of local heuristics based on the properties of vertices. Near-perfect line drawings are assumed as input to his program.

Regions, i.e., surfaces bounded by simply closed curves, are linked according to rules associated with certain types of vertices. Linkages may be inhibited by neighbors of certain types. The linked regions, known as nuclei, are further linked to form maximal nuclei under the following three rules:

1) If two nuclei are linked by two or more strong links they are merged.

2) If two nuclei are joined by a strong link and a weak link they are merged.

3) If a nucleus consists of a single region, has one link with another nucleus, and no other links with other nuclei it is merged with the second nucleus.

Each rule is applied until no maximal nuclei can be formed, before the next rule is considered. The final nuclei constitute the isolated objects.

Guzman's program sometimes makes errors by clumping objects when exterior lines are missing. He describes how his program could be used with stereo perceptions to obtain depth information.

Falk [1970]

Falk's program utilizes a vertex labelling scheme to catalogue interpretations of vertices and segment a scene relative to links formed between these vertices. His work is more general than Guzman's in that correct segmentation can occur despite missing or partial lines. After body separation, completion routines using heuristics based on collinearities and extension vertices are called to determine occlusion relations and insert missing lines and line segments.

Like Roberts, Falk uses a restricted set of models for matching objects in the scene. The number of sides, faces, and vertices of the model for different views are stored. These properties are compared with those of objects in the scene to compile a list of possible candidates for matching. Final choice of an object is based on a comparison of feature vectors extracted from physical properties computed from the image using hypotheses of ground plane and object support.

As objects are identified and located they are back-projected in a manner similar to the technique used by Roberts. The projected drawing can be compared for closeness of tolerances between original and back-projected lines.

Winston [1970]

Winston took off from Guzman's work and used local clues to recognize objects and to discover structural relations between objects and groups of objects. Basically, his system starts with line drawings and uses Guzman's program to segment the scene. From this point the program has three choices. An attempt may be made to identify the entire scene by matching it with a known model or series of models. Another goal can be to find an instance of some particular model in the scene. The third alternative is to use the scene description to help form new models of structural concepts.

Winston's system learns through presentation of scenes to the program. Conceptual models are formed an a network type data structures by combining simpler concepts or relations between certain types of objects. For example, from the relations of "support" and "marry" and three brick-like objects the concept of "arch" can be derived. An interesting point in Winston's work is that the process of model acquisition can learn as much or more from near-miss situations as from correct examples. In the case of an arch, for instance, a near-miss presentation can indicate that the two supports are not permitted to "marry".

Brice and Fennema [1970]

Brice and Fennema came up with a new approach to segmentation in the world of blocks. They attempted the direct transformation of a gray-scale picture to regions, bypassing the edge-finding procedures.

Atomic regions are initially formed by collecting all connected points of the same intensity. Points p1 and p2 are said to be connected if there exists a sequence of points, the first of which is p1 and the last of which is p2, and if the consecutive points are neighbors. By "neighbors" is meant the four non-diagonally adjacent points. These atomic regions are then merged by melting boundaries if they meet certain criteria.

Two heuristics are used to guide the merging of regions. Strong boundaries are never disolved, but even if the boundary is weak, regions are joined only if the resultant boundary does not grow too fast. Since interest is in the weak part of a boundary, define W to be the number of boundary vectors having a strength less than some threshold, t1. Then two regions are merged if W/PM is greater than some threshold, t2, where PM = min(P1,P2); P1 is the perimeter of the first region, and P2 is the perimeter of the second. If t2 is small, many regions may be joined. If it is large, two regions are merged only if one of the regions almost surrounds the other.

The second heuristic joins regions solely on the basis of the strength of the boundary that separates them. Two regions are merged if W/I is greater than some threshold, where W is defined as before and I is the intersection of the two regions. Although this heuristic is more natural than the first, it is too local to be used alone. If it were applied before the first heuristic, the result would be to wipe out almost all of the regions.

To make the segmentation more amenable to scene analysis, a line-fitting program is applied to the ouput of the region grower. The operation consists of three passes, each of which applies increasingly larger masks to successive points and fits a line approximation.

The scene analytic portion of the program attempts to identify the output of the line-fitter, using local clues, with a two-dimensional description of the object. At this point Brice and Fennema are working with imperfect data. Lines are missing or broken, and objects are occluded. Basically, the scene analyzer extracts easily recognized regions first (e.g., wall and floor), groups the regions, using a Guzman type technique, into objects, and then tries to recognize the faces of objects. If recognition fails, it proposes lines, regroups regions, and begins again.

Semantic information is used to extract wall and floor regions. For example:

1) Floor and wall are separated by a baseboard of known height.

2) Floor and walls are light in intensity.

3) Wall is high in the picture.

4) Floor is low in the picture.

As the authors point out, there are several weaknesses to their system. The criteria for region growing are very simple, and sophisticated techniques utilizing feature vectors need to be developed. The line-fitting process is not nearly as accurate as it could be. Only fairly simple scenes can be analysed with the present unsophisticated recognizer.

Kelly [1970]

Kelly's work is the first substantial system to treat naturally occurring visual scenes. His program chooses, from a collection of pictures of people, those pictures that depict the same person. The program works by finding the location of features such as eyes, nose, or shoulders in the images and classifying people on the basis of measurements of distances between pairs of such features.

Kelly uses a number of methods developed by previous researchers. The position of the body is found by subtraction of the background. The top of the head and the feet are found by template matching. The outlines of the head, neck, and shoulders are found by edge-detection operators. The eyes are found by dynamic threshold setting followed by smoothing and template matching. The nose is located by dynamic threshold setting. The mouth is located with a line detection operator. All of these methods are applied heuristically in a manner based on an implicit model of the structure of the human body. After the measurements between features are extracted, pattern classification techniques are used to identify the body.

Two facts are worthy of emphasis in this work. The first is the basic goal-directed nature of the search for objects believed to be present. This implies the use of context or semantics to aid the process. The program searches for basic parts that can be most reliably detected, e.g., the head. Once one part is identified, this can lead to a search for another portion known to bear a certain relation to the first part, e.g., the eyes. As more parts are found, more confidence is gained that a certain object has been found. In addition, as recognition proceeds the work to be done diminishes. The goal-oriented behavior of the system is directed and driven by a model of the object desired.

The second factor to be noted is Kelly's use of planning to reduce the search space. A new reduced picture derived by averaging and application of an edge-operator is prepared from the original. This gives us a simplified model to work from. The objects that remain in the reduced picture are likely to be important features. Since the reduced image is smoothed, noise is diminished. Objects are now tentatively identified in the reduced picture which serve as a plan to verify the presence of edges in the original. The planning scheme has the advantage of speed and does lead to clean and complete edge outlines in a complex environment.

Barrow and Popplestone [1971]

Barrow and Popplestone, working on the robot project at the University of Edinburgh, also departed from the previous preoccupation with planar-surfaced objects. They constructed a system that will recognize a small range of objects including a cup, a wedge, a hammer, a pencil, and a pair of spectacles. The digitized pictures are initially analyzed for regions within a small range of brightness. Regions are merged if the average contrast across the two boundaries is less than some threshold. The last step in the process is weeding out very small regions (which are probably spurious), and those with weak boundaries (probably part of the background which is not represented).

The next step is to construct a feature vector for the regions. Each property is chosen so that it is invariant for a limited range and class of movements of the object in a field of view. Properties that are calculated include compactness and shape. Relations include bigger, adjacent, distance, convexity, above, and beside. Once these features are extracted, the remaining problem is to find the best match of a subset of the picture regions with a subset of the model regions. This is done by a graph search of region descriptions, utilizing prior information from the partial match developed so far. Further correspondences are considered only if they are especially promising. The best match so far encountered is remembered, and when no more promising lines of development are available, this will constitute the interpretation.

Model generation is accomplished through a learning sequence. An object is placed before the camera and analyzed into regions. The regions are exhaustively described in terms of properties and relations. The system is then provided with a view of the object which is the correct response, and with correspondences between regions in the picture and in the view. Comparisons are made after updating to note discrepancies.

The system descibed above is weak from several aspects. Only single objects are recognized. Shadows may result in additional regions being generated. Edges may be blurred by tricks of lighting. Finally, occlusion can give different property and relation measures for the objects.

Waltz [1972]

Waltz' system reconstructs three-dimensional descriptions from line drawings which are obtained from scenes composed of plane-surfaced objects under various lighting conditions. In this description, shadow lines and regions are identified, regions which belong to the same object are grouped; support, in-front-of, and behind relations between objects are denoted; and information about spatial orientation are noted.

The techniques of Clowes (1971) and Huffman (1971) are used to label vertices in accordance with their possible interpretations. Each label at a vertex assigns a specific label to each one of its lines. In this way most of the possible edge interpretations, and even lighting conditions on the side, are covered. A filter program with a set of combination rules is now applied to check the inter-consistency of two sets of vertex labels for each line. Inconsistent labels are deleted. A surprisingly large number of unique labels were found in this manner. A full tree-search for consistency is performed if any of the labels are still not unique. The resulting labelling determines segmentation of the scene by case analysis. If the result is not unique, then several interpretations are possible, which would also be the case with humans. From this point the program goes on to treat certain cases of missing line segments and to derive support relations and orientation data.

This work differs from those previously described in several ways. In the first place a much broader range of scene types, but fewer object types, can be dealt with. Ambiguity is also dealt with in a natural manner by eliminating the impossible cases rather than selecting the most probable. Another point of departure is that the program is algorithmic and does not require back-up facilities if the filter program finds an adequate description. Lastly, the use of a descriptive language and powerful case analysis can be used to understand previous work in the field (e.g. Guzman).

Shirai [1972]

Shirai constructed a system to recognize polyhedra in a scene, working directly on the digitized picture. This process is chiefly of interest in the use made of heterarchical structures. Data is analyzed and lines are looked for with a general concept of "body" as a guide. The information already gained is used to further complete an object. This is in contrast to the previous schemes that have proceeded in a hierarchical fashion to extract successively higher abstractions.

Basically, the program looks for lines at concave junctions or at other suggestive places. Once evidence of a line is found, the program tracks along that line looking for vertices or extensions, with the global context of the object available. Implications of the evidence found so far is assimilated as the process tracks the lines. Decisions as to objects are made as sufficient evidence is obtained.

Bajcsy [1972]

This research represents the only significant example that we have come across of utilization of texture analysis to segment complex natural scenes. Analysis of the power spectrum produces measurements for orientation, contrast, size, spacing, and, in periodic cases, the locations of texture elements. The local descriptors are defined over windows of various sizes. Region growing is based on non-contextual properties of texture and color. The non-local properties of the transform give poor edge and position information.

Sakai et. al. [1973]

A face recognition system was constructed by Sakai, Nagao, and Kanade using a hypothesize and test paradigm. A Laplacian edge-operator is used to obtain a line-like picture from the input gray-level picture. Context-dependent masks are used to locate easily recognized portions of the face (e.g., top of the head). As certain portions are found and analyzed, new subroutines are called which locate and analyse more difficult features. If a portion is not located, constraints are relaxed or another feature is looked for, depending on how much has been identified thus far. Models of features are used to provide the necessary global context.

Grape [1973]

Grape has come up with yet another system to identify convex planar-faced objects. His chief contribution is the use of global models to guide locally based decisions in the parsing of scenes. His program operates satisfactorily in the presence of such adverse conditions as noise, shadows, glare, and missing line segments.

The preprocessing phase of the program consists of utilizing the edge follower of Pingle and Tenenbaum (1971) to extract edges from the picture, and following with a line extraction program which fits lines to edges by a least-square method. Some conservative line extension is performed here, but the resulting output may have missing line segments. Recognition then proceeds by linking lines together by possible vertices. Cross-reference tables are formed which map the relationships between lines in terms of intersections and collinearities. Links are now created between scene elements and model elements. The links are investigated in order of decreasing complexity until a complete object is found or the links are exhausted, which in the latter case results in the best match being chosen. The final phase of the program is object completion where lines still not accounted for are considered with partially matched objects to see if they can complete the recognition.

Yakimovsky [1973]

Yakimovsky developed a system to analyze complex natural scenes, utilizing a semantic base to segment the scenes. Initially, sample points are selected from the quantized picture which are assumed to be representative of different regions. Separation between sample points of distances that range from 5 to 20 are used. Local operators are applied to determine dominant color and intensity around the sample points. Points are assumed to be in the same region if a vector of available features differs by less than some conservative threshold. A non-semantic region grower, which melts the weakest boundary in the current image, is now applied to the picture. Boundary strength in this phase is based on average differences between boundary sample points, feature vectors, and the length of the boundary. At each step the boundary and region structures are updated. The new region is merely the union of the points included in both of the former regions. The boundary structure update is more elaborate since it is ordered and thus requires a detailed algorithm. The merging process stops when the weakest boundary in the current image surpasses some threshold.

The next step in the process consists in region growing on the basis of the world model. The model is input through a learning process which imposes a statistical measure on the features of the designated regions. A probability estimate, that a specified region will have a certain interpretation given the feature meaurements, is formed. These features include such properties and relations as size, vertical position, horizontal position, boundary touching (top, bottom, sides of) the picture frame, average light intensity, color saturation, color hue, and some rough shape measurements. The boundary strength is calculated as the Bayesian probability that, given the properties of the boundary and two regions defining it, the boundary separates sub-parts of images of different objects. The process stops when the weakest boundary surpasses a certain threshold or when a good interpretation for the current segmentation is reached.

Yakimovsky's system was shown to work quite well on two picture domains. The first domain consisted of road scenes as may be seen while driving. The second domain was left ventricular angiograms.

Tenenbaum et. al. [1973, 1974]

As we have progressed through our survey, we have noticed a gradual shift of interest to complex natural scenes. Tenenbaum is currently developing a system which uses sensory data from several sources to extract features of objects. His current aim is not to exhaustively describe a scene, but to locate pre-specified objects. He feels that eventually the system will use planning in selecting appropriate methods to extract the most meaningful features for discriminating and locating objects.

The wealth of information and complexity of detail that make many of the techniques of the world of blocks unusable in naturally occurring scenes can be exploited in real-world environments to give a variety of attributes. Easily extracted features should be used first to distinguish an object, resorting to the more expensive properties only if necessary. Some of the features used by Tenenbaum include color hue, color saturation, height, depth, and surface orientation.

The search for an object is intended to proceed in two phases, called acquisition and validation. Acquisition proceeds by sampling the image for characteristic attributes of the desired object based on a model. If planning is good, obviously irrelevant areas of the scene will be rapidly disqualified so as to concentrate efforts in most promising locations. The features tested will be in order of the most discriminatory first. Contextual knowledge can be used to direct the search. For instance, sampling may be localized to the vicinity of known objects (e.g., a tabletop is located within a certain range of height from the floor). It may be simpler to look for more easily distinguishable objects for which the desired object has a known relation. For example, telephones are small and known to lie on desks.

Validation consists of using more computationally expensive features to verify that the candidate regions obtained in the acquisition phase is the genuine article. Additional evidence may be gained by looking for shape or textural attributes and by verifying additional surfaces and known contextual relationships. Validation proceeds as a sequential decision process; after each feature is considered, a decision must be made whether to accept the original acquisition hypothesis, to reject it (and continue sampling), or to continue the validation process.

Implementation of the proposed system is proceeding on two fronts. To verify experimentally the basic premise of distinguishing objects by easily extractable features in constrained contexts, an interactive system has been constructed that allows the investigator to apply specified primitive operators to graphically designated areas of the scene and to observe the results in pictorial form. The operators extract a variety of local attributes (e.g., height, hue, saturation, surface orientation, range) from input arrays of color and range data. The attributes can be extracted by pointing to the image for a local value or by outlining a region to obtain an average value. By outlining the principal objects in a scene the investigator can obtain the information necessary to develop a perceptual strategy for distinguishing them. The adequacy of a given set of attributes can be tested by requesting the system to indicate, graphically, all points in a scene satisfying the specified predicate.

Concurrently, a system for automatic planning and execution of distinguishing feature strategies is being implemented. Initially, objects will be described directly in terms of their distinguishing features. Basic planning will involve first determining a subset of attributes sufficient to distinguish the goal object in a given context and then ranking those attributes to determine a cost-effective testing sequence. The utility of a direct search is contrasted with the total effort required for indirect acquisition through contextually related objects. The system can dynamically alter its strategy during execution as utility estimates get updated by results of the tests.

Tenenbaum's system is by far the most ambitious project to date in terms of general applicability to complex natural scenes. The current interactive program indicates that most objects in an office environment can be distinguished by a small number of the available features. The basic concept holds great promise for the investigation of theories and strategies applicable to general vision systems.

Lieberman [1974]

Lieberman has constructed a complete system that identifies a limited number of objects in relatively uncomplicated outdoor scenes. Segmentation is primitive and provides only a very rough partitioning. His main contribution lies in the use of complex texture analysis to make some three-dimensional inferences concerning objects in the picture. He is, for example, able to identify the ground plane in a scene with a large meadow. He also makes a good case for the use of semantic nets to model objects of an image and the relationships that they bear to one another.

Our work does not rely heavily on any one of the individuals cited above. We are rather indebted to all those researchers who have gone before us to show the way. They have provided a wealth of operators, material, and insights to be built upon and utilized for advancement of the field.

2 MODEL, METHODOLOGY AND MATERIALS

This chapter gives some motivational background for the work undertaken in this body of research and is not critical to the understanding of the chapters that follow. In addition we furnish some brief details on materials available to us. More specifically, we discuss the model that provides the necessary framework for a continuing study of the perception task, some methodologies which assist in undertaking large difficult tasks, and the materials required to carry on the work.

The Model

Although the work presented in this dissertation is not, in its present stage, dependent upon any larger system organization, we would like to discuss the model that provides the framework for ongoing research. In our opinion the vision models proposed in the past suffer from a number of deficiencies that preclude them from contributing in a major way to solutions for complex perceptual tasks. Hierarchical structures are not flexible enough to adapt to systems which will have to make use of many diverse kinds of knowledge. They suffer from a lack of communication as well as the inability to make graceful error recovery. Heterarchical systems have the problem of deciding which module is going to communicate what details to which other modules. This leads to the problem of restructuring each time a new knowledge source is added to the system. To overcome these deficiencies the vision group at Carnegie-Mellon University has decided to make use of a model based on the hypothesize-and-test paradigm that has found successful application in the Hearsay speech understanding system (Reddy et al., 1973a, 1973b; Lesser et al., 1974).

The chosen model is organized around independent sources of knowledge which cooperate through a global data base. Figure 2.1 provides a conceptual view of the system and the types of knowledge which we feel are necessary for a general vision system. As research continues and a better understanding of the system is obtained details of configuration are likely to change. The data base is organized in an hierarchical ordering of multiple representations (figure 2.2). Ideally, each knowledge source can consult the base to see if enough information is present to hypothesize new representations or verify results proposed by other modules. Practically speaking, initial configurations of the system require a controller to evoke correct responses.

A system organization of this type has several important features. First, a framework is provided within which research can be conducted without undue concern given to interfacing issues. The same features of relative module independence that make this possible also allow reformulation of indivdual knowledge sources, or even their removal, without major restructuring of the system. With the additional requirement that removal of any one knowledge module does not fatally cripple the system, knowledge sources can be evaluated for their specific contributions. Another feature of the model is that all reasonable region formulations can be hypothesized and treated systematically. This is an important advance over systems which allow only a single representation for any specific portion of the picture. Relatively simple

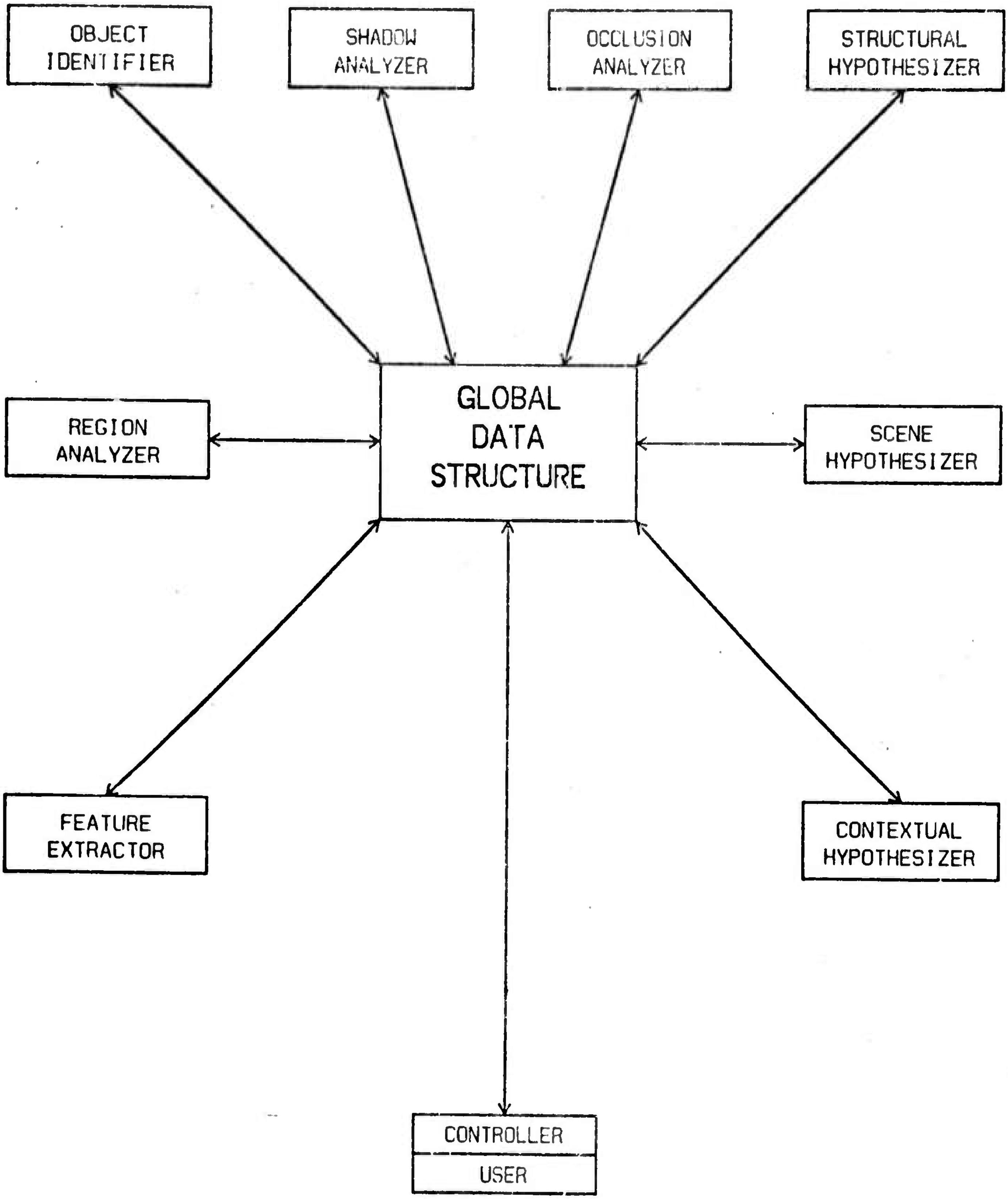


Figure 2.1. A model for computer vision.

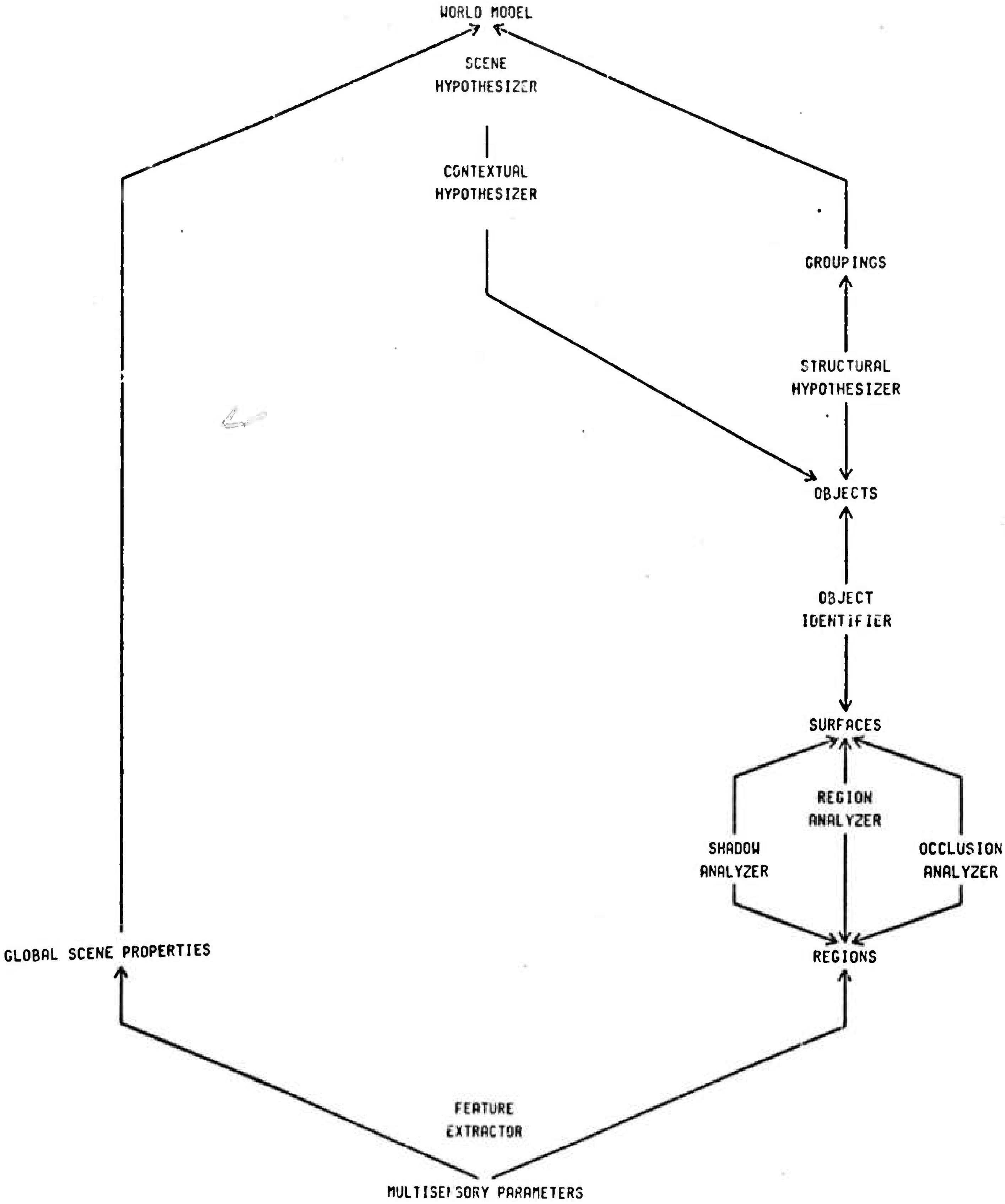


Figure 2.2. Organization of the global data base.

error recovery can be achieved by shifting the focus of interest to an alternate representation for subsequent analysis. This last property is critical in perceptual systems which start with errorful data and then compound the problem with imperfect mechanisms which analyze the data.

The model which furnishes the above features provides the necessary requirements for constructing complex AI systems. Sources of knowledge can be developed independently and assimilated into the system structure as they become available. The research in the next two chapters was carried out with this eventuality in mind.

The Methodology

In many cases, we do not have sufficient theoretical understanding of a task to implement the sources of knowledge necessary to its solution. Such is the case for computer analysis of natural scenes. This raises the question of knowledge acquisition, which has proven to be a source of difficulty for many AI systems. Since we do not, in general, know how to build systems which have the ability to "learn" by making inductive inferences from their environments, we must consider other options for developing the necessary knowledge. In the case of segmentation, where much experimentation has gone before, we can attempt to explore the capability of existing operators and seek to discover ways of extending the operations to natural scenes.

When there is no previous body of research upon which to build, an attractive alternative is to pursue the policy of what Woods and Makoul call incremental simulation (1973). This method utilizes human personnel to fill the roles of part or all of a knowledge source. The experimenter, for instance, takes the place of the shadow and occlusion module. He starts by removing occlusion, shadows, or highlights from scenes presented to him. As he begins to understand the issues he automates some of the necessary responses. New information is imparted to the knowledge source as it becomes available. In time the human user gradually replaces himself with a computer program.

In both approaches to the problem discussed above, an initial computer system is necessary to provide primitive constructs for manipulating the data. This in itself is a non-trivial task. In our own case, considerable time was spent in implementing interactive systems to provide suitable responses for the experimenter. For segmentation, a system to access the digitized picture data was the first step. Upon this foundation we built a large number of picture operators which could be evoked by the user as he saw fit. In the case of occlusions, shadows, and highlights a graphic subsystem that would provide the required visual response to the user was constructed. In addition to this, a number of operators that could manipulate the underlying data structure were supplied. This gave the experimenter the necessary tools to investigate the problems of "restoring" affected areas to their actual state.

An important feature of the methodology discussed is that it provides for partial implementation of the model described in the previous section. This means that it is not critical that all knowledge modules be available to study the functioning of specific sources of knowledge or the system in general.

Materials

In this section we will briefly describe the various data and tools available for research.

The Computer System

Figure 2.3 illustrates the computational facilities available at Carnegie-Mellon University at this time. Eventually we expect to move the bulk of the segmentation system to the PDP-11/40 where a dedicated machine and specialized processors will be available.

The PDP-10 and its peripherals serve as the main processing device. At this time the majority of software is resident here.

The Graphic Display Units are vector display devices which find many uses in our system. One use is to exhibit various size windows of gray-scale representations of the digitized data for further processing in an interactive mode. The units are also programmed to produce vector drawings of regions resulting from various stages of the segmentation and recognition phases.

The Video Monitor displays high resolution, video, color or black and white pictures from input matrices of gray-level density. Resolution is currently dependent upon memory speed and capacity limitations. Performance will improve when the unit is interfaced to the 11/40 machine and will realize full potential with its connection to the C.mmp multi-miniprocessor machine. Eventually the monitor will provide most of the services currently furnished by the grahics displays.

The Xerox Graphic Printer furnishes hard-copy gray-scale representations of video pictures. It also has the capability of reproducing binary pictures or results of edge operations performed on various sensory parameters.

The SPS-41 is a high-speed multi-processor best utilized for computationally expensive picture operations.

The Sensory Data Base

The sensory data base consists of a number of files of binary data derived from twenty-seven photographs of natural scenes. A color print of each scene was quantized into 256 density levels through red, green, and blue filters to yield three files of 600 X 800 X 8 bit resolution.[1] Care was taken to ensure a broad experimental base by collecting images which display a wide range of visual stimuli such as color, texture, shape and structural complexity. Indoor scenes range from those of severe simplicity with single objects of modern furniture to a view of an office containing a

[1] The pictures were digitized at the Engineering Control lab at the University of Southern California. Our thanks are given for their cooperation in this research effort.

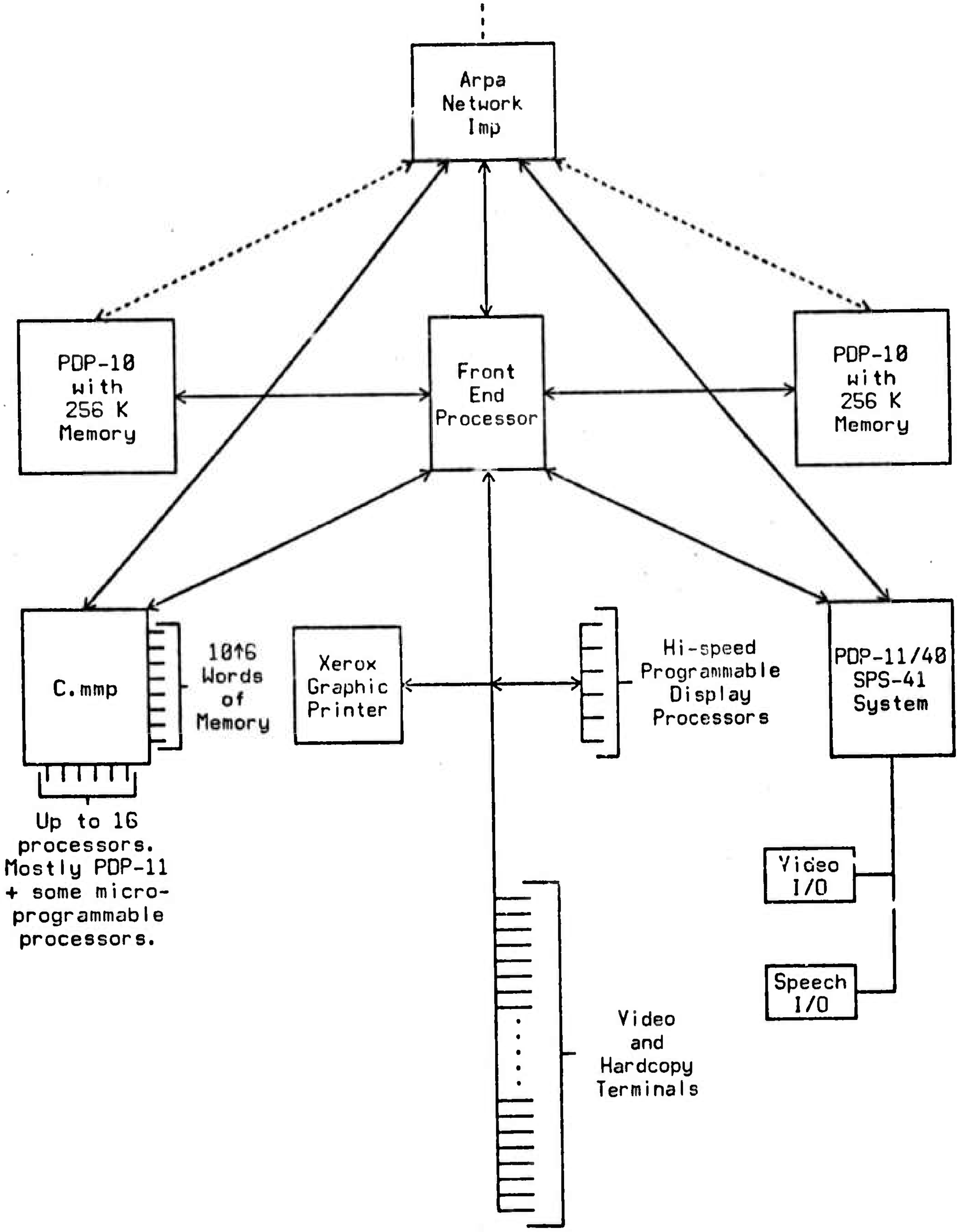


Figure 2.3. Computer facilities.

large number of items of all shapes and sizes. Outdoor compositions include animal and human subjects, panoramic views, skylines, large edifaces, expanses of greenery, small buildings, and automobiles.

From this library of subjects six images were selected for experimental purposes. They are shown in figures 2.4.a through 2.4.f and are presented in increasing order of structural complexity. We tried to ensure that the collection was representative of a wide class of everyday scenes. The girl was included to compare our segmentation with results that have been obtained in the face recognition systems. The room was thought to be a good scene for initial investigation. It contains a fair degree of complexity but lacks heavily textured areas of any size. The house provided a logical follow-up to the room: it possesses rich color properties and relatively large areas of strong texture. The car presented problems in respect to its lack of regular shapes and its reflective surfaces. The bear clearly lacks areas of definitive shape; here the difficulties in segmentation are aggravated by the background of "colorless" rocks. It was thought that the skyline scene presents the ultimate challenge to the segmentation process and should produce the most interesting results; the amount of detail and complexity is almost overwhelming. We were to discover later these difficulties were compounded further by a lack of clearly defined color properties. All in all, we believe the selection presents a broad enough range of scenes to ensure that their successful segmentation constitutes a significant advance in the field of computer vision.

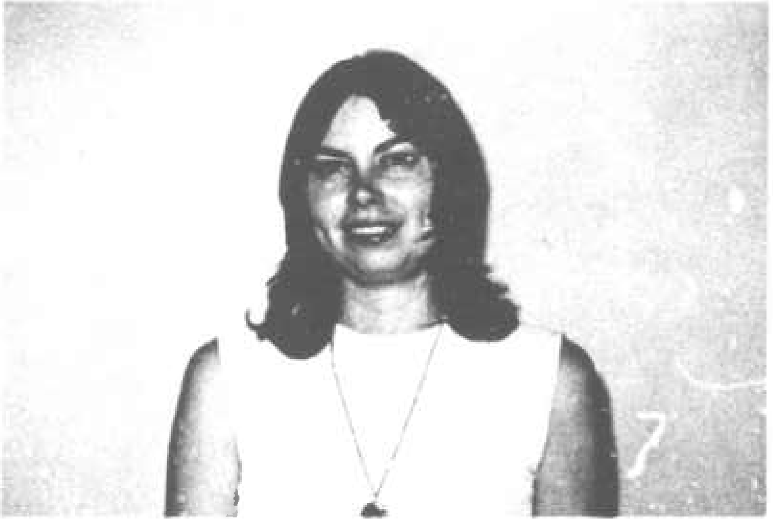


Figure 2.4.a. Girl scene.

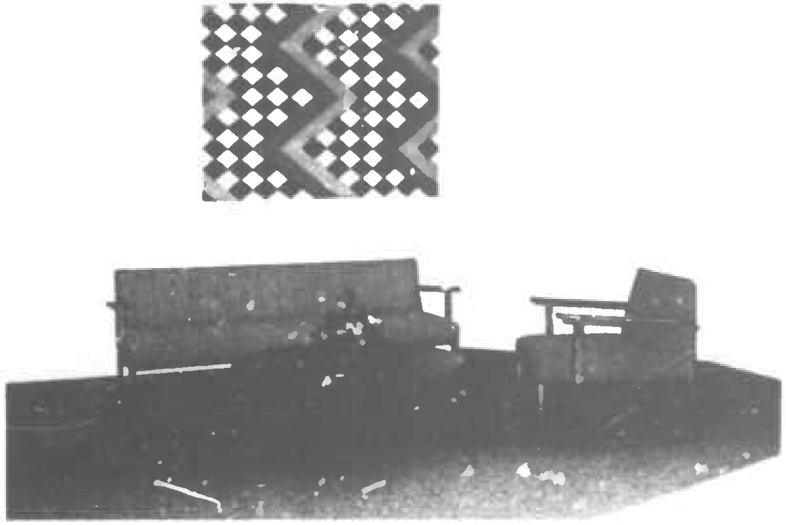


Figure 2.4.b. Room scene.

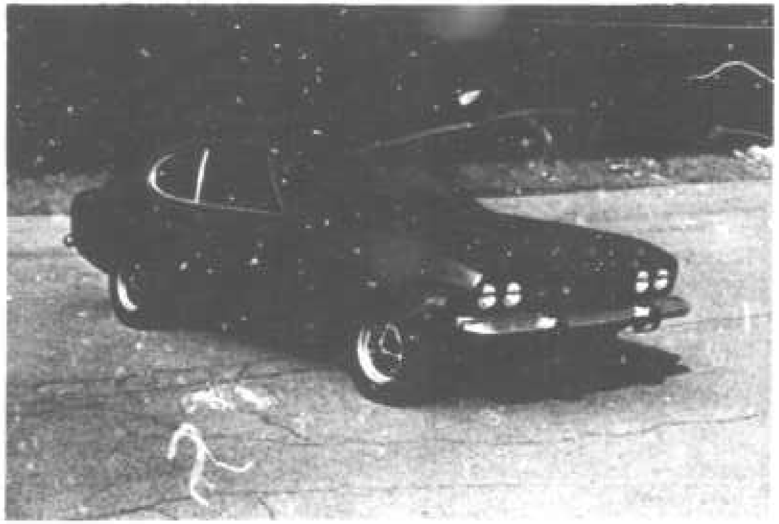


Figure 2.4.d. Car scene.

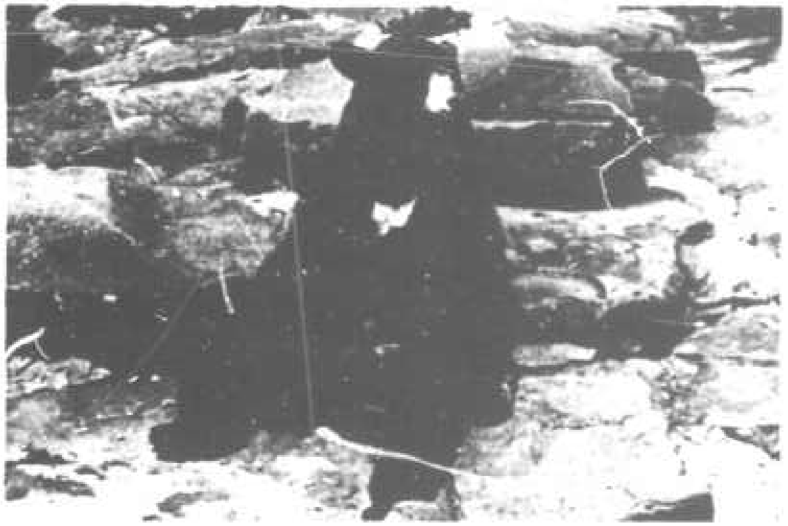


Figure 2.4.e. Bear scene.

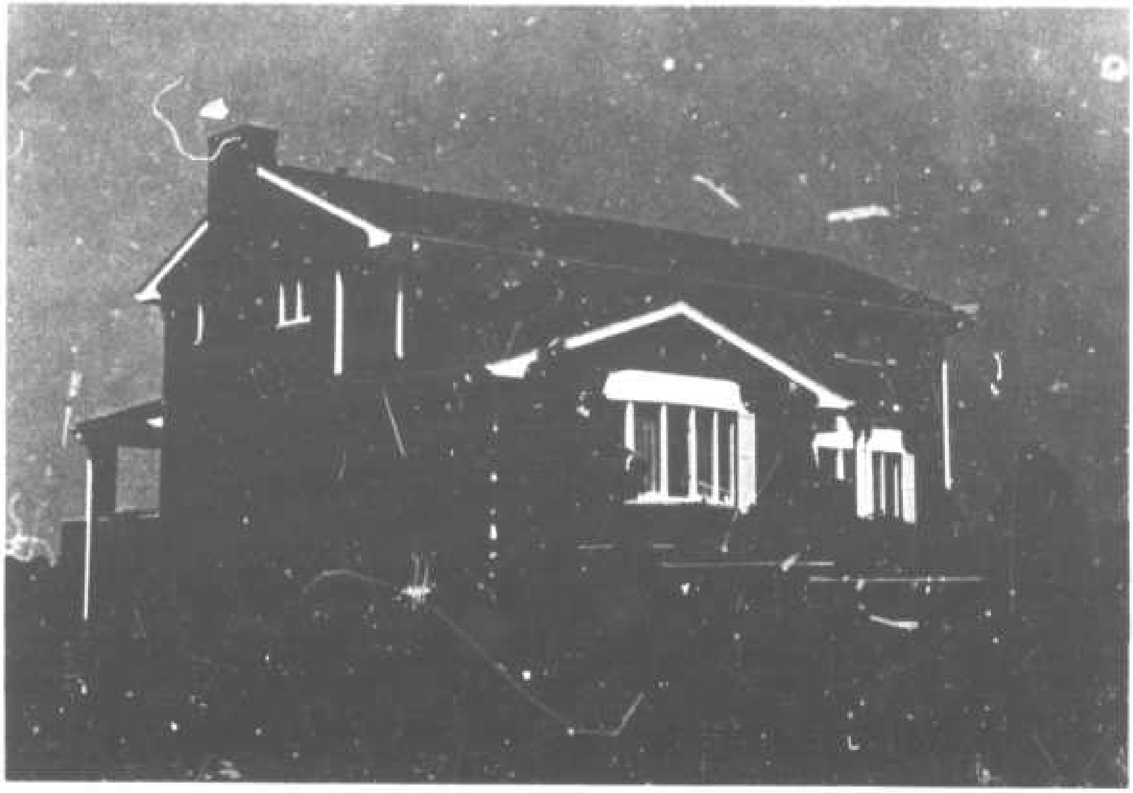


Figure 2.4.c. House scene.

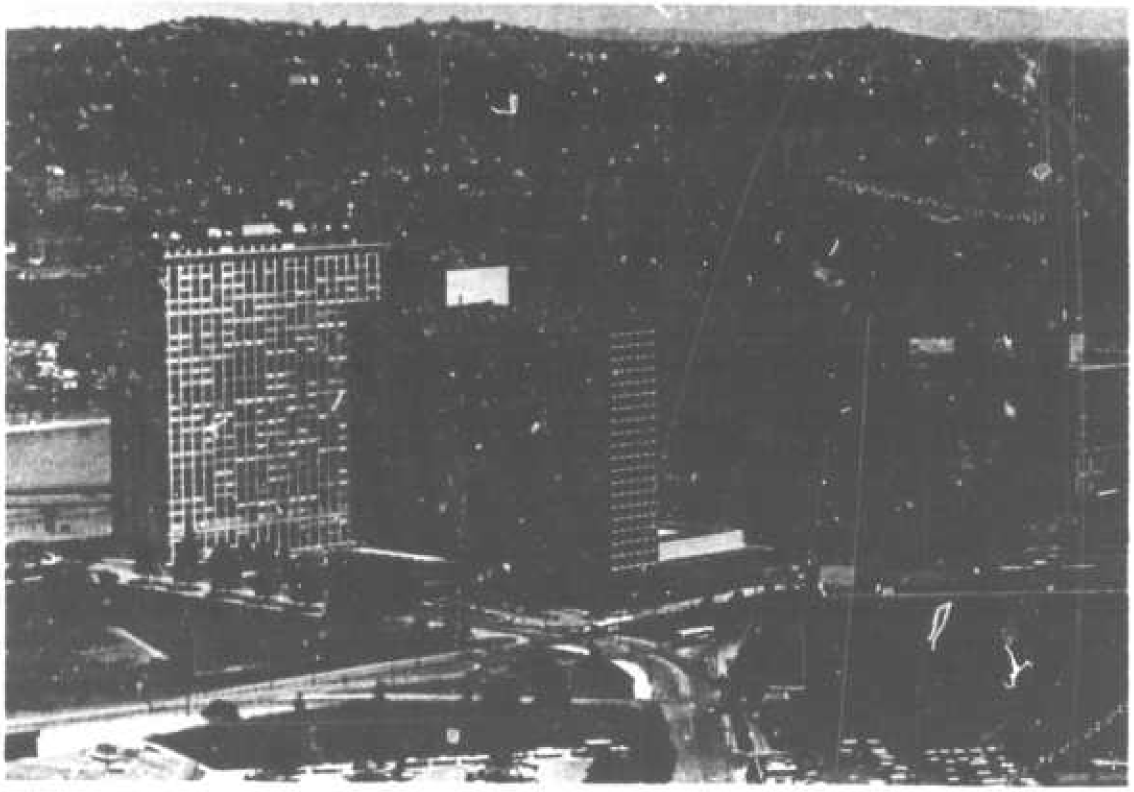


Figure 2.4.f. Skyline scene.

3 SEGMENTATION

Segmentation, as we have come to use the term in the area of computer vision, is the partitioning of an image into some number of isolated areas which possess uniformity along some dimension. It constitutes the single greatest stumbling block to progress in the area of computer vision. This is as true today as it was a decade or more ago when early AI researchers first discovered the complexity of the problem. In this chapter we want to point out the importance of proper segmentation in the analysis of natural scenes, why the problem is so difficult to solve, and why the problem will remain formidable for some time to come. We will also discuss some of the major techniques that have been used in the past to achieve segmentation and why no one of them, by itself, is likely to be successful in its application to natural scenes of even moderate complexity. Finally, we shall present our own investigation into the problem and give our step by step derivation of a unified process which has attained some degree of success in partitioning a series of six pictures of moderate to great complexity.

Necessity of Segmentation

It has come to be implicitly understood in the area of scene understanding that object recognition is incomplete without some more or less accurate delimitation of its shape.[1] In this respect, segmentation has come to be an indispensible concomitant of the scene analytic process. The end result of analysis must always be to partition the scene into regions which depict objects of interest. The question is how such a segmentation can be achieved. There have been two predominant control structures employed in the past to realize this end. The most familiar application has been the bottom-up approach used by most block world recognition systems (Roberts, 1963; Winston, 1970; Grape, 1973). Usually, an initial partition is obtained by some edge-detection and edge-following techniques. Properties of the segments thus extracted are then evaluated, and regions are continually refined and/or combined to arrive at a final partition which bounds objects of interest within the image. The second, much less commonly employed method of segmentation has been the utilization of goal-directed techniques to actively locate regions which possess a number of specific properties. The best examples of this approach are the face recognition programs of Kelly (1970) and Sakai et al. (1973).

Goal directed techniques alone will not prove to be sufficient for general understanding of complex natural scenes. An initial partition is necessary for at least two reasons. In the first place, segmentation is required to provide needed impetus for higher level analysis. In a system which handles a wide range of scenes, clues will be needed to establish proper context. In the second place, given an accurate segmentation, it is much easier to recognize objects on the basis of derived features than to actively search for the desired properties. Indeed, in some cases it would be

[1] The word "object" has come to have a special connotation for computer vision. It is often used to refer to a portion of a real world entity as bounded by the picture frame of reference (e.g., sky or meadow).

extremely difficult to establish accurate models which could motivate a top-down segmentation. For example, how would one specify the necessary features to constrain a goal directed search for rocks in the bear scene (figure 2.4.e)? It is not sufficient to look for something gray at ground level. It would likewise be a hopeless task to try to specify shape parameters. It is possible, however, to establish a rough first order subdivision of the rocks by segmentation techniques (see the section on results). It would be at this point that top-down mechanisms could exercise their proper function: to establish identifications, to refine first level approximations, to verify low-level hypotheses, and to probe for objects that have been missed by the partioning process.

Investigation of the Problem

The problem in achieving an adequate segmentation is dependent upon devising techniques to detect properties of uniformity among the picture elements, and then isolating the clusters of points so discovered. Several operators have been devised to accomplish this end. The best known ones are edge-detection, region growing and thresholding. Since we were not prepared to come up with a new technique we were constrained to adapt one of these methods. We had no particulary good ideas for a new kind of operator so we elected to explore the capabilities of existing mechanisms when applied to scenes of widely differing compositions. In the paragraphs that follow we briefly describe our experiences with these operators and why we rejected all but one as a basis for general segmentation.

Edge-detection

The operator which has commanded the major degree of attention is that of edge-detection. This method attempts to capture the discontinuities which occur at junctures of dissimilar portions of the picture. The many kinds of edge operators that have been expounded in the literature range in complexity from the simple Roberts cross operator to the sophisticated Hueckel operator (Roberts, 1963; Duda and Hart, 1973; Hueckel, 1973; Rosenfeld, 1969, 1969a, 1972, 1972a, 1973). Typically, edges are detected by the use of gradient operators which examine and compare intensity values within a small region of the picture. Others have been formulated to average intensity over neighborhoods bordering the point under consideration. These methods are less sensitive to noise but are also less sensitive to small regions. Rosenfeld has devised a scheme which uses averaging in varying size neighborhoods oriented in the vertical, horizontal, and diagonal directions (Rosenfeld and Thurston, 1971; Rosenfeld et al., 1972a; Hayes and Rosenfeld, 1972). Heavy edges produced by several sizes of neighborhoods can be thinned by suppression of non-maxima. Another method uses subtraction of averages over paired neighborhoods and multiplication of results in order to thin edges (Rosenfeld, 1970). MacLeod proposed an operator that calulates an edge weight for every point of the picture by multiplying the gray-level value of each point in a surounding neighborhood by the value of the corresponding point of a mask and then summing (MacLeod, 1972). The Hueckel operator fits a gray-level function derived from a circular area within the image to that member within a set of ideal edge lines whose Gaussian error of approximation to the input is minimal (Hueckel, 1973).

Segmentation

Past experience has shown the various edge-detection schemes to be of marginal success when applied to block world environments. How could we expect any one of them, then, to be successful when applied to natural scenes? It was felt that applying edge analysis to a variety of sensory sources of data might result in extracting edge elements from one parameter that were absent in another. In this way we hoped to avoid many of the missing lines that have been so troublesome in the past.

Of the many operators available we chose to make use of the Sobel operator (Duda and Hart, 1973), which for a 3x3 window,

$$\begin{matrix} a & b & c \\ d & e & f \\ g & h & i \end{matrix}$$

yields a gradient at point "e" which is defined as

$$|(a+2b+c)-(g+2h+i)| + |(a+2d+g)-(c+2f+i)|.$$

We found this edge detector to give reasonable results for our purposes. After deriving a gradient matrix corresponding to points in the picture, non-maxima were suppressed for a small neighborhood around maximum gradient values. A threshold was then applied in the usual way to obtain an edge picture. The threshold was manually selected to eliminate texture "noise" yet retain as many desireable edge elements as possible. Figure 3.1 shows the result when this sequence of steps is applied to the room scene of figure 2.4.b. Clearly, the missing lines present quite a problem. Hopefully, other parameters could fill in the missing portions.

Three sources of sensory data were directly available from the original digitization process. These were red, green, and blue density information as reflected through filters of the appropriate color. Using these sets of data we were able to extract two new parametes, hue and saturation, by means of a series of transformations borrowed from Tenenbaum (1973). These of course are not actual psychological measurements but their psychophysical analogs. The terms "hue" and "saturation" are used because they are more meaningful to most individuals. A sixth parameter, intensity, was obtained by averaging the three sensory inputs at each picture element (pixel). Three additional parameters, which correspond to television industry standards, "Y", "I", and "Q" (USC, 1973), were also transformed from the input data. This gave us a total of nine measures of the sensory data from which to extract edge information.

Extracting edges from the remaining parameters and pooling the results with the original intensity information gave the result shown in figure 3.2. Each parameter cast a "vote" at those pixel locations for which it could contribute an edge element. It was found that a vote of 3 was optimal in the final determination of the presence of a "real" edge at any picture point. Figure 3.2 shows some small evidence of noise but it can be easily handled. What is more important are the missing edges that still exist. The kind of heuristics that drive edge-followers in the block world could have some application in this scene. There are also possibilities of compensating for the missing

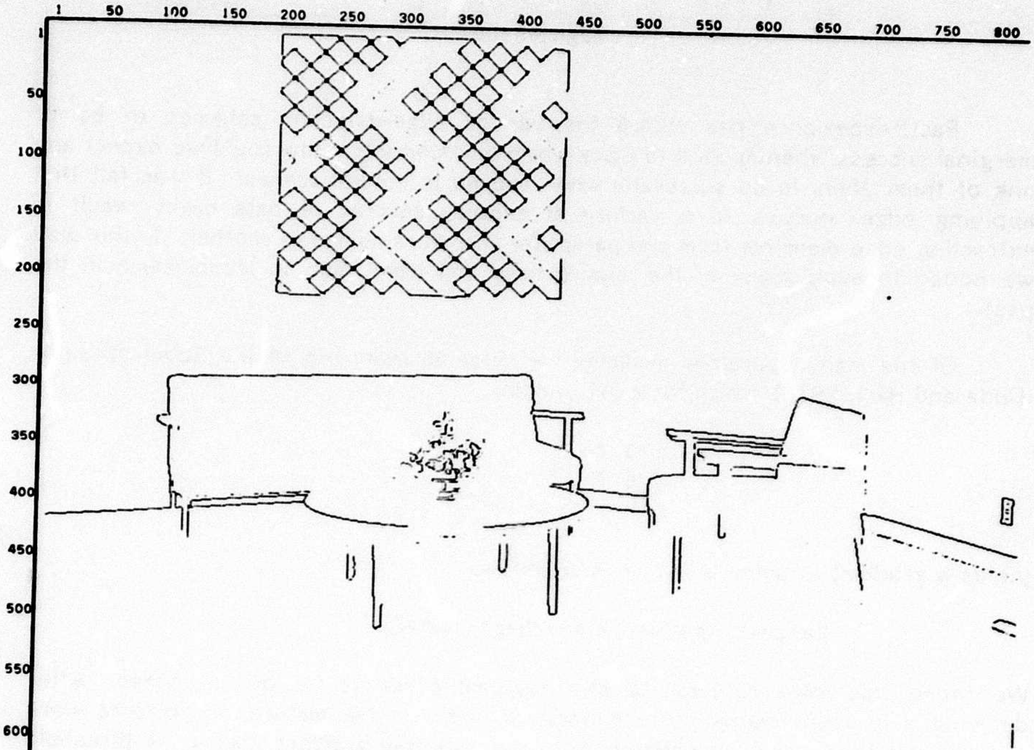


Figure 3.1. High-level intensity edges for room scene.

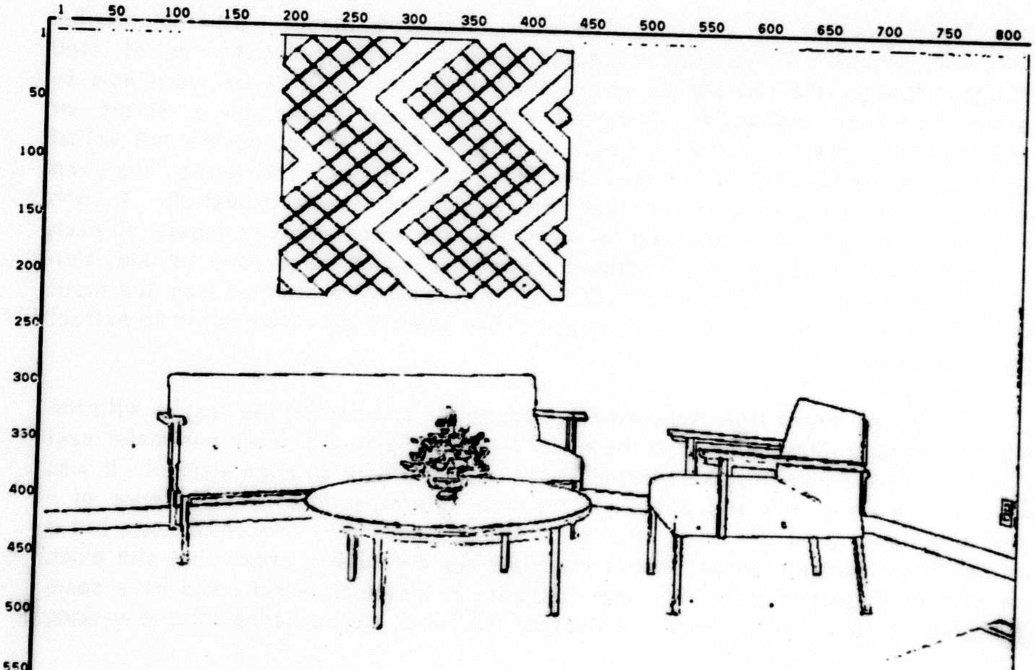


Figure 3.2. Combined intensity edges.

edge segments by line extension algorithms and region growing techniques to close the region.

The problems mentioned above become largely academic after applying the same edge-detection sequence to the bear (figure 3.3) and the skyline (figure 3.4). We did note that the edge information contained in these pictures could be used to make texture approximations by determining the number of edge elements per unit area, as proposed by Rosenfeld (Rosenfeld and Troy, 1970a). However, even if texture could be removed the resulting edge picture would present insurmountable problems to complete analysis.

Region-growing

Region-growing is a process which seeks to merge regions on the basis of similar attributes. Brice and Fennema (1970) were the first to make use of the scheme. They proposed a straightforward region-growing techniqe which merged regions on the basis of boundary strength determinations. Their program did fairly well with simple block structures. Barrow and Popplestone (1971) formed regions on the basis of uniformity (within a slight tolerance) of light intensity. They were able to recognize simple, single, real objects. Yakimovsky (1973) greatly elaborated this idea into a strategy which has proved to be successful in segmenting natural scenes. He used of a syntactic region-grower which made use of a number of region and boundary properties to effect a first level segmentation. To achieve his final partition he developed a semantic region-grower which utilized a number of features to improve upon the results of the low-level effort. He accomplished this by employing a probabalilistic model to determine the best merging of existing regions which would fit the world model.

Yakimovsky's work was rather complete in its research of the use of region and boundary attributes in determining merging criteria. We could not really expect to improve upon the operator in this respect. There were, though, several possible avenues of investigation that could have led to improvement. They were: the use of all pixels in a large scale picture for determining region properties; the growing of regions one pixel at a time; and performing the growing process in the context of edge elements extracted from the same scene. The scheme worked rather well for areas which were sharply defined by edges, but suffered the same deficiencies that Yakimovsky noted when edges were missing at a boundary, i.e., stopping short or growing into a neighboring region. Although investigation could have been pushed further by considering more discriminating similarity measures, more promising results in threshold applications caused us to abandon the technique at this point.

The General Segmentation Process

The following discussion treats, at some length, the construction of a general procedure for segmentation of natural scenes. We first consider some aspects of the thresholding operator which make it attractive when functioning in complex images. This is followed by a detailed explanation of the development of the basic algorithm.

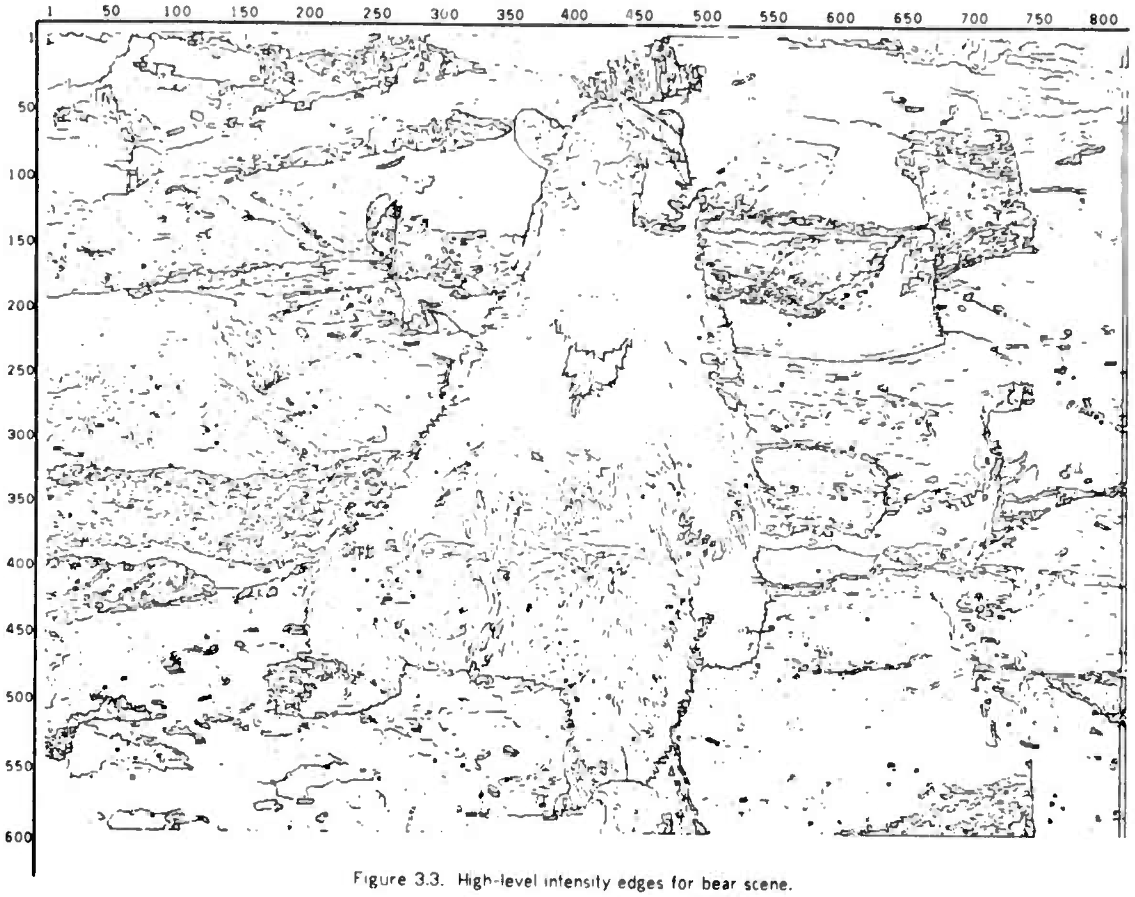


Figure 3.3. High-level intensity edges for bear scene.

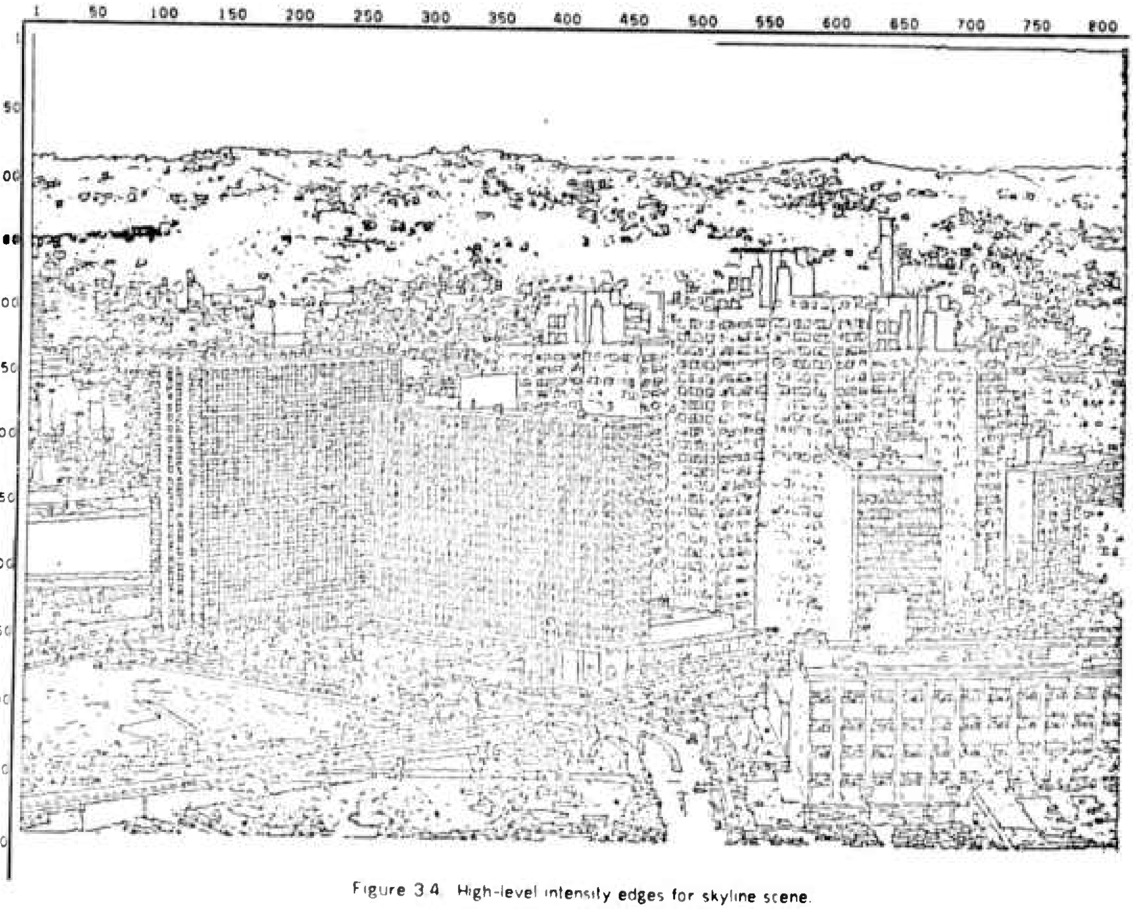


Figure 3.4. High-level intensity edges for skyline scene.

Segmentation by Thresholding

Investigation into the capabilities of the thresholding operator was conducted in parallel with the explorations mentioned in the previous section. Initial experiments convinced us that this operator could give more accurate results and would be appropriate for a wider range of scenes. Thresholding, or slicing, of picture intensity matrices has been used to good effect in various image processing applications for some time now. It has been utilized on raw picture data in images of simple scenes of high contrast with some success (Mendelsohn, 1968; Rosenfeld, 1969; Shirai, 1972). However, it has not found wide use in the larger scene understanding systems, probably because of scene complexity (natural outdoor scenes), or the presence of too many objects of similar gray level (block scenes). More recently, though, some researchers have begun to explore the capability of the operator in more difficult task domains. Tomita, Yachida, and Tsuji (1972), for example, have investigated some possibilities in thresholding complex picture properties to obtain shape and texture information. Another practical application has involved the efforts of the Jet Propulsion Laboratory in a study of simulated Mars environments for their robot vehicle (O'Handley et al; 1974). We have carried this exploitation even further in our construction of a segmentation scheme based upon the properties peculiar to the thresholding technique.

Let us examine briefly some of the aspects of a simple thresholding procedure and see what basis it provides for a general segmentation process. Initially, the problem will be discussed with respect to a simple digitized image of light intensities. What we hope to accomplish in a thresholding application is an isolation of portions of a picture by isolating, at any one time, only those points in a scene lying between certain gray levels. An immediate problem that comes up is how to choose the proper cutoff values. In the case of a binary level picture, such as might be presented by black type upon a white page, the solution is quite straightforward. Even in scenes which present several gray scale values, but only two predominant shades, there is no real difficulty. What options do we have, however, when we are faced with scenes of greater complexity? In figure 2.4.b, for instance, we would suspect that one should be able to isolate the various portions of the white walls quite easily. But what cutoff value should we give? What range of intensity do the brighter points of the picture have? How can a machine even know when there is something which can be partitioned out of the scene?

If we consider a histogram of the intensity values for the scene in figure 2.4.b we might see some indications of a sharply differentiated area. Figure 3.5 shows such a plot of frequency of occurrence versus intensity. As expected, a large number of points of high intensity ranging from approximately 190 to 240 in gray level are present. Examining a binary representation of these points after thresholding (figure 3.6), we see that we have quite accurately determined the points of the wall as well as the highlighted portion of the rug and the white background of the design on the wall. The point that we want to stress is that we have been able to accurately choose cutoff values for a thresholding operation by simply consulting a frequency function for a single parameter of the image in question. We have been able to avoid completely any arbitrary selection of critical values. What is more, the means of choice is dynamic in that it is based on a frequency histogram that varies with the

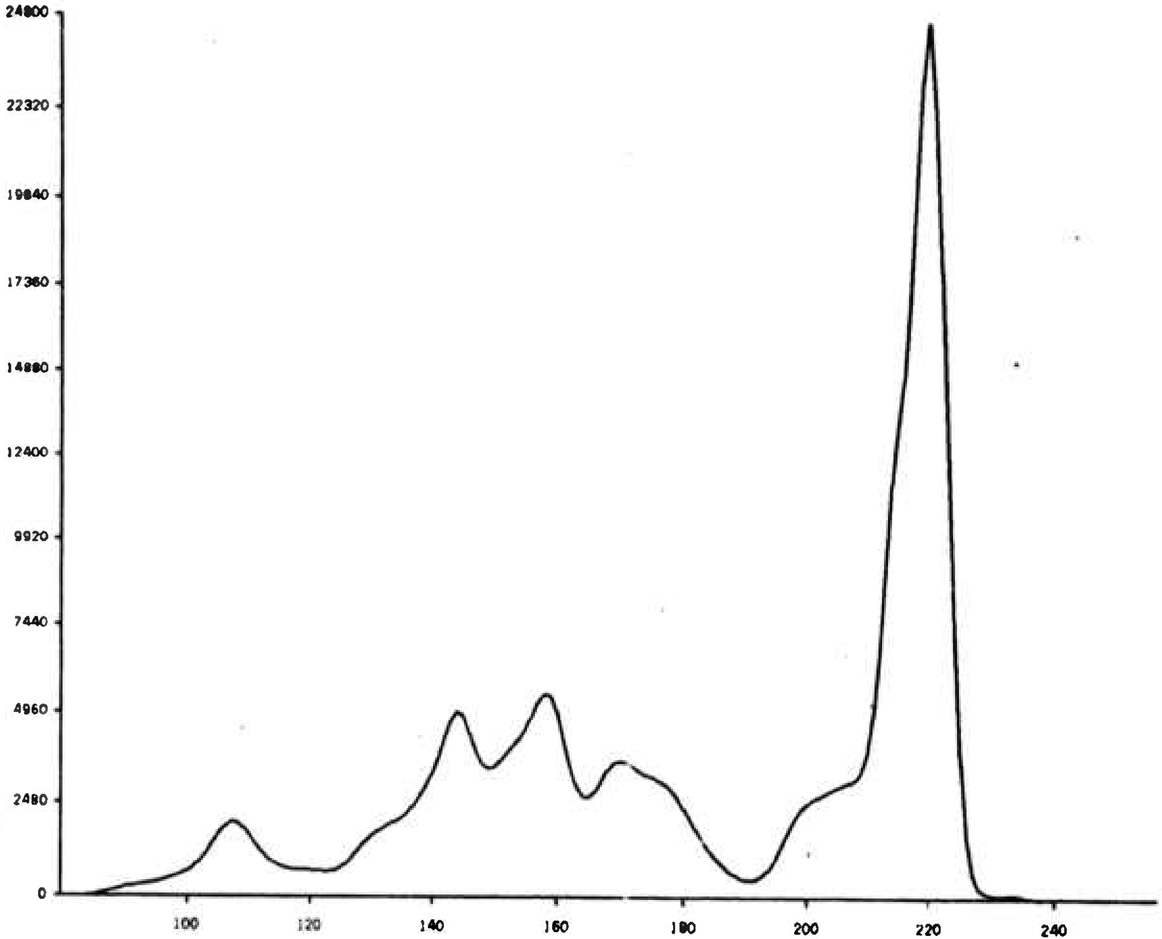


Figure 3.5. Intensity histogram for room scene.

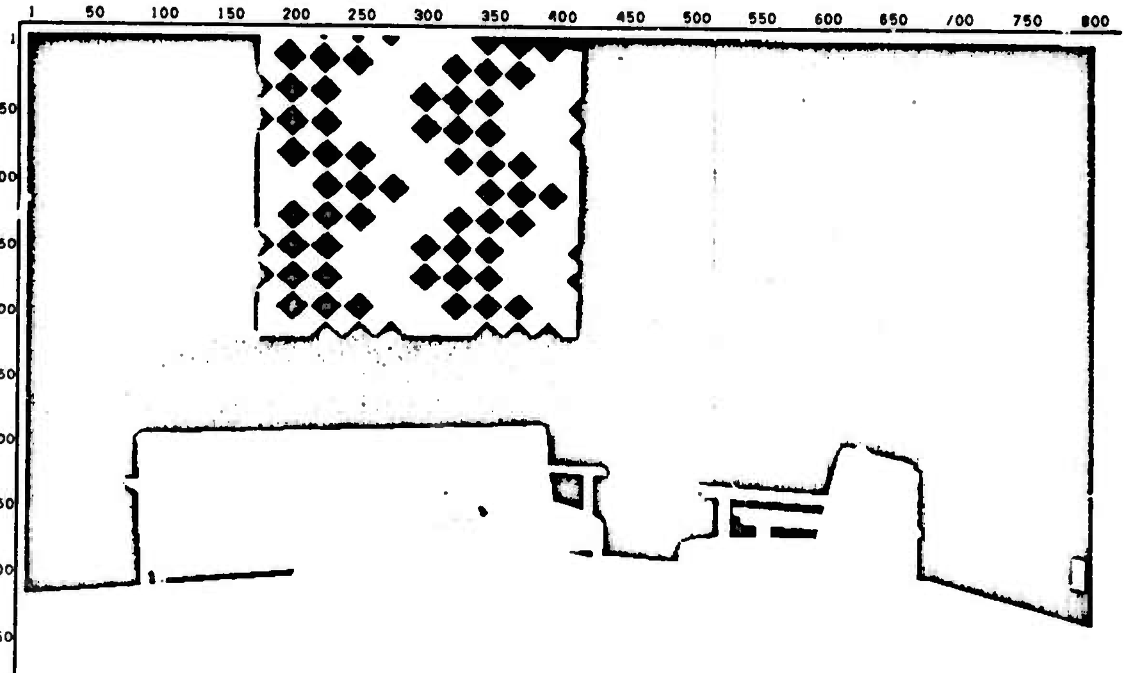


Figure 3.6. Room scene thresholded on intensity parameter (190 to 236).

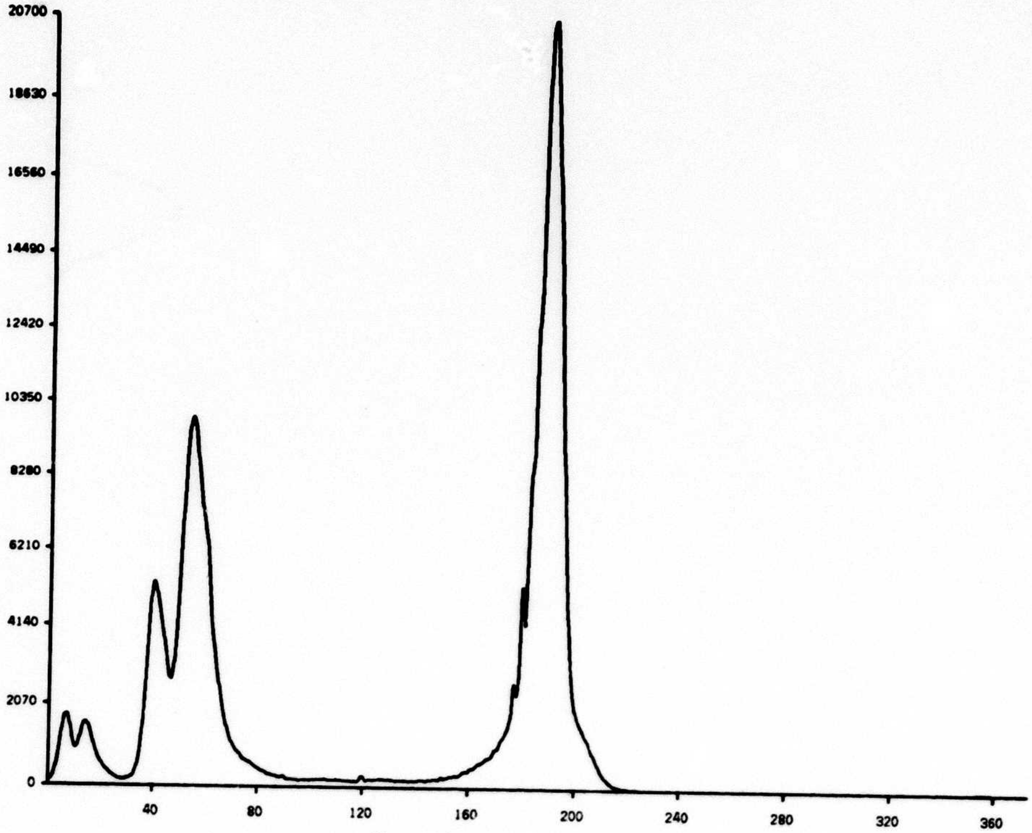


Figure 3.7. Hue histogram for room scene.

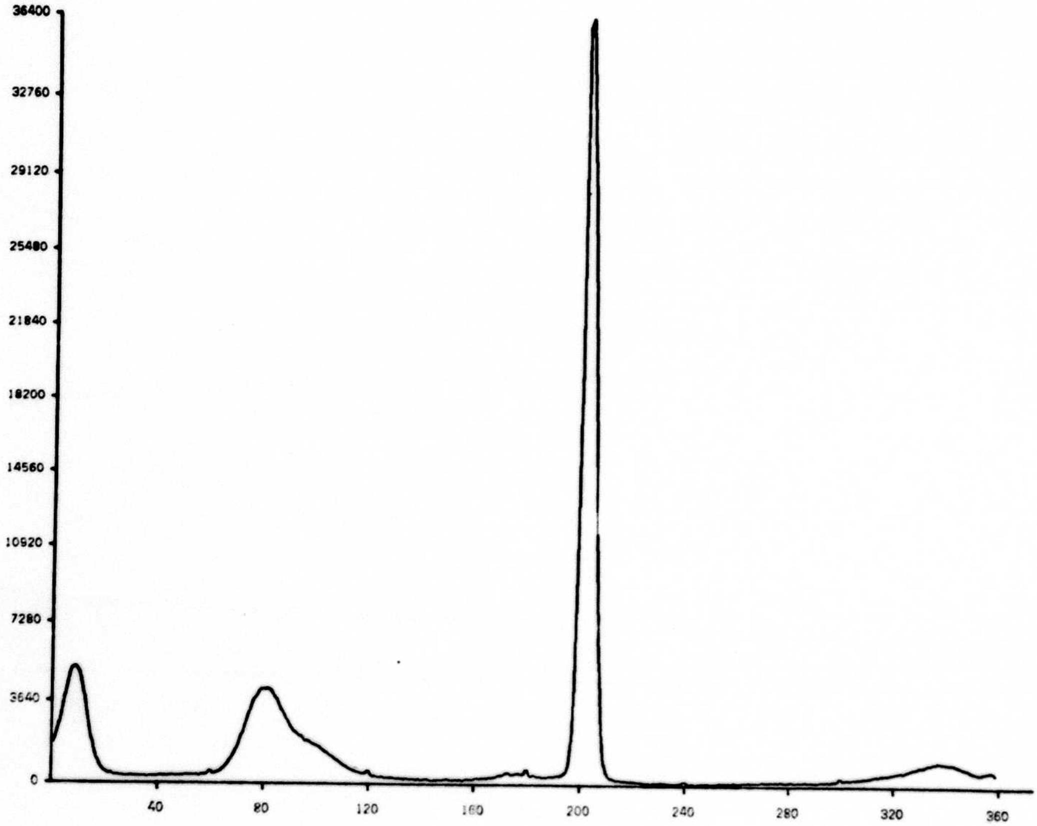


Figure 3.8.a. Hue histogram for house scene.

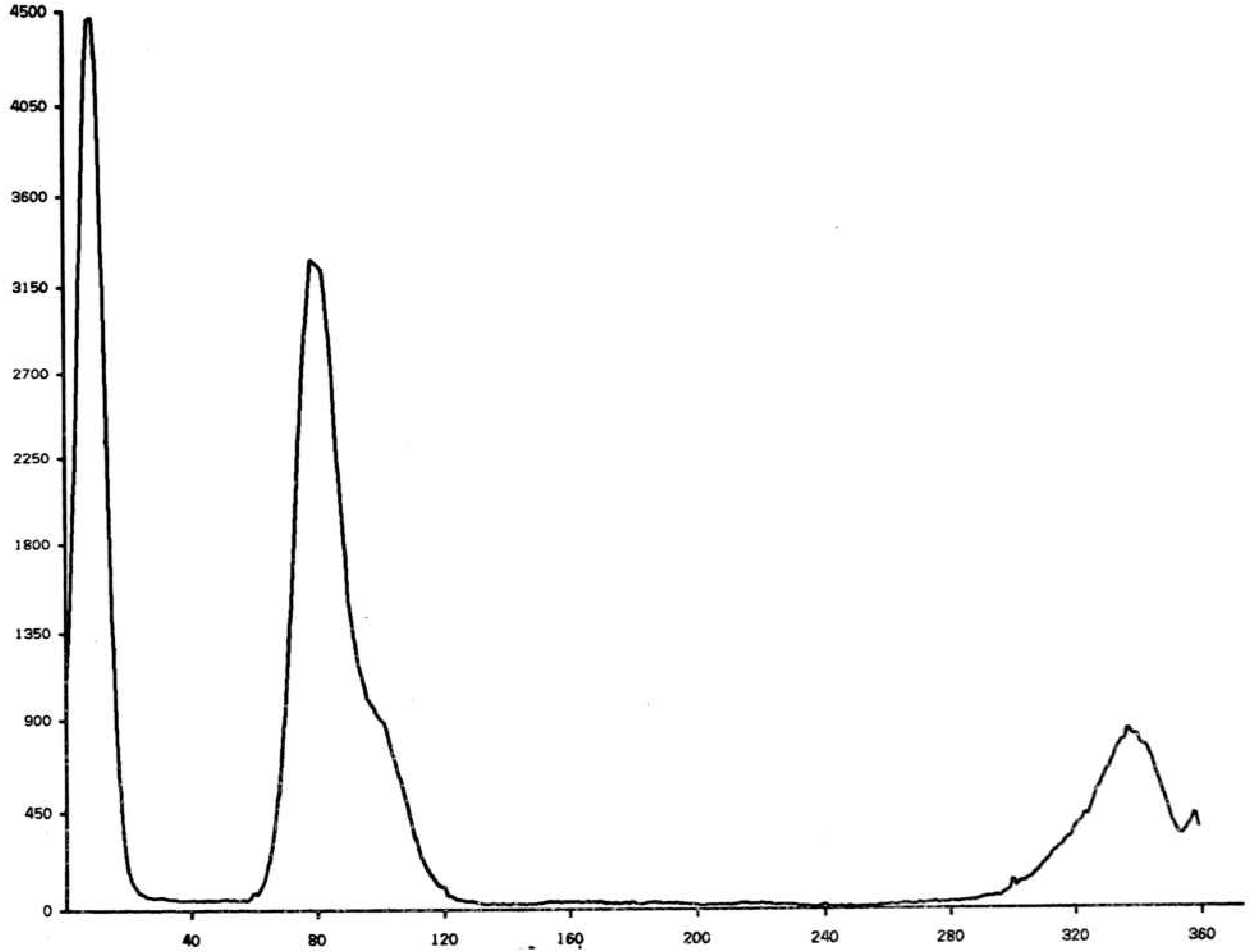


Figure 3.8.b. Hue histogram for house scene (sky removed).

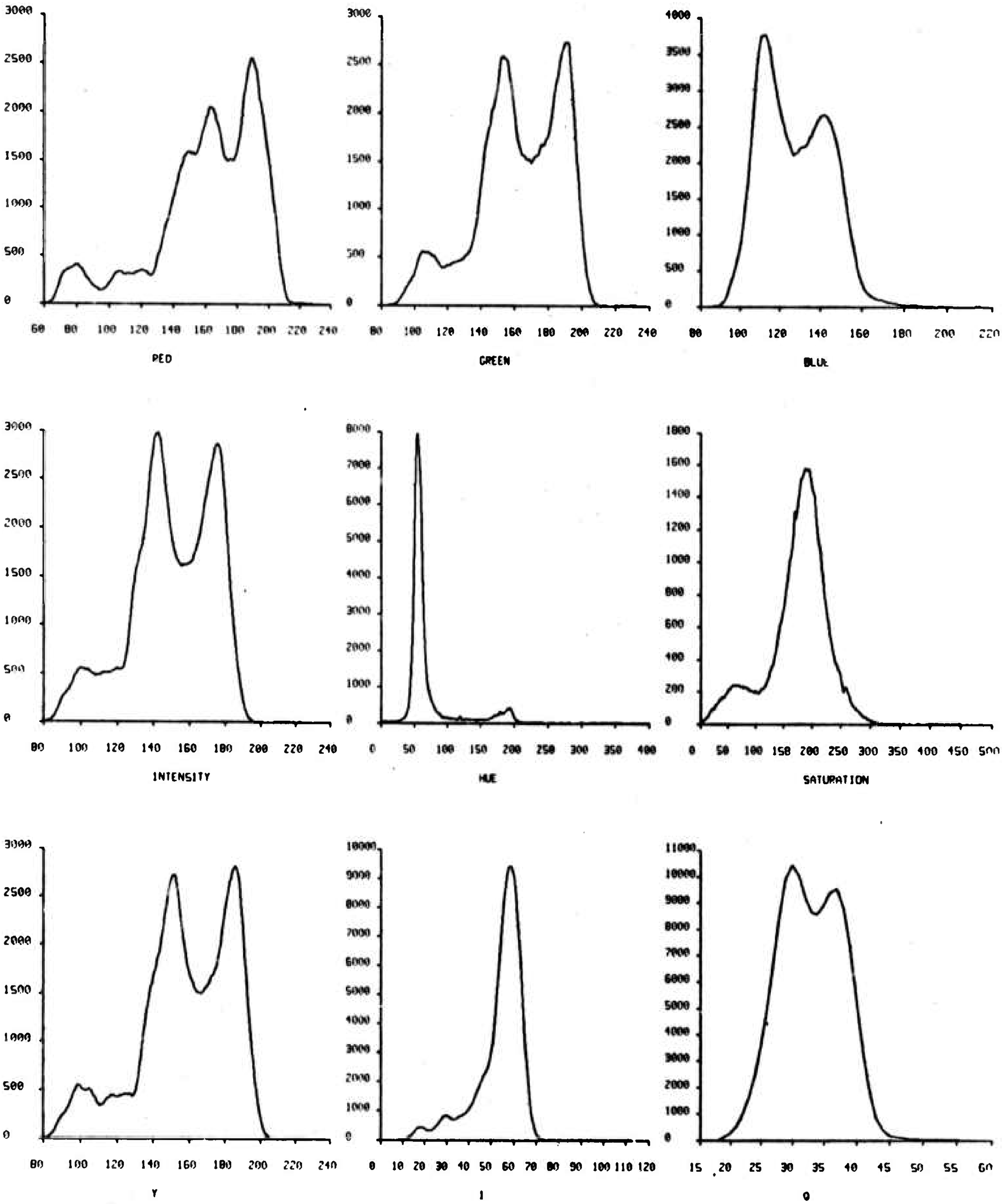


Figure 3.9. Nine property histograms for rug in room scene.

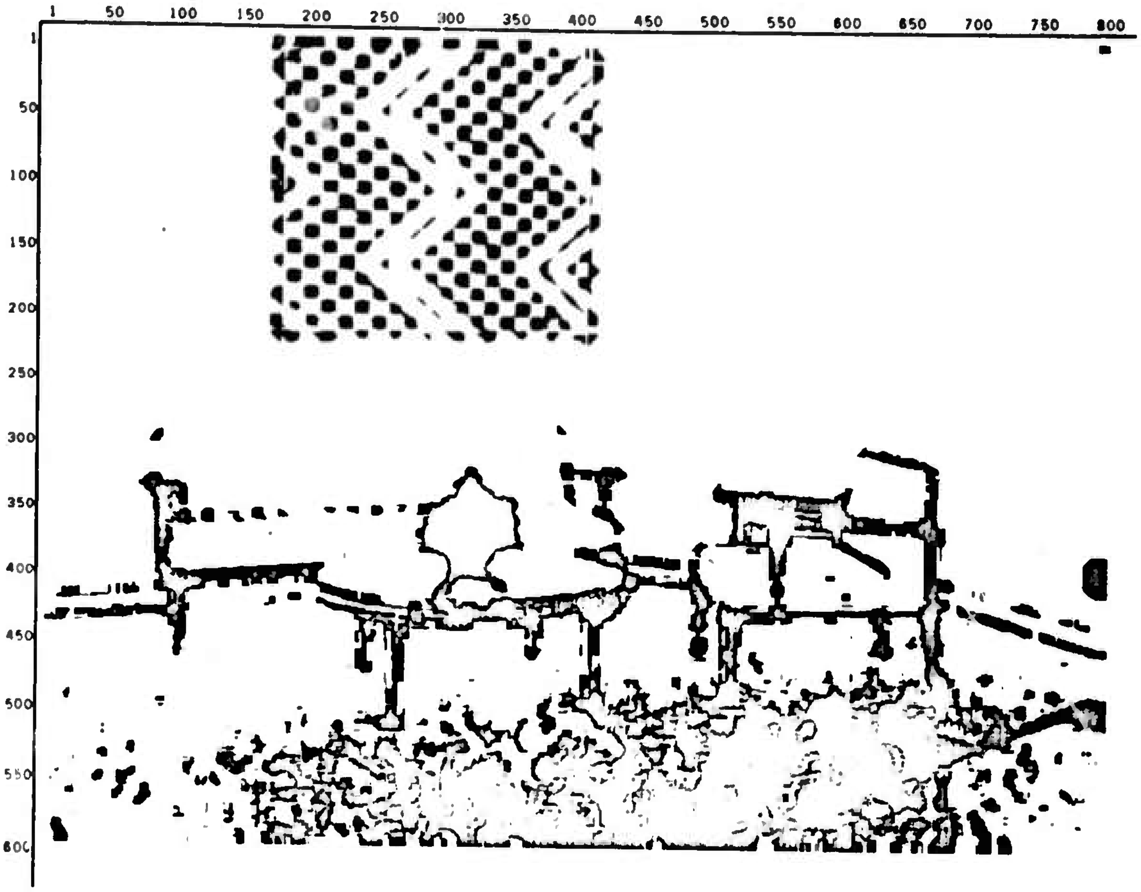


Figure 3.10. Moderate texture areas from intensiy parameters.

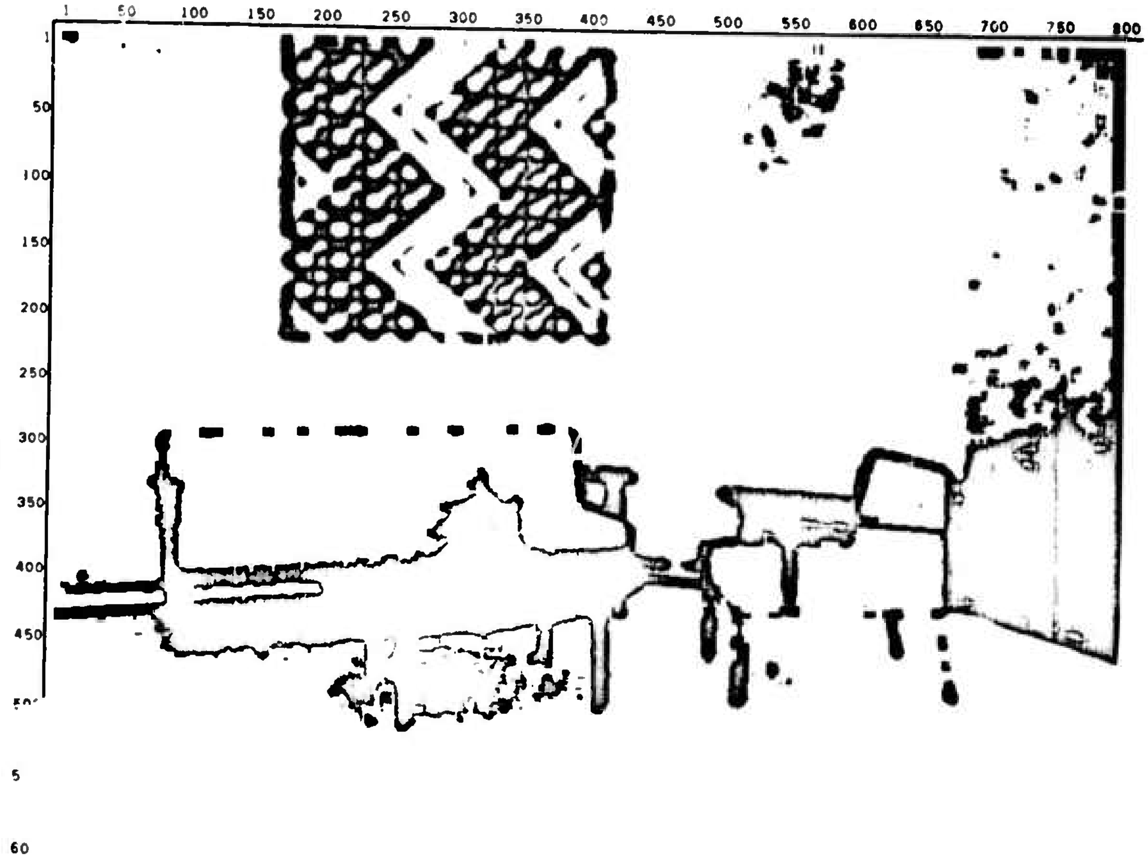


Figure 3.11. Moderate texture areas from hue parameters.

parameters of the picture. If another image of the scene had different lighting qualities, we would expect the frequency distribution changes in intensity to be reflected in a similar peak that has shifted slightly to the right or left.

When we examine the remainder of the histogram in figure 3.5 we see indications of additional regions of interest but, with the exception of a low range mode, no clear-cut delimitations of intensity values. Were we limited to this single source of image data we would be forced to conclude that a threshold-based segmentation operation is of limited utility due to overlap of values from regions of similar properties. We have, however, access to the same variety of sensory data that we used for edge detection (red, green, blue, intensity, hue, saturation, Y, I, Q). We have found that quite often one or more parameter will be sensitive to data that appears uniform in the other dimensions.

Thus, we have the option of examining nine histograms to determine the most sharply defined feature as measured by some parameter. Thresholding on limits provided by the minima bounding the best peak will furnish us clusters of points which are uniform for the given feature. We can then extract the region(s) so isolated and eliminate it(them) from further consideration. This elimination of extracted points can result in features which were formerly obscured becoming more distinct.

Our discussion, so far, promises that with sufficiently varied sources of sensory data, we might expect to do at least as well with thresholding as any edge detection scheme for fairly homogeneous types of scenes. After all, strong edges in any scene require adjoining regions which are highly contrasted along some parameter (e.g., light intensity, color, etc.). This would indicate that sharply defined features could be obtained for some number of histograms of the measurable parameters. But what can we expect for images which contain areas of moderate or strong texture? Will the process totally disintegrate as was the case for edge detection? An examination of figures 3.7 and 3.8, which are histograms of the hue for the room scene of figure 2.4.b and the house scene of figure 2.4.c respectively, reveals some aspects that indicate moderate texture might be treated fairly successfully. Modes shown in these figures argue that the distribution around maximum points seems to be approximately Gaussian in nature. Our intuitive appreciation of relatively large regions of homogeneous colors as reproduced in picture form seems to support this assumption. The white walls of figure 2.4.b show a large number of bright points with smaller areas that are in light shadow or highlighted. The effect is a gradual one which is consistent with a bell-shaped curve. For many types of moderate texture which occur in natural scenes, slightly flattened, and more widely distributed modes will appear. The roofs in the house scene (figure 2.4.c) which have values of hue between 300 and 360 in figure 3.8.b are examples of this. More remarkably, an object which is textured along several parameters may be uniform in some dimension. This can be clearly observed in figure 3.9 which shows histograms for the rug area of the room scene. Figures 3.10 and 3.11, which indicate moderately textured areas of the scene along the parameters of intensity and hue respectively, demonstrate the phenomenon even more graphically.

Areas of heavier texture present more dificult problems. In cases such as these no uniformity is evident in such properties as hue and intensity. We may not even have a uniformity of textural pattern. If one should consider the areas of shrubs and bushes in figure 2.4.c, at least two distinct shades of green could be discovered. One of them is approximately the same color as the grass and lies between 50 and 150 in figure 3.8.b The darker shade occurs between 150 and 240 on the same graph. If one or the other of these sets of values were thresholded on, regions with holes and discontinuities would be forthcoming. Although it might be possible to combine portions of textured areas on the basis of higher level knowledge, it is preferable to segment such areas in their entirety.

Such an achievement has, of course, been a long desired goal in computer vision and the subject of a good deal of research. The few investigators that have examined the problem for natural scenes have developed techniques for classifying some categories of texture but have been notably short of success in accurately delimiting areas of similar pattern. Perhaps the most effective work in this direction has been that of Bajcsy (1972). Analyzing spectral data, she was able to achieve good classification and a fairly coarse segmentation. The nature of the operation, measuring data over windowed areas of the image, precludes well-defined boundary separation.

Our own efforts in this respect have concentrated along two lines of approach. Primarily, we attempt to isolate highly textural areas by elimination of surrounding regions. If neighboring areas are homogeneous or of moderate texture it may be possible to segment them out by the thresholding procedure described above. What would be left then is the area of strong texture with sharply defined boundaries. The operation requires that we have an estimation of the heavily textured areas of the scene. This is necessary because it is quite possible that we could have locations within these areas which have attributes similar to some region which we are trying to isolate. Using a rough approximation of the strong texture region as a mask we can eliminate the unwanted point clusters from consideration. Another motivation for a preliminary estimation of highly textured regions is to provide a halting criterion -- we must be able to determine when further thresholding will result only in fragmentation of textured areas.

If there are not sufficient discriminatory parameters to allow segmentation by this method a more direct approach is used. A rough approximation of the strong texture area is extracted as before. For various reasons, some of which we have already mentioned and others which we will discuss later, it is desirable to refine this estimate. This can be accomplished by utilizing a variation of the basic thresholding scheme. The pixels in the approximating region can be used to mask out corresponding points of the nine parameter matrices. By averaging the sensory data for each parameter over a sufficiently large window we can blur the texture effect. The smoothing is often sufficient to permit differentiating properties of the heavily textured areas from those of surrounding regions. A refinement of boundaries can be achieved by taking histograms of the averaged parameter values and thresholding closed regions as we did for the direct values. An instance of this technique will be seen later when we describe some of the actual results.

The Basic Algorithm

In this section we will describe in some detail a general procedure which meets with a fair degree of success in segmenting six dissimilar natural scenes. Rather than attempt to put forth the final complex version of the process at this point, we will begin by presenting the comparatively simple procedure with which we started. As we progress through the processing of the pictures we shall explain modifications and additions to the basic procedure that we found necessary. In this way we hope to offer an orderly development of the process which will be more understandable to the reader.

Since our purpose was to investigate several aspects of a general vision system in terms of a research model, we rely on human interaction to provide several functions. It was felt to be more productive to use available time to investigate the feasibility of a widely applicable segmentation algorithm rather than generate a fully automated process of limited scope. For this reason the human experimenter furnishes most of the control structure. He does this by applying various available operators in the sequence required by a flow chart. User intervention is also demanded in some of the subroutines which are more time consuming to program and/or utilize. As we progress through our presentation we shall more closely define the investigator's active role.

The algorithm illustrated in figure 3.12 appears to be more cumbersome and complex than it actually is. This is due, in part, to the difficutly in depicting a recursive procedure in flow chart form. The main concept is basically simple, viz., a continuous application of a thresholding operator to a picture until all areas possessing uniformity along some dimension are isolated. As the process advances, closed regions are generated from two causes. They may be the result of the thresholding operation; in such a case they are further refined by thresholding if additional histograms indicate the need of it. The results of this procedure are completely processed segments. Closed regions may also come about by isolation of portions of the image as processed segments are removed. When this occurs the basic procedure is recursively applied to any subimages thus separated. This structure is similar to the recursive descent scheme proposed by Tomita, Yachida, and Tsuji (1972) for partitioning simple images. There are other similarities between our system and the one they describe -- the use of multiple sources of data and the use of a threshold operation based on histograms.

Let us now step through the algorithm as we apply it to one of our pictures. The room scene shown in figure 2.4.b was chosen for an initial application because it is generally homogeneous, rich in color, and reasonably complex in structure. Such properties provide an adequate test of the mechanism without taxing it to the point where the functioning of the basic algorithm becomes overwhelmed.

The first step in the procedure involves the derivation of the sensory parameters. This is accomplished in the manner described in the subsection describing edge analysis. Step 2 requires analysis of data to approximately locate areas of strong texture and to extract the overall textural properties which guide further evaluation. Figure 3.13 gives more detail on the operations required. As we mentioned earlier, use is being made of a primitive type of texture indicator (Rosenfeld

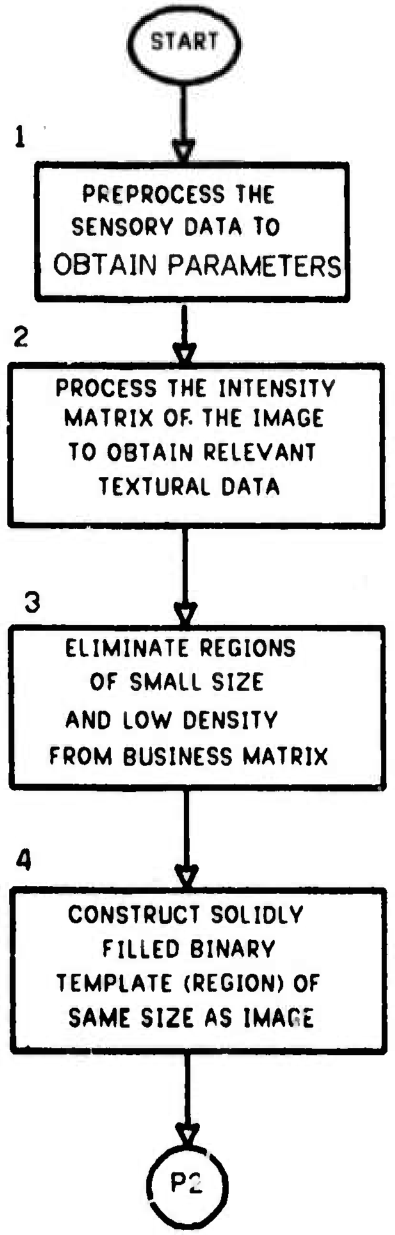

START
1
PREPROCESS THE SENSORY DATA TO OBTAIN PARAMETERS
2
PROCESS THE INTENSITY MATRIX OF THE IMAGE TO OBTAIN RELEVANT TEXTURAL DATA
3
ELIMINATE REGIONS OF SMALL SIZE AND LOW DENSITY FROM BUSINESS MATRIX
4
CONSTRUCT SOLIDLY FILLED BINARY TEMPLATE (REGION) OF SAME SIZE AS IMAGE
P2

Figure 3.12. The segmentation algorithm.

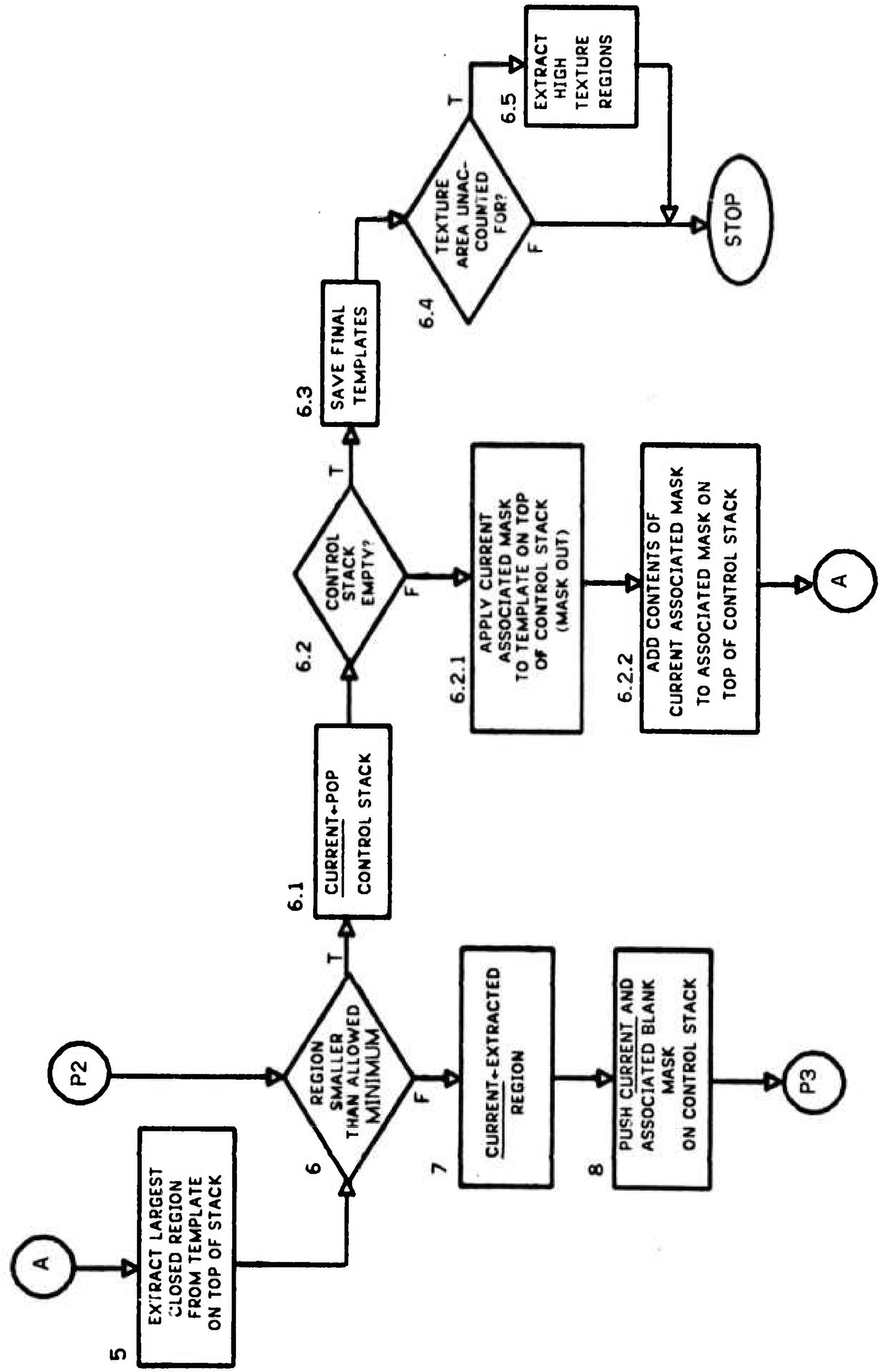

A
5
EXTRACT LARGEST CLOSED REGION FROM TEMPLATE ON TOP OF STACK
P2
6
REGION SMALLER THAN ALLOWED MINIMUM
T
F
7
CURRENT←EXTRACTED REGION
8
PUSH CURRENT AND ASSOCIATED BLANK MASK ON CONTROL STACK
P3
6.1
CURRENT←POP CONTROL STACK
6.2
CONTROL STACK EMPTY?
T
F
6.2.1
APPLY CURRENT ASSOCIATED MASK TO TEMPLATE ON TOP OF CONTROL STACK (MASK OUT)
6.2.2
ADD CONTENTS OF CURRENT ASSOCIATED MASK TO ASSOCIATED MASK ON TOP OF CONTROL STACK
A
6.3
SAVE FINAL TEMPLATES
6.4
TEXTURE AREA UNACCOUNTED FOR?
T
F
6.5
EXTRACT HIGH TEXTURE REGIONS
STOP

Figure 3.12.(continued). The segmentation algorithm.

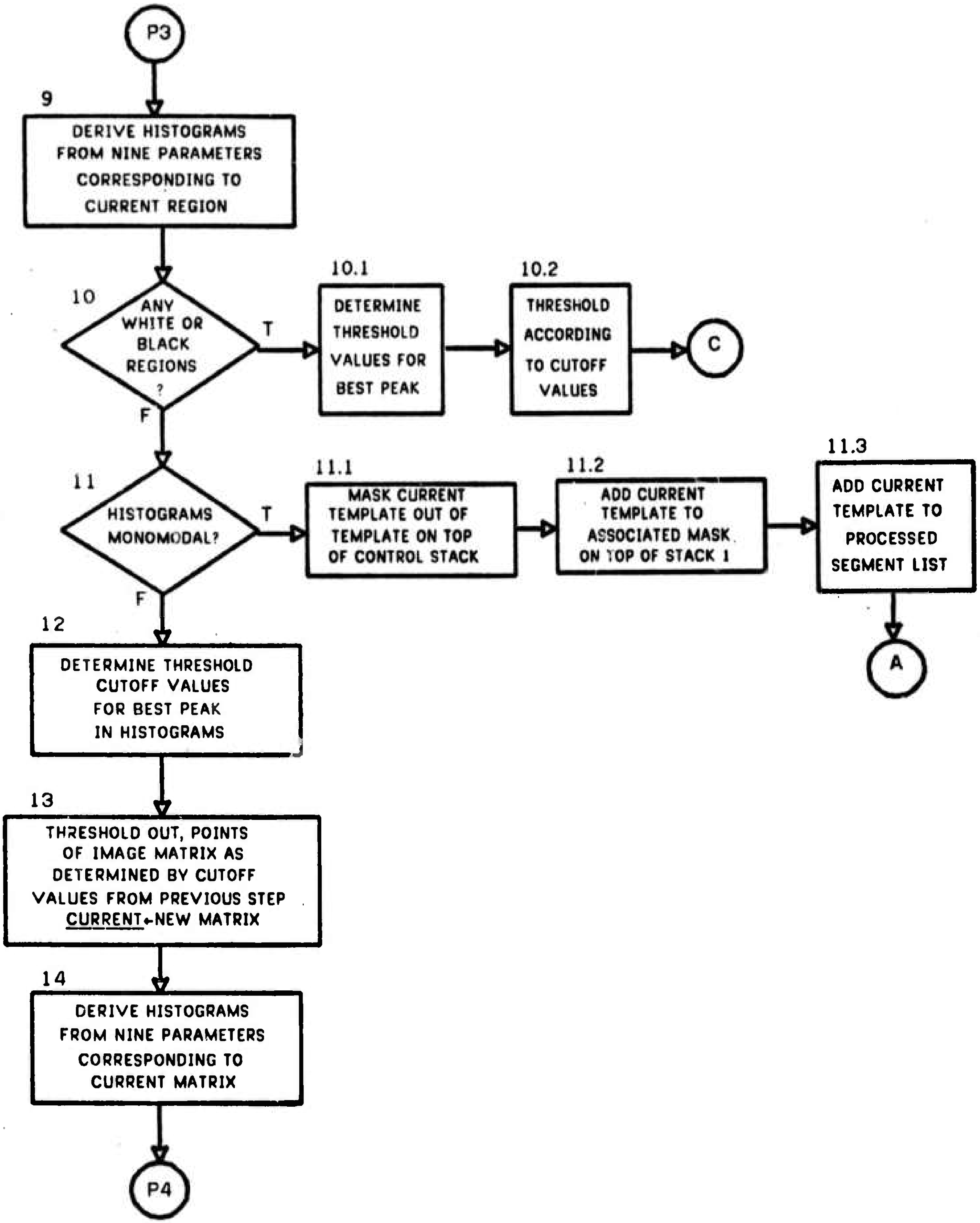

P3
9
DERIVE HISTOGRAMS FROM NINE PARAMETERS CORRESPONDING TO CURRENT REGION
10
ANY WHITE OR BLACK REGIONS ?
T
F
10.1
DETERMINE THRESHOLD VALUES FOR BEST PEAK
10.2
THRESHOLD ACCORDING TO CUTOFF VALUES
C
11
HISTOGRAMS MONOMODAL?
T
F
11.1
MASK CURRENT TEMPLATE OUT OF TEMPLATE ON TOP OF CONTROL STACK
11.2
ADD CURRENT TEMPLATE TO ASSOCIATED MASK ON TOP OF STACK 1
11.3
ADD CURRENT TEMPLATE TO PROCESSED SEGMENT LIST
A
12
DETERMINE THRESHOLD CUTOFF VALUES FOR BEST PEAK IN HISTOGRAMS
13
THRESHOLD OUT, POINTS OF IMAGE MATRIX AS DETERMINED BY CUTOFF VALUES FROM PREVIOUS STEP CURRENT←NEW MATRIX
14
DERIVE HISTOGRAMS FROM NINE PARAMETERS CORRESPONDING TO CURRENT MATRIX
P4

Figure 3.12.(continued). The segmentation algorithm.

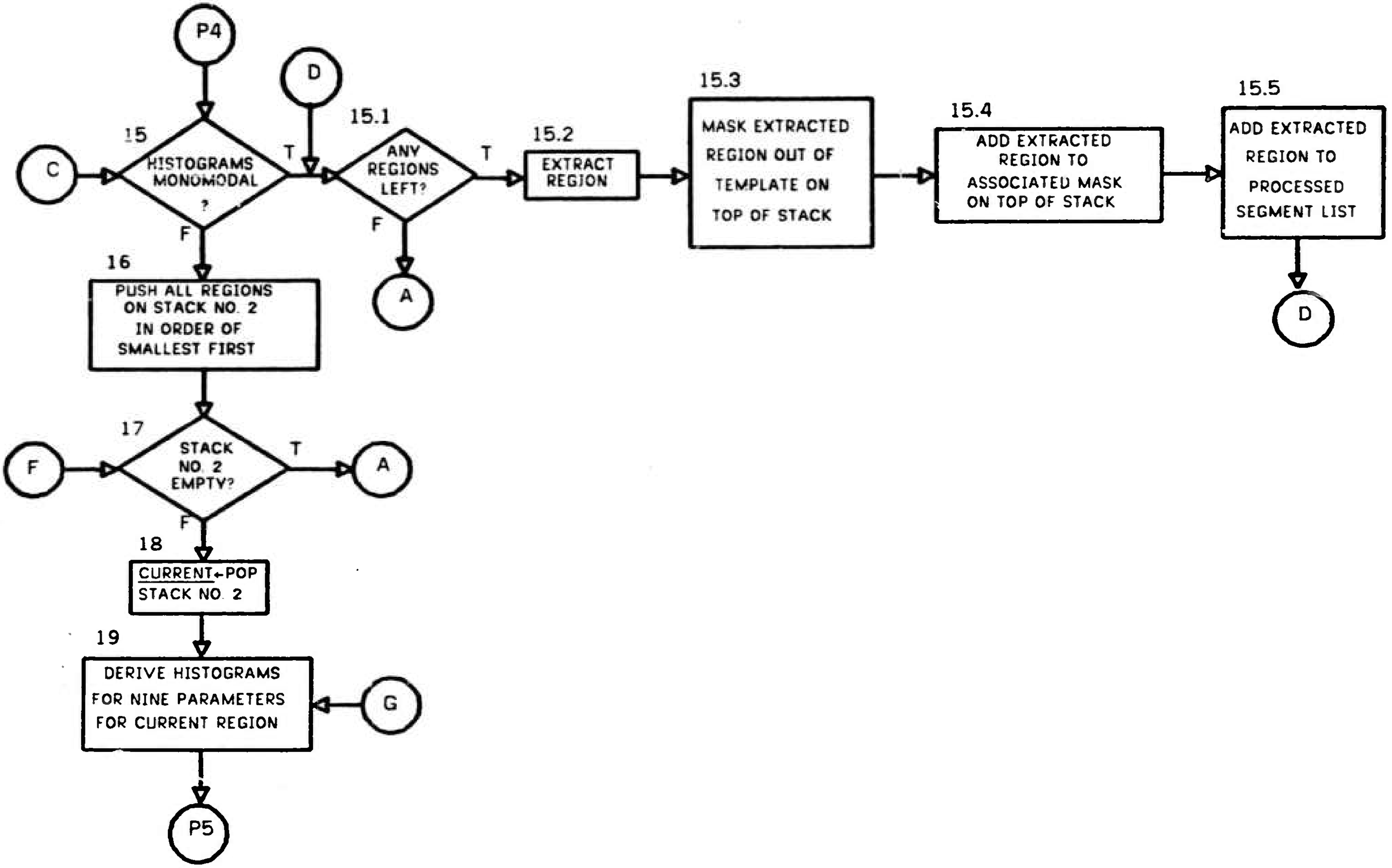

P4
D
15
C
HISTOGRAMS MONOMODAL ?
T
F
15.1
ANY REGIONS LEFT?
T
F
A
15.2
EXTRACT REGION
15.3
MASK EXTRACTED REGION OUT OF TEMPLATE ON TOP OF STACK
15.4
ADD EXTRACTED REGION TO ASSOCIATED MASK ON TOP OF STACK
15.5
ADD EXTRACTED REGION TO PROCESSED SEGMENT LIST
D
16
PUSH ALL REGIONS ON STACK NO. 2 IN ORDER OF SMALLEST FIRST
17
F
STACK NO. 2 EMPTY?
T
A
F
18
CURRENT←POP STACK NO. 2
19
DERIVE HISTOGRAMS FOR NINE PARAMETERS FOR CURRENT REGION
G
P5

Figure 3.12.(continued). The segmentation algorithm.

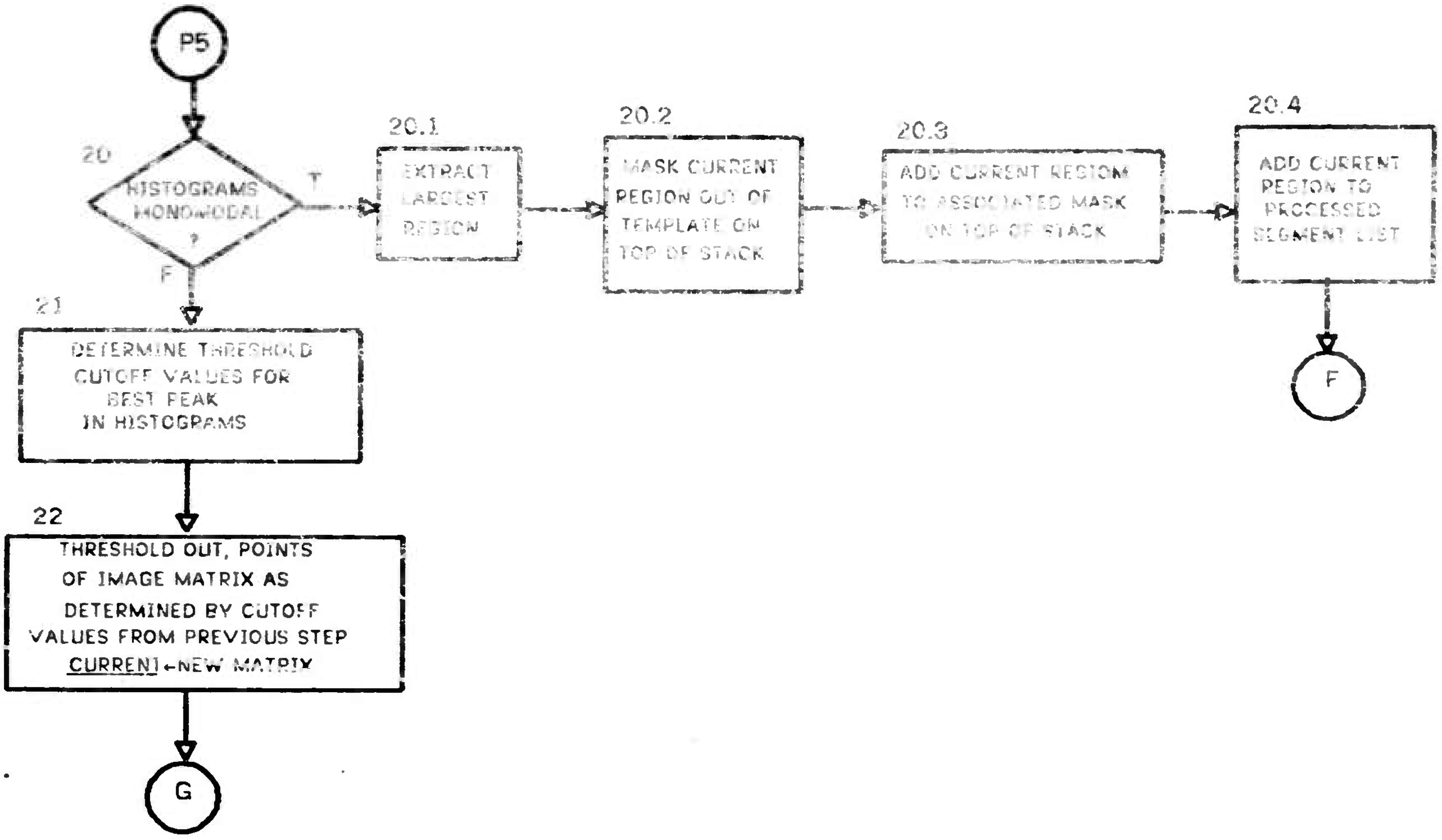

P5
20
HISTOGRAMS MONOMODAL ?
T
F
20.1
EXTRACT LARGEST REGION
20.2
MASK CURRENT REGION OUT OF TEMPLATE ON TOP OF STACK
20.3
ADD CURRENT REGION TO ASSOCIATED MASK ON TOP OF STACK
20.4
ADD CURRENT REGION TO PROCESSED SEGMENT LIST
F
21
DETERMINE THRESHOLD CUTOFF VALUES FOR BEST PEAK IN HISTOGRAMS
22
THRESHOLD OUT, POINTS OF IMAGE MATRIX AS DETERMINED BY CUTOFF VALUES FROM PREVIOUS STEP CURRENT←NEW MATRIX
G

Figure 3.12.(continued). The segmentation algorithm.

and Troy, 1970a). Strong texture is often characterized by sharp and rapid changes in sensory data values. We noticed this phenomenon earlier in the section on edge analysis (figure 3.3). Utilzing this feature we can estimate texture by an evaluation of the number of edge points per unit area above a specified threshold. To this end we first utilize the Sobel operator (Duda and Hart, 1973) on the intensity parameter to derive a matrix of gradient values. The next step necessitates a determination of a suitable threshold cutoff value. To avoid a completely subjective choice we arrived at the ad hoc, but sucessful, method of making a selection based on the statistical characteristics of the frequency distribution of the gradiant matrix. A value taken at one-half the standard deviation above the mean proved to be satisfactory. We processed the intensity data, alone, for two reasons. In the first place any additional operations are very expensive for images of the size we are analyzing and we already require a large time expenditure. Secondly, we believe that strong textural patterns in any one parameter will be reflected in the intensity data.

Thresholding the gradient matrix at the selected value gives us a binary image which is very similar to the type obtained for edge extraction (figure 3.3). What we want to locate in this image are those regions, like the vase, which have a relatively high number of edge indicators in a compact area. To discover such places an operator is employed which counts the number of edge points in some window area. A new matrix is contructed which contains, at each entry, the count obtained from the window centered on that point. We call the result a "business" matrix. Only those pixels of the derived matrix which possess a given count (busy factor) need be considered. Utilizing a 9x9 window we have found it useful to retain those points possessing a busy factor greater than or equal to 25. This figure was arrived at by arbitrarily deciding that any window with more than two and one-half lines running through it was indicative of the occurrence of a textural pattern. Note that it requires an aggregate of such points to define a textured region of significant size.

Application of the busy and threshold operators results in the binary image shown in figure 3.14. The highly textured region of the vase and flowers is adequately delimited. There is also evidence of at least two points of weakness in the process. First, there are a lot of long narrow regions which have two or more edges in spatially close proximity. These areas give the same evidence of texture with the exception that they are not repetitive nor of long spatial extent. Ways of eliminating such regions will be discussed shortly.

The second phenomenon requiring explanation is the large number of small squares in the upper center portion of the figure. These are the results of the juxtaposed squares of different colors present in the design of figure 2.4.b. We observe that this is symptomatic of one of the fundamental properties inherent in the concept of texture, viz., textural patterns can be of any size. One way of handling the difficulty is increasing the window size, or equivalently, reducing the image. Figure 3.15 shows the effect of reducing the picture by a factor of two and applying the required operators. As one might expect additional indicators of texture arise due to the compacting of detail. Specifically, the arms of the chair have now been brought into sufficient proximity to define one large area of texture where formerly there were two. Rather than pursue this avenue and cope with new difficulties that might surface, we elect to remain with a single window dimension. The textural pattern

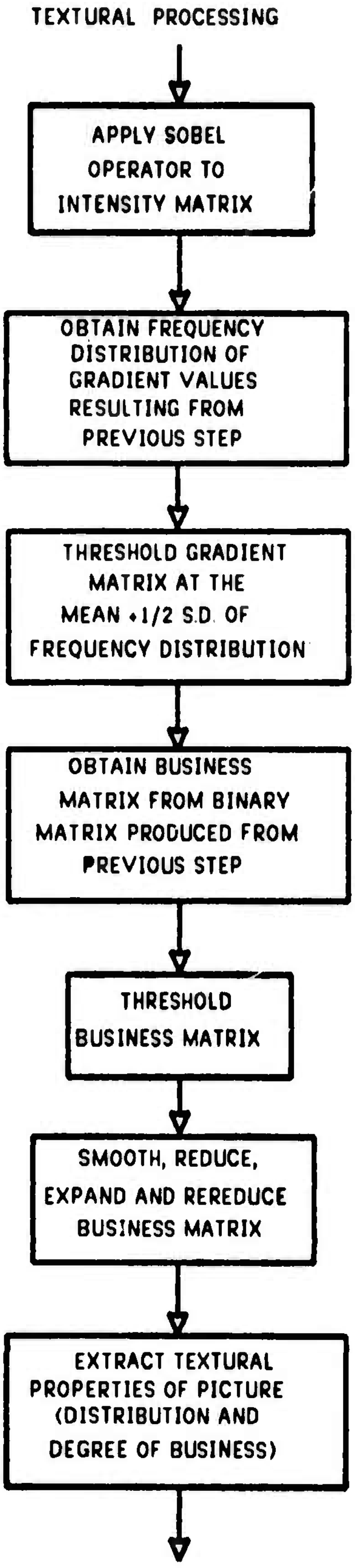

TEXTURAL PROCESSING
APPLY SOBEL OPERATOR TO INTENSITY MATRIX
OBTAIN FREQUENCY DISTRIBUTION OF GRADIENT VALUES RESULTING FROM PREVIOUS STEP
THRESHOLD GRADIENT MATRIX AT THE MEAN +1/2 S.D. OF FREQUENCY DISTRIBUTION
OBTAIN BUSINESS MATRIX FROM BINARY MATRIX PRODUCED FROM PREVIOUS STEP
THRESHOLD BUSINESS MATRIX
SMOOTH, REDUCE, EXPAND AND REREDUCE BUSINESS MATRIX
EXTRACT TEXTURAL PROPERTIES OF PICTURE (DISTRIBUTION AND DEGREE OF BUSINESS)

Figure 3.13. Algorithm for texture preprocessing.

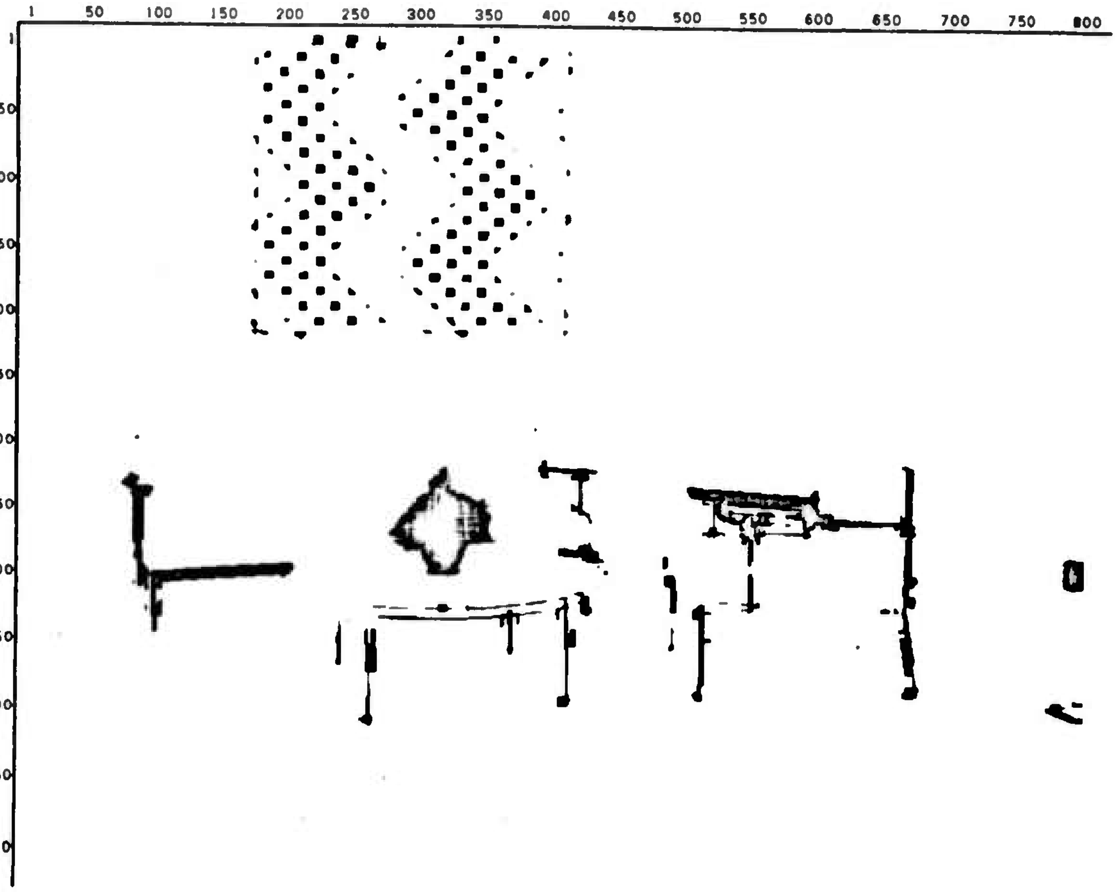


Figure 3.14. Busy mask for room scene.

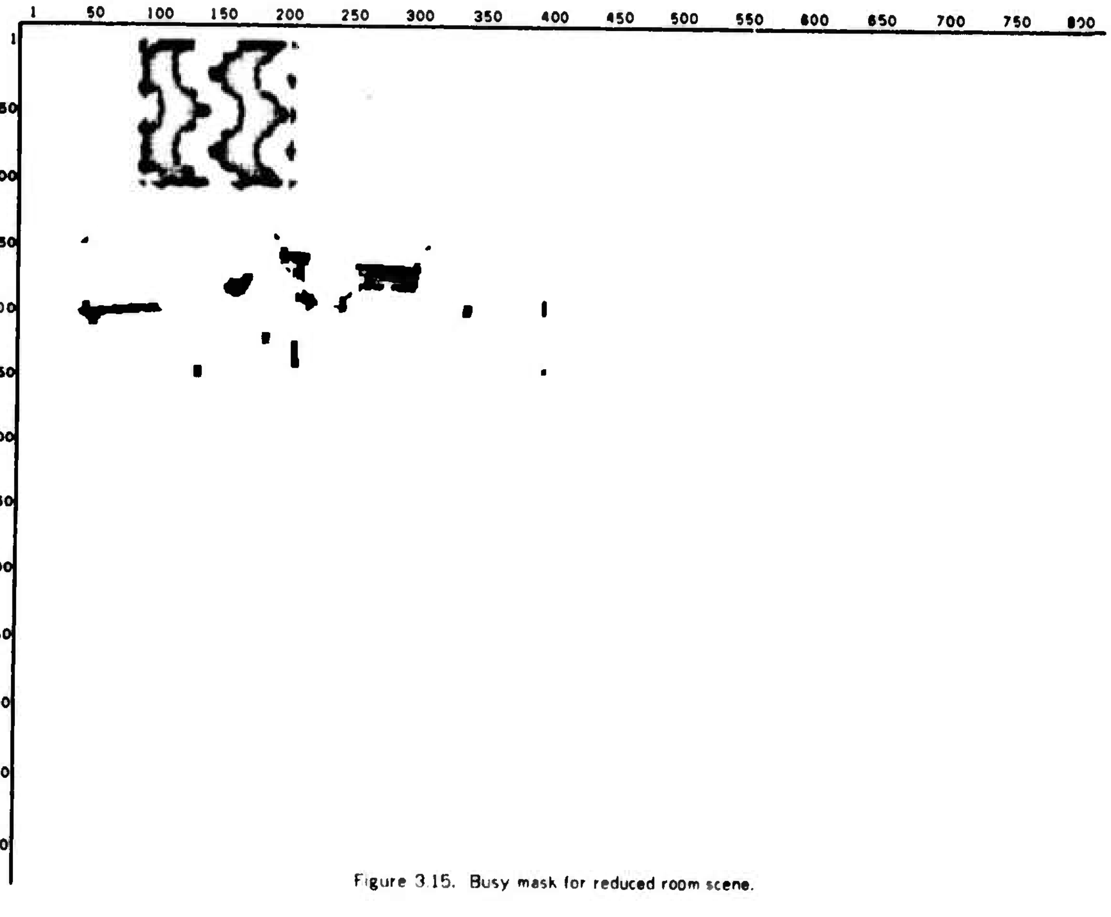


Figure 3.15. Busy mask for reduced room scene.

presented by the design on the wall proves to be of sufficient size and regularity to be handled by the general thresholding technique.

We have yet to compensate for the various "erroneous" indicators of texture. They can be eliminated by judicious application of several operators and heuristics. A smoothing or filling operator is initially used to eliminate isolated points and fill in small holes. A reduction operator (Ejiri et al., 1973) then serves to contract the picture, thereby eliminating some of the smaller and/or narrower regions (figure 3.16). The matrix is then expanded back to normal size and again enlarged uniformly in all directions to fill minor holes. A second contraction brings the picture back to normal size (figure 3.17). If image textural properties, as described in the next paragraph, indicate that it might be useful, the matrix is then processed to obtain the best estimate of the "busy" areas. This step is only worthwhile if the image is basically homogeneous in nature. The resultant matrix of the bear, as shown in figure 3.18, would not yield to this process.

Several somewhat crude textural attributes can be extracted which help to direct later analysis and which also play a role in identification of scene type in the recognition module. The measures indicate degree and distribution of heavy texture. We calculate the percentage of the total number of pixels comprised by busy points, to derive some idea of picture composition. A locus estimate is formed by computing the relative percentage of business in each quadrant of the image. Our last determination, which measures amount of dispersion, is based on the chi-square formula as given below.

$$\text{(1)} \qquad \text{chi-square} = \sum (E-x)^2/E,$$

where E is the estimation of the mean of the distribution of busy points for some given window size in the business matrix and x is the observed value. The image has been divided into 60x80 squares which gives a total of one-hundred squares, ten in each direction. Uniformly distributed points should give a low value for this measure. The chi-square value varies with total numbers of busy points so the final term is normalized by dividing by that number. The result is a somewhat crude but effective indicator for our range of pictures. Figures obtained for the picture under analysis are:

strong texture points (fraction of total) = .0018
fraction in upper left quadrant = .0000
fraction in upper right quadrant = .0000
fraction in lower left quadrant = .0018
fraction in lower right quadrant = .0000
modified chi-square = 16.75

Our modified chi-square result is relatively high, which is what we would expect from the small number of highly concentrated strong texture points in the image.

On the basis of observed textural parameters, step 3.1 of the basic procedure is performed next. We extract each region of the business matrix by the use of a connected point algorithm as described in Rosenfield (1969). The use of this

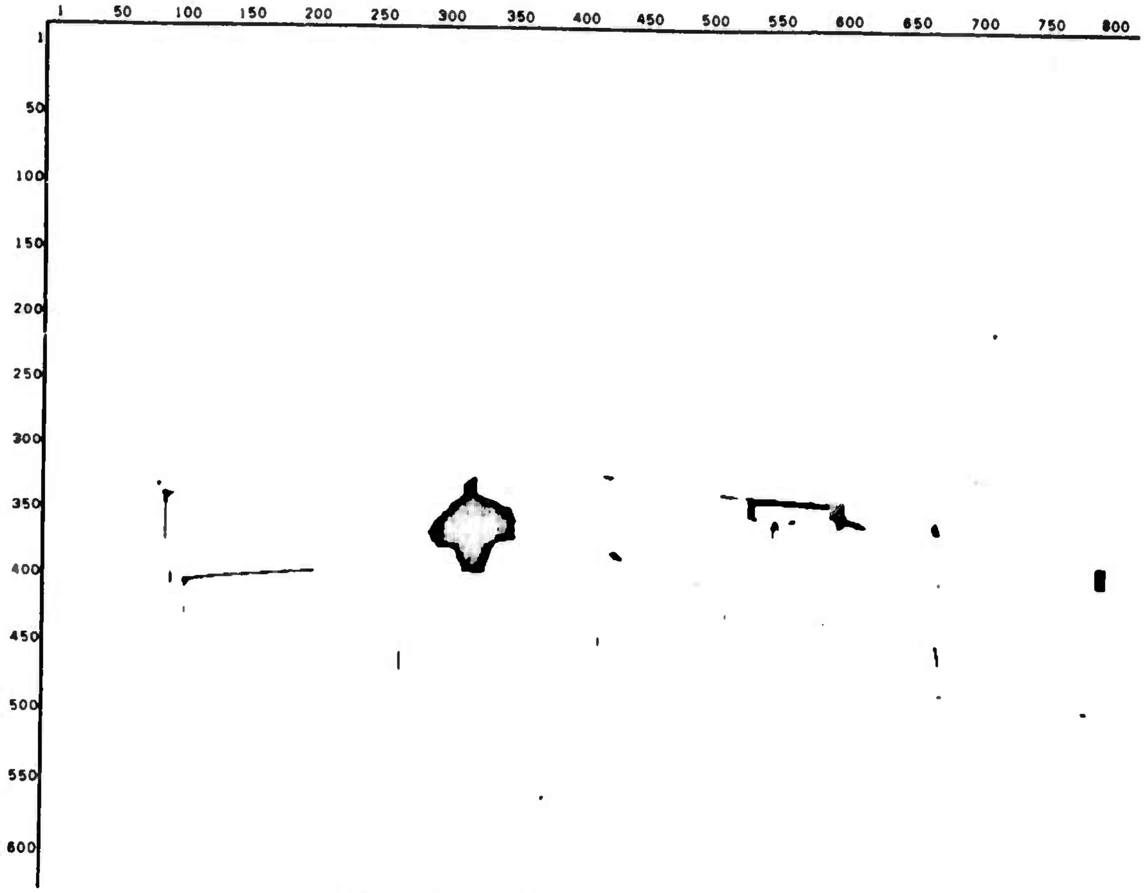


Figure 3.16. Contraction of busy mask for the room.

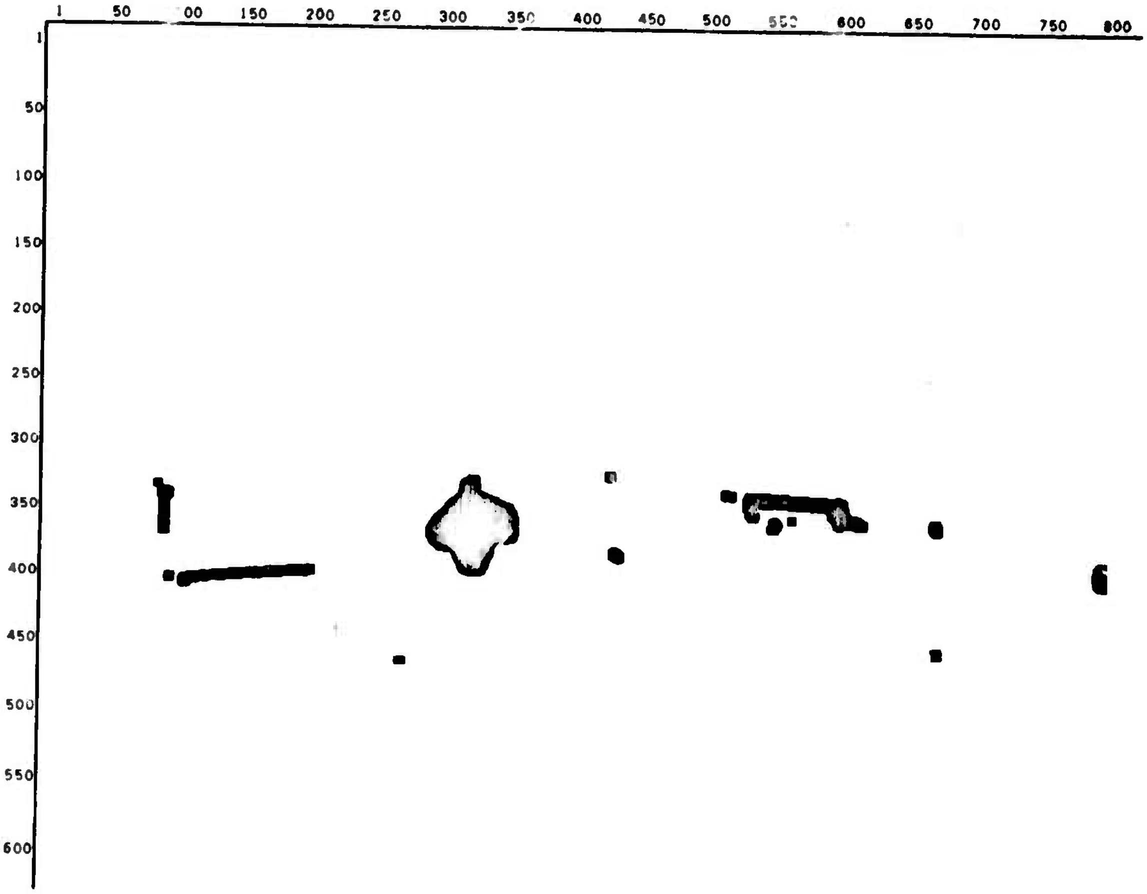


Figure 3.17. Expansion of busy mask for the room.

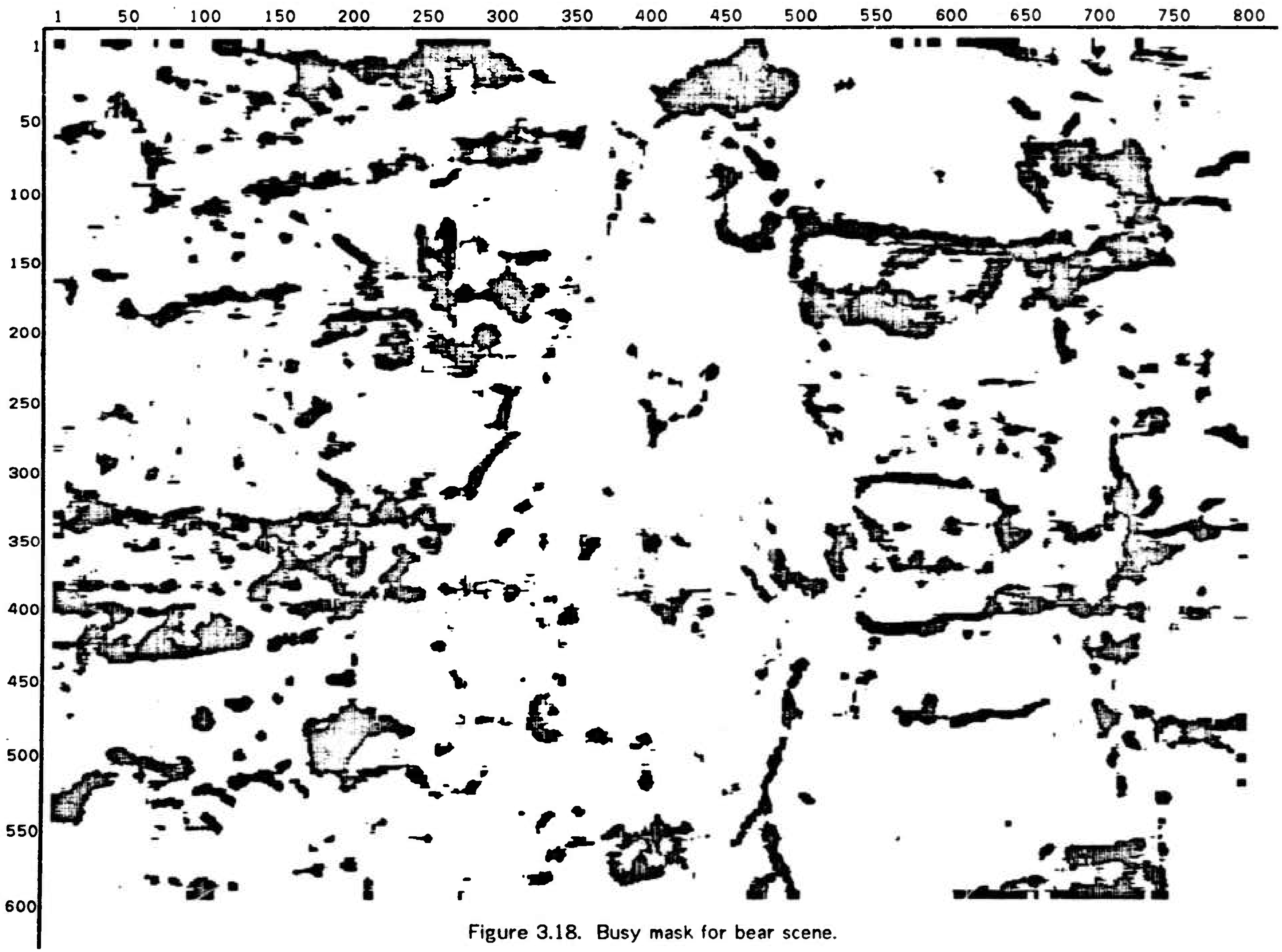


Figure 3.18. Busy mask for bear scene.

subroutine gives us an exact copy of the region's perimeter which we then fill. Since our efforts so far have given us a mask which is completely filled, an "anding" operation is employed to "punch out" the corresponding holes in the new construct. The operator compares bits in the mask with the proper bits in the original matrix. The final outcome is a binary picture matrix which is equivalent to extracting a window from the original matrix which contains precisely the desired regions. This window, however, will not possess any bits from the original matrix which are exclusive of the target area.

The derived region must now meet three criteria to qualify for inclusion as a busy area: it must be of sufficient size, of sufficient extent in either dimension, and of sufficient density. For a homogeneous region the minimum requirements are:

size = .2% of picture frame (960 pixels),
dimension = 25 pixels
density = 30% of mask frame.

The size requirement eliminates any candidate regions which might have been caused by small size or accident of position (e.g., the junction of the small checks in the design on the wall of figure 2.4.b). The dimension criterion cancels the texture effect given by thin objects (e.g., arms of the chair). It is determined for horizontally or vertically oriented regions by simply examining the dimensions of the image matrix containing the region under analysis. As noted above the matrix is a minimum bounding rectangle for the region, oriented with sides parallel to the x-y axes. The density parameter is computed as the percentage of area of the extracted mask which is turned on. The 30% limitation eliminates business due to thin objects oriented in a non-vertical or non-horizontal direction.

The last task of the preprocessing phase of the algorithm is the very simple one of constructing a completely filled binary template, equal in dimensions to the size of the picture (600x800). We use the term template in the same sense that we have used the expression mask. They both refer to binary matrices or pictures which represent areas of the total scene. They are simply black and white images which, for any (i,j), p(i,j) is either 1 or 0. We follow the common practice of using the set of cells for which p(i,j)=1 to represent the figure and the set of cells for which p(i,j)=0 to represent the background. Although it is usually the case that the highest possible value of any pixel of any given picture has the highest intensity we have been depicting the points with value 1 as black. This gives the figure positive emphasis. Besides the general meaning given to the terms in question, we attach a very specific connotation within the context of the basic algorithm. Templates are binary images which always represent points of a subpicture under some phase of analysis which have not yet been segmented out. Masks on the other hand will always contain one or more completely processed segments.

The first evocation of the main segmentation procedure skips step 5 and compares the size of the initial template to the allowable minimum. This decision is necessary to avoid collecting those small regions which might result from imprecise boundary determination or fallout from complex scenes. For homogeneous scenes a limit of .1% is set. Having accepted the template for further processing an associated blank mask is constructed and pointers to both are pushed on the control stack. The template on the top of the stack is always representative of the subpicture currently under analysis, in this case the entire picture. All regions extracted for the current level are copied into the associated mask.

The next step in the process requires the derivation of histograms of the sensory data corresponding to the turned on bits of the current template. Figure 3.19 shows the graphs for the nine parameters for the office scene. From these a decision is made as to the peak which best indicates area(s) of uniformity. The first choice is to look for signs of black or white surfaces. These receive particular emphasis because of the special role they play visually. Quite often they will occur as small or narrow regions which separate larger areas. An example of this would be the sections of white trim in figure 2.4.c. We attempt to locate such surfaces by looking for peaks in the 0-60 and 200-250 ranges of the intensity parameter. The peaks furthest to the left or right are selected first.

If no signs of white or black areas are present in the high or low ranges of the histograms other suitable indicators of uniformity must be sought. In selecting modes which represent interesting areas of the image, certain features are desired. Peaks which have minima that come close to the baseline and which have neighboring modes of similar height, indicate sharply defined uniform regions. Table 3.1 gives the conditions which attempt to model acceptability under these criteria. Notice that the requirements for the priority 3 condition is much less stringent than for the others. This is allowed because of specialized knowledge which tells us that achromatic points lie in the 0-150 range (approximately) of the saturation parameter. Any kind of minimum around this range is an indicator of a cutoff value for the neutral color points of the picture. The point that we wish to emphasize here, is that there exists some method for selecting useful peaks from the histograms. The numbers in table 3.1 are meaningful only for the pictures actually analyzed. As the algorithm is improved and more pictures are investigated values will change and better methods evolve.

Modes possessing the correct characteristics are isolated in the order of precedence given in the table. The search is successful and halts when all peaks for all histograms of a given priority have been found. If no adequate peaks are found the histograms are determined to be "monomodal". If success is achieved with more than one candidate, arbitration is in order. Selection of the best peak is based upon a point count given for certain qualities:

peak with lowest average value of minima = 2 points
peak with maximum to highest minimum ratio of n:1 = n points.

If a tie results from the point count the peak with the highest maximum value is chosen.

Implicit in the peak selection process was the determination of relative minima and maxima. We have developed a subroutine which meets with fair success in this endeavor. Small discontinuities, inflection points and temporary plateaus can be suppressed. Human adjustment is sometimes necessary since the program does not always select the best extreme point. This will sometimes happen when the curve

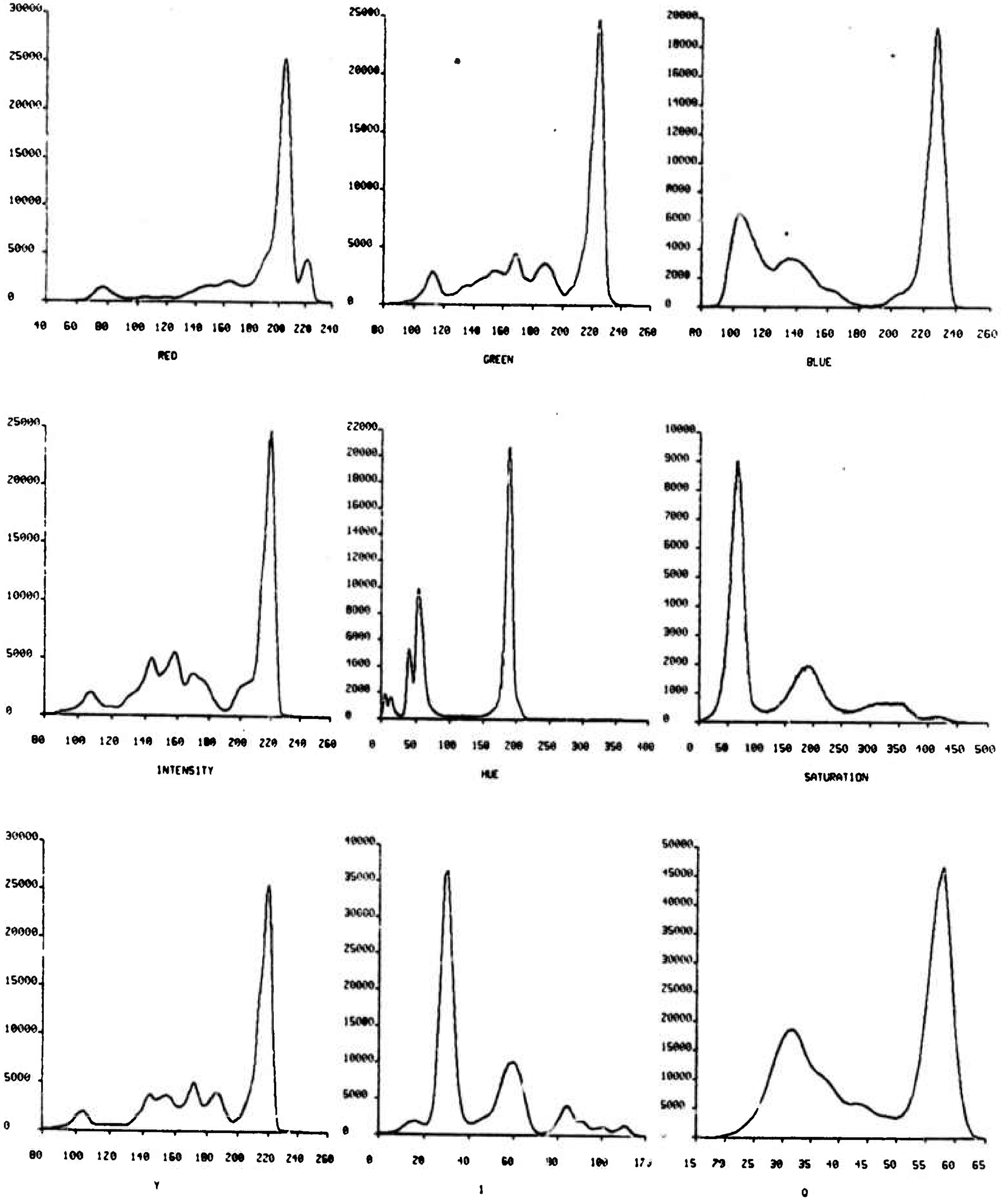


Figure 3.19. Nine property histograms for the room scene.

<table>
<tr><th rowspan="2">PRECEDENCE</th><th colspan="2">CONDITIONS</th></tr>
<tr><th>HOMOGENEOUS</th><th>NON-HOMOGENEOUS</th></tr>
<tr><td rowspan="3">1</td><td colspan="2">Both minima ≤10% of maximum frequency value for the histogram.</td></tr>
<tr><td>Max/min ratio* of 4:1</td><td>Max/min ratio of 2:1</td></tr>
<tr><td colspan="2">At least one minimum separates another peak with a max/min ratio of at least 2:1.</td></tr>
<tr><td rowspan="3">2</td><td colspan="2">Both minima ≤25% of maximum frequency value for the histogram.</td></tr>
<tr><td>Max/min ratio of 5:1</td><td>Max/min ratio of 4:1</td></tr>
<tr><td colspan="2">At least one minimum separates another peak with a max/min ratio of at least 2:1</td></tr>
<tr><td rowspan="2">3 **</td><td colspan="2">A local minimum divides two peaks both of which have max/min ratios at least 2:1</td></tr>
<tr><td colspan="2">Maxima are within 10% of each other.</td></tr>
<tr><td rowspan="3">4</td><td colspan="2">One minimum in the 0-200 range of saturation histogram.</td></tr>
<tr><td colspan="2">Max/min ratio of at least 2:1.</td></tr>
<tr><td colspan="2">The minimum in the 0-200 range has another peak to the opposite side with a min/max ratio of 1.2:1.</td></tr>
<tr><td rowspan="2">5</td><td colspan="2">Both minima are ≤10% of maximum.</td></tr>
<tr><td colspan="2">10% of total area under the curve lies to the side of one of the minima; ie. outside the peak area.</td></tr>
</table>

* The ratic of the maximum is to the minimum bounding the peak which is of highest value.

** This case applies to histograms which are essentially bimodal. Both peaks are conditionally acceptable.

Table 3.1. Table of conditions for peak acceptability.

possesses a long irregular tail or a number of spikes due to the mixed pixel phenomenon. This latter problem is most troublesome in the case of the hue parameter. It is caused by averaged intensity values produced by the digitization process when the quantization window straddles the junctures between two surfaces of different properties.

For the scene under analysis, we discover that there are several large peaks which satisfy our conditions. Step 10.1 determines that the best peak is provided by the blue sensory data and that the proper cuttoffs are 190 and 241. The threshold limits are approximated by the user. If the minima come very close to the base line, they are given as the desired values. If the peak is the result of a priority 3 condition (table 3.1), the value of the cutoff on that side of the peak is given as the intensity at which the minimum occurs. This is done because of the special implications noted above for a minimum in the given range of the saturation histogram. In all other cases a Gaussian extension of the sides of the curve to the base line is estimated. In view of our initial assumptions this seemed a reasonable procedure to follow.

Applying the threshold operator to the parameter matrix for the blue data yields the results shown in figure 3.20. We note that the homogeneous nature of the wall has provided a very clean result. Application of step 14 produces histograms of a monomodal nature, as can be seen in figure 3.21. Notice that if the tail as shown for the blue data histogram in our estimation of threshold limits had been cut off, the results shown in figure 3.22 would have been obtained. This is clearly a less desirable region and illustrates a danger in always assuming the Gaussian extension. We try to avoid this when we can by taking a better alternative when it is available.

Now a smoothing operator need only be applied and the same extraction process cited earlier in the preprocessing phase for busy areas. Each region that is extracted at step 15.2 must be evaluated for proper size (at least .1% of the picture) and non-inclusion of points common to the heavy texture regions. If the latter condition is not found to hold true the previously derived business matrix is used to mask out the overlapping area. Repeated application of the loop (steps 15.1 through 15.5) produces the template on top of the control stack (figure 3.23) and the associated mask (figure 3.24). A black and white view of the applicable pixels of the original picture are shown in figures 3.25 and 3.26. respectively.[2]

Complete derivation of all the uniform regions at this level of processing brings us to step 5 of the algorithm. From the template of figure 3.23 we extract the largest unprocessed region (figure 3.27). The template and an associated blank mask which mark the new level of analysis are pushed on the control stack. Histograms are derived from the sensory source files using the template to mask out the relevant pixels (figure 3.28). As there are no remaining white or black regions we proceed to the decision box of step 11. Since there are a number of candidate modes a best choice must be determined. Peaks exist for the red, intensity, hue, Y, and I parameters. This time, hue provides the most suitable peak. Thresholding the sensory data at the estimated limits of 0 and 30 produces the cushions of the chair as shown in

[2] All of the black and white photos in this section were produced from a video monitor. Low resolution due to lack of memory bandwith gives the somewhat coarse effect.

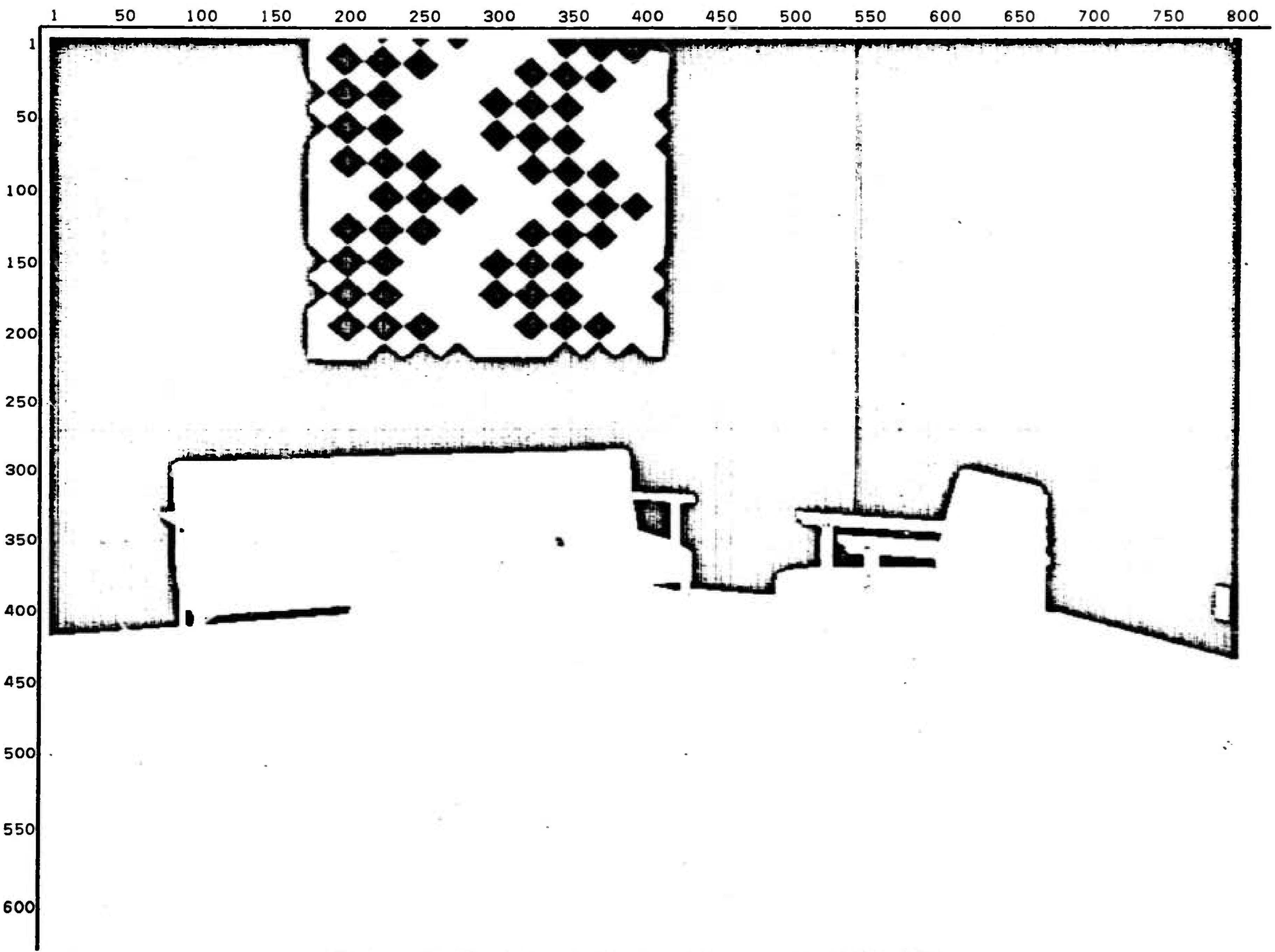


Figure 3.20. Points thresholded from blue parameter (190 to 231).

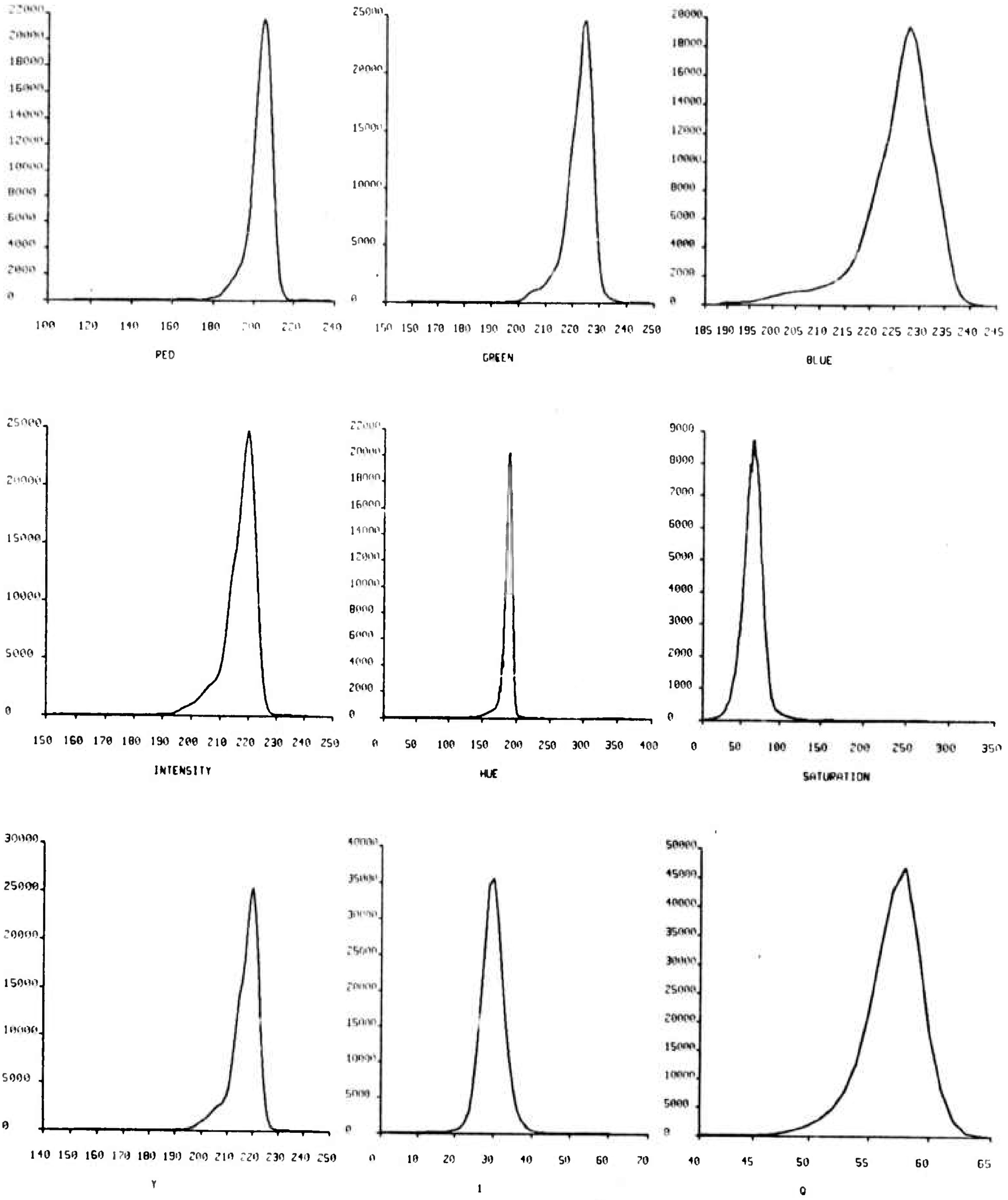


Figure 3.21. Nine histograms for points in figure 3.20.

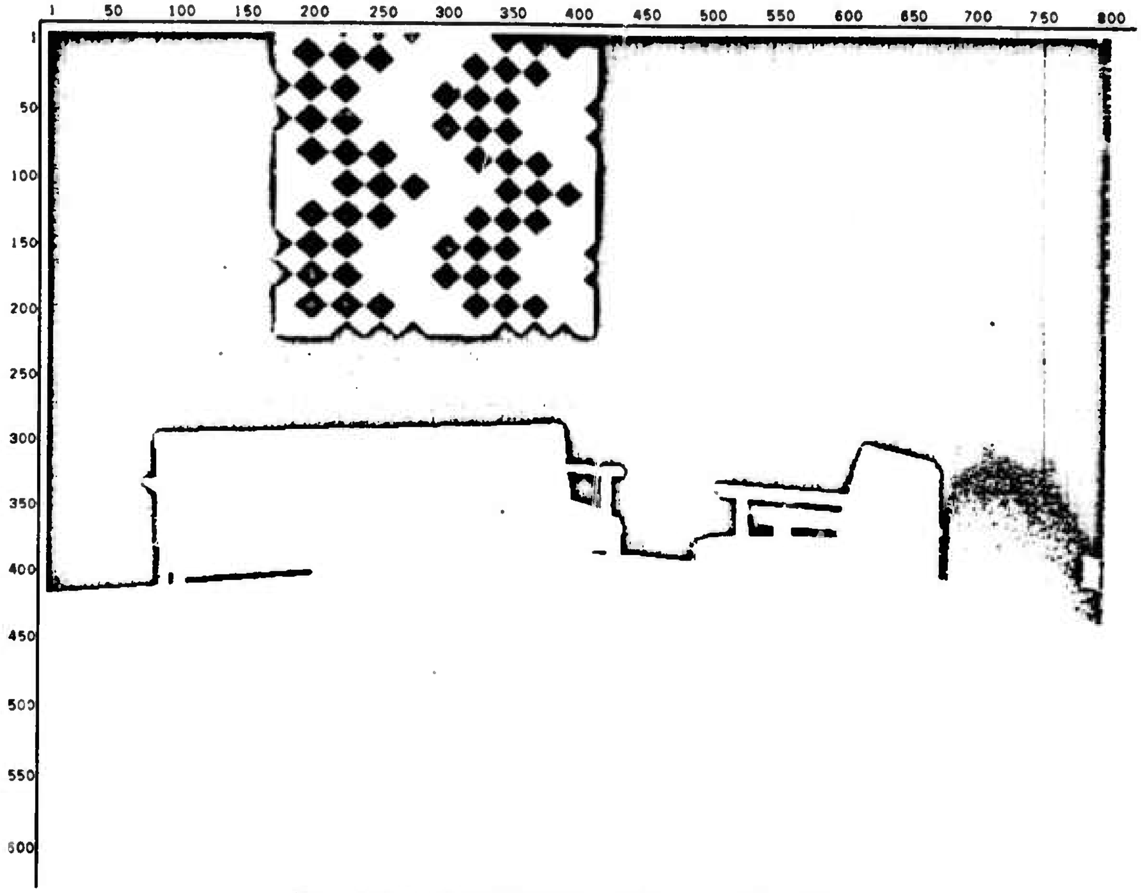


Figure 3.22. Points thresholded from blue parameter (210 to 230).

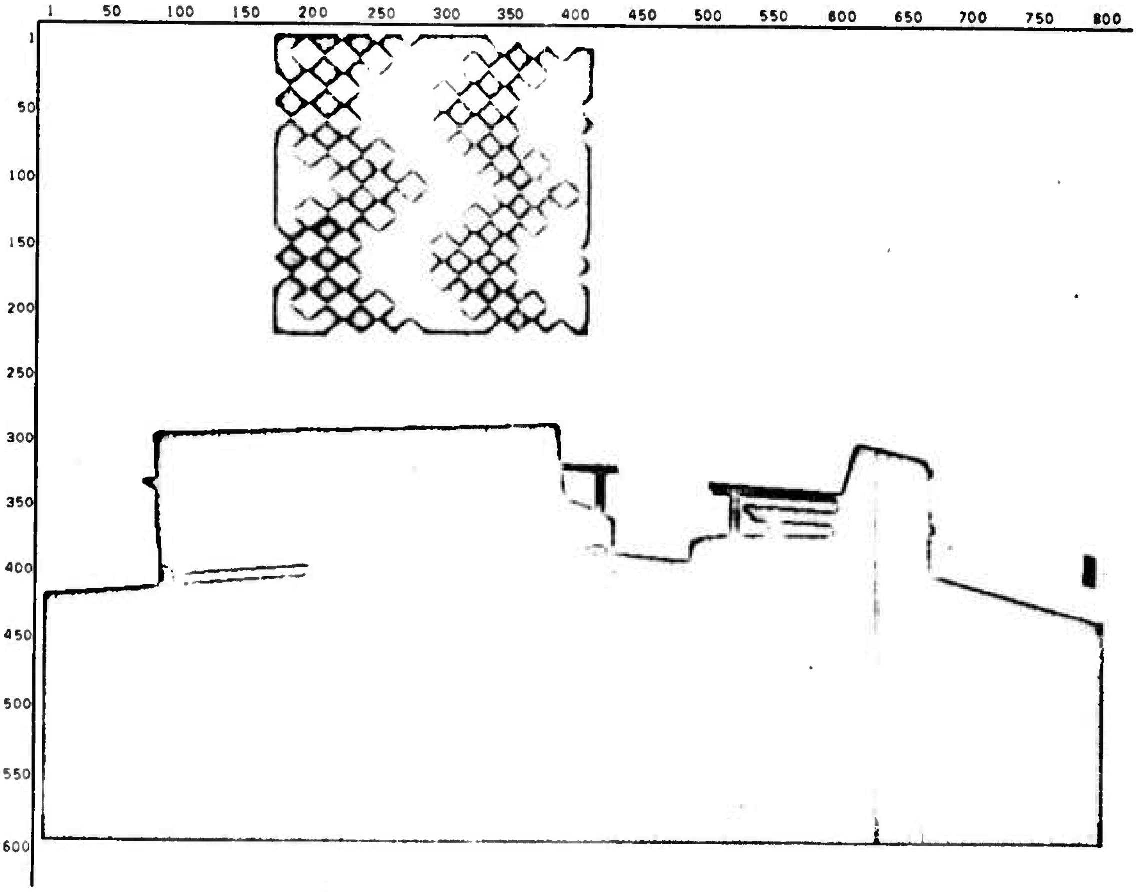


Figure 3.23. Portions of room remaining after masking.

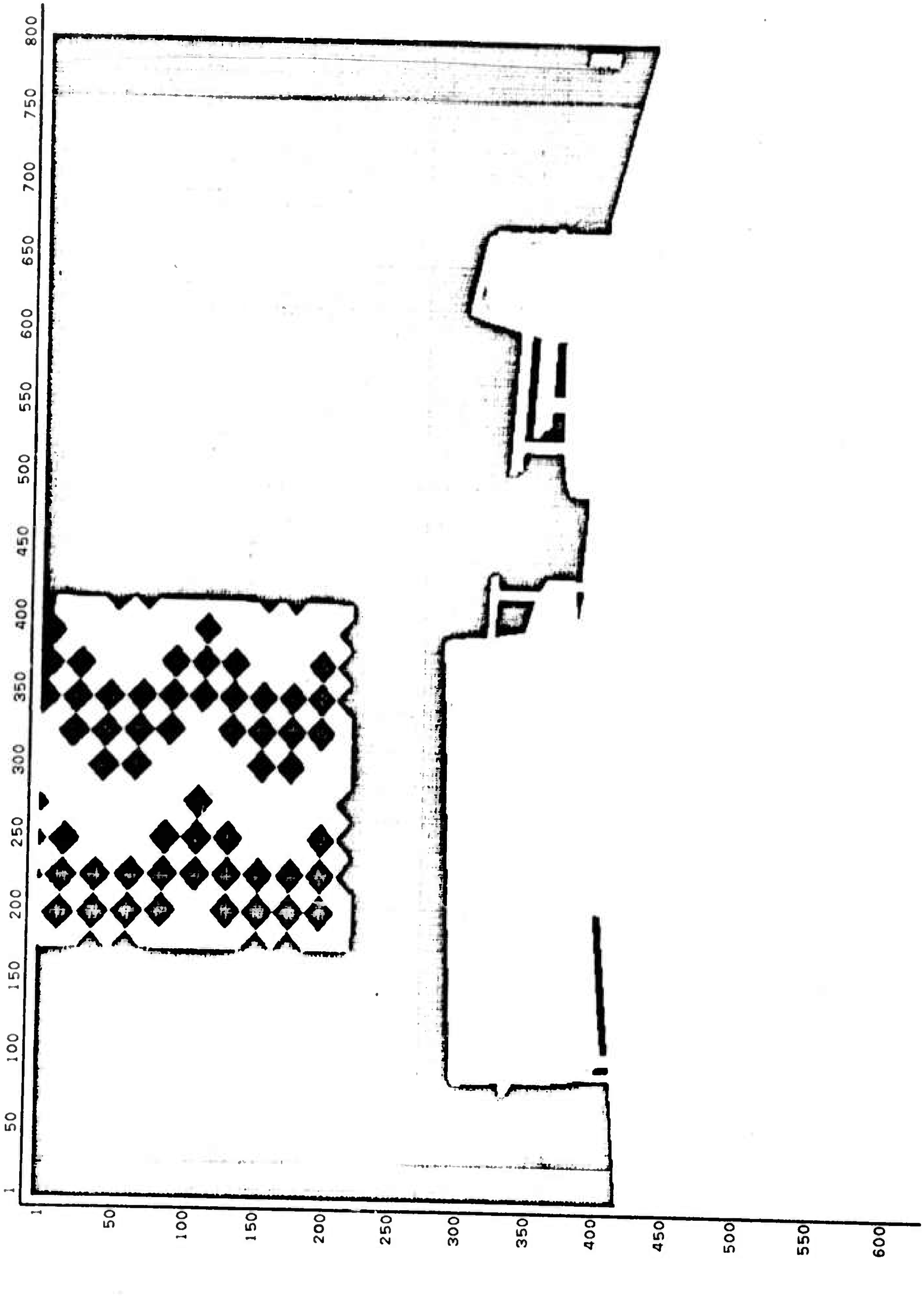


Figure 3.24. Regions extracted from first level of recursion.

Figure 3.25. Black and white picture of points corresponding to figure 3.23

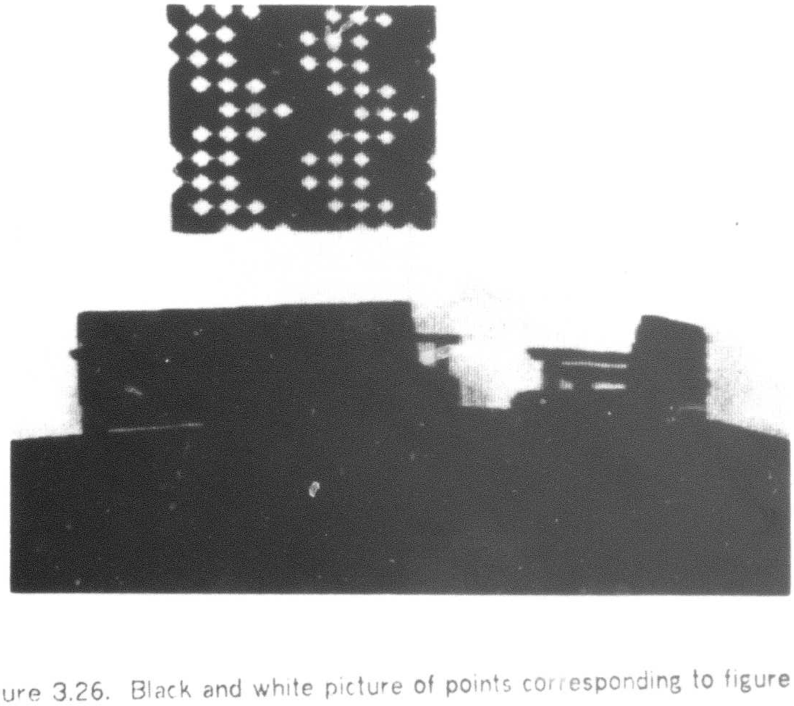


Figure 3.26. Black and white picture of points corresponding to figure 3.24.

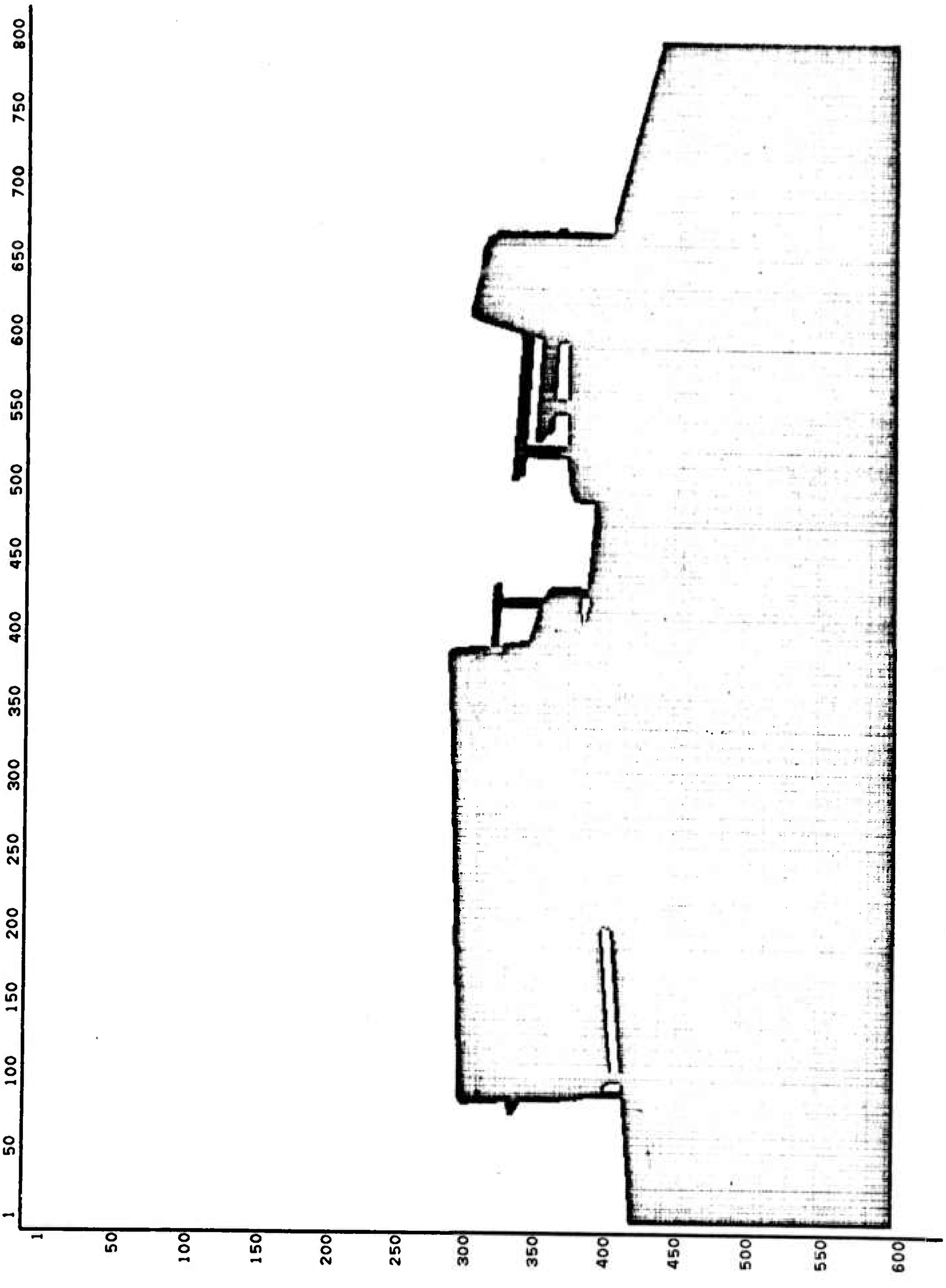


Figure 3.27. Largest remaining unprocessed portion of the room.

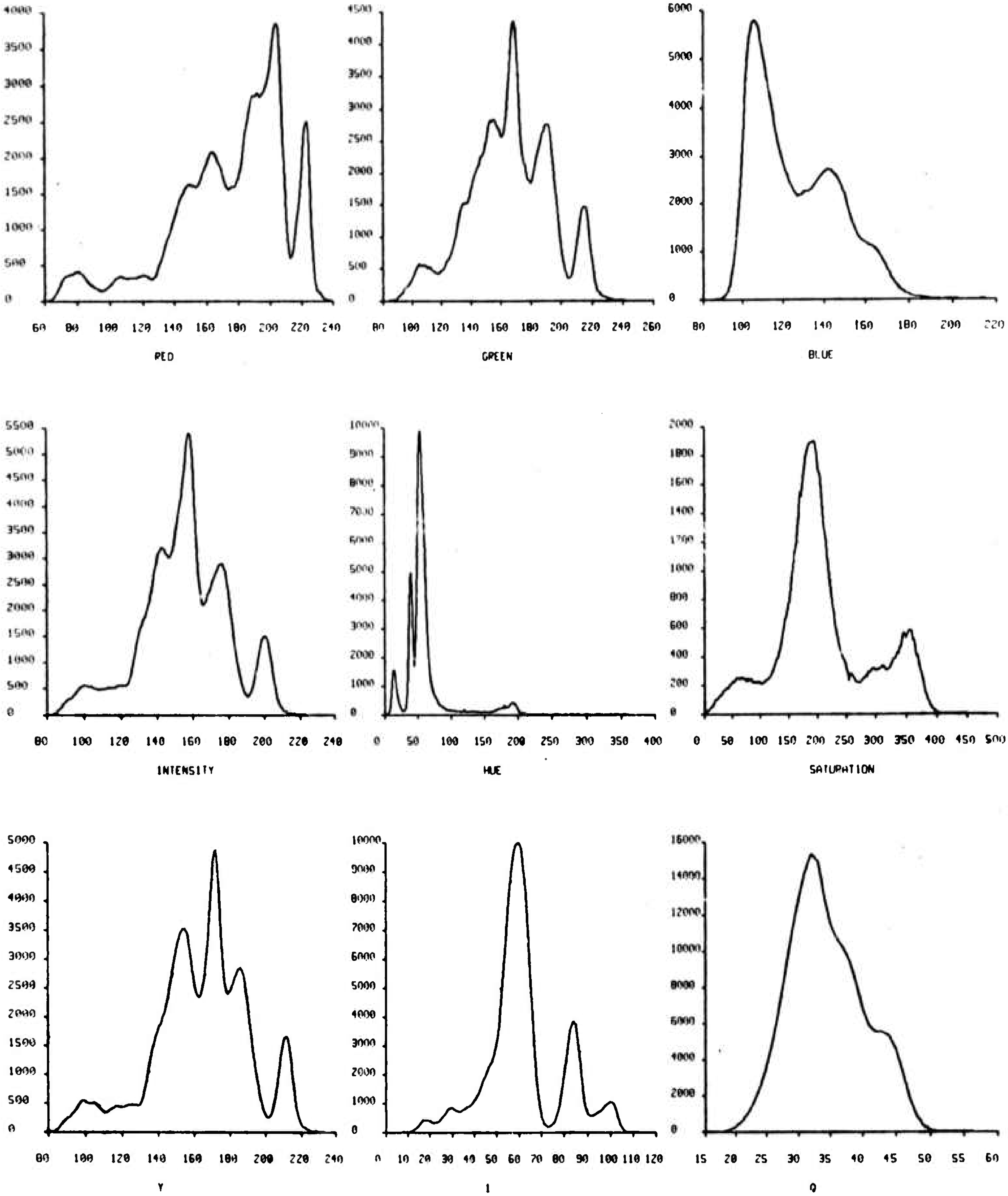


Figure 3.28. Nine parameter histograms for template in figure 3.27.

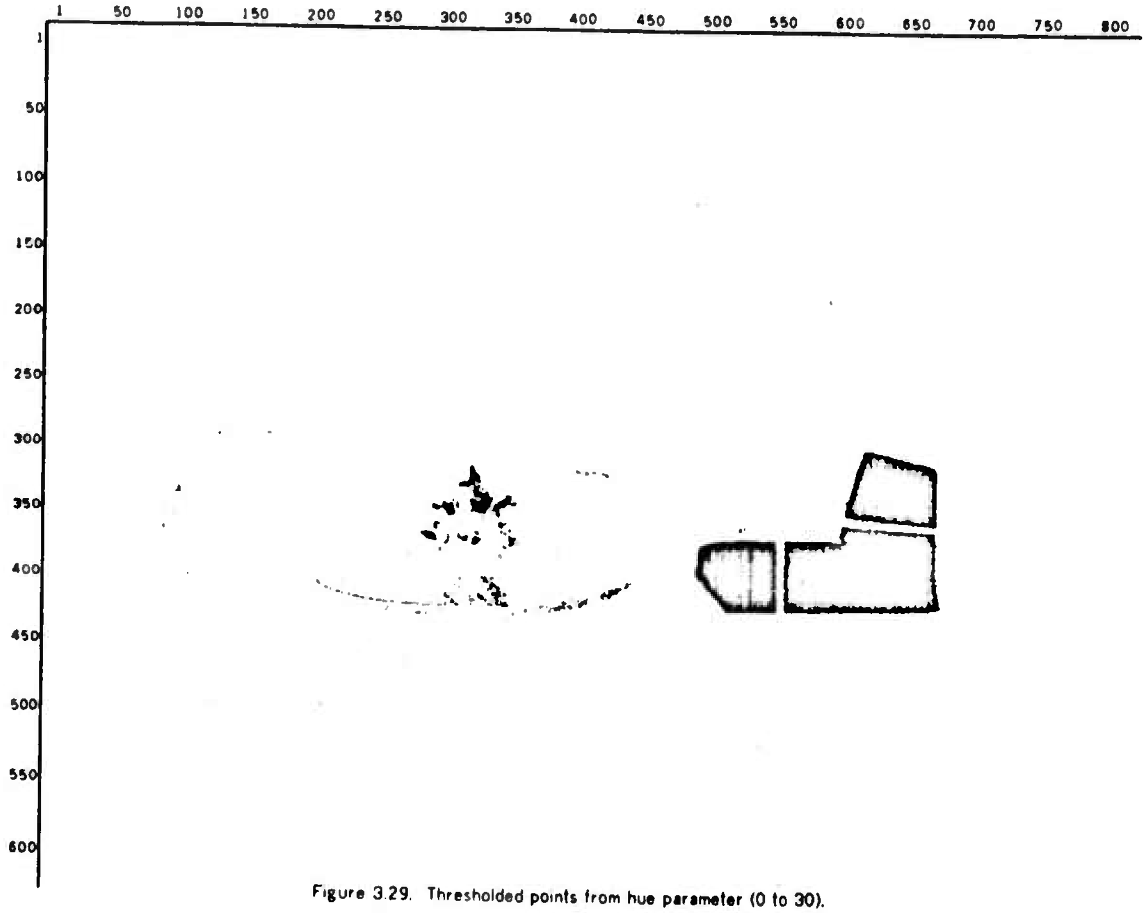


Figure 3.29. Thresholded points from hue parameter (0 to 30).

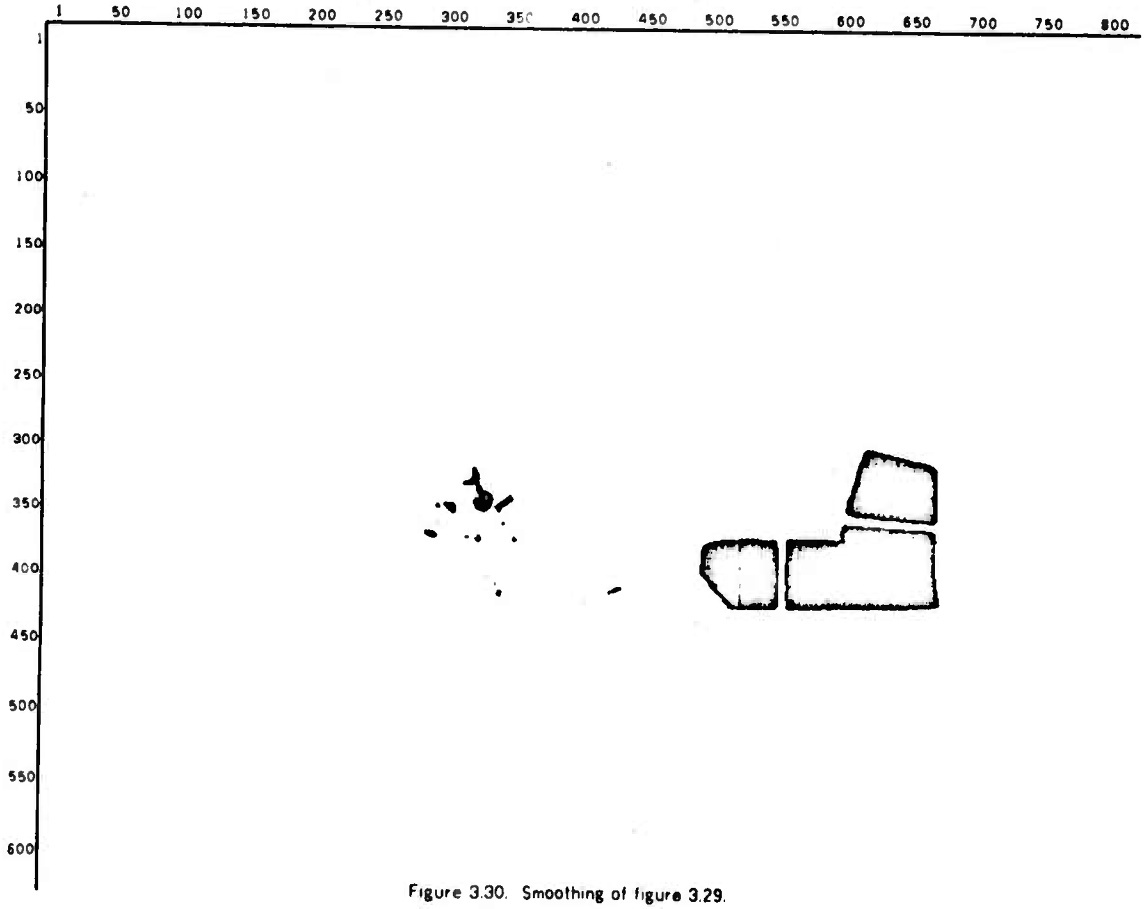


Figure 3.30. Smoothing of figure 3.29.

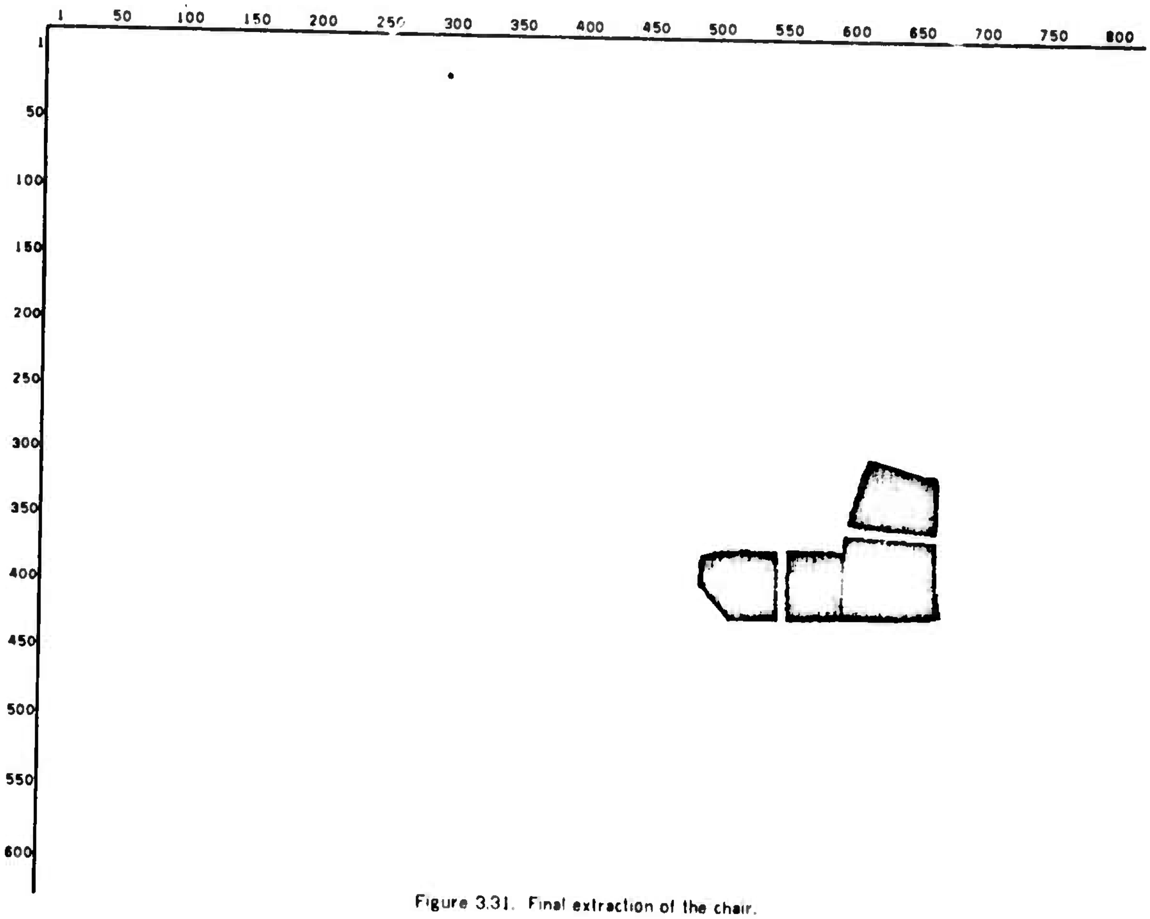


Figure 3.31. Final extraction of the chair.

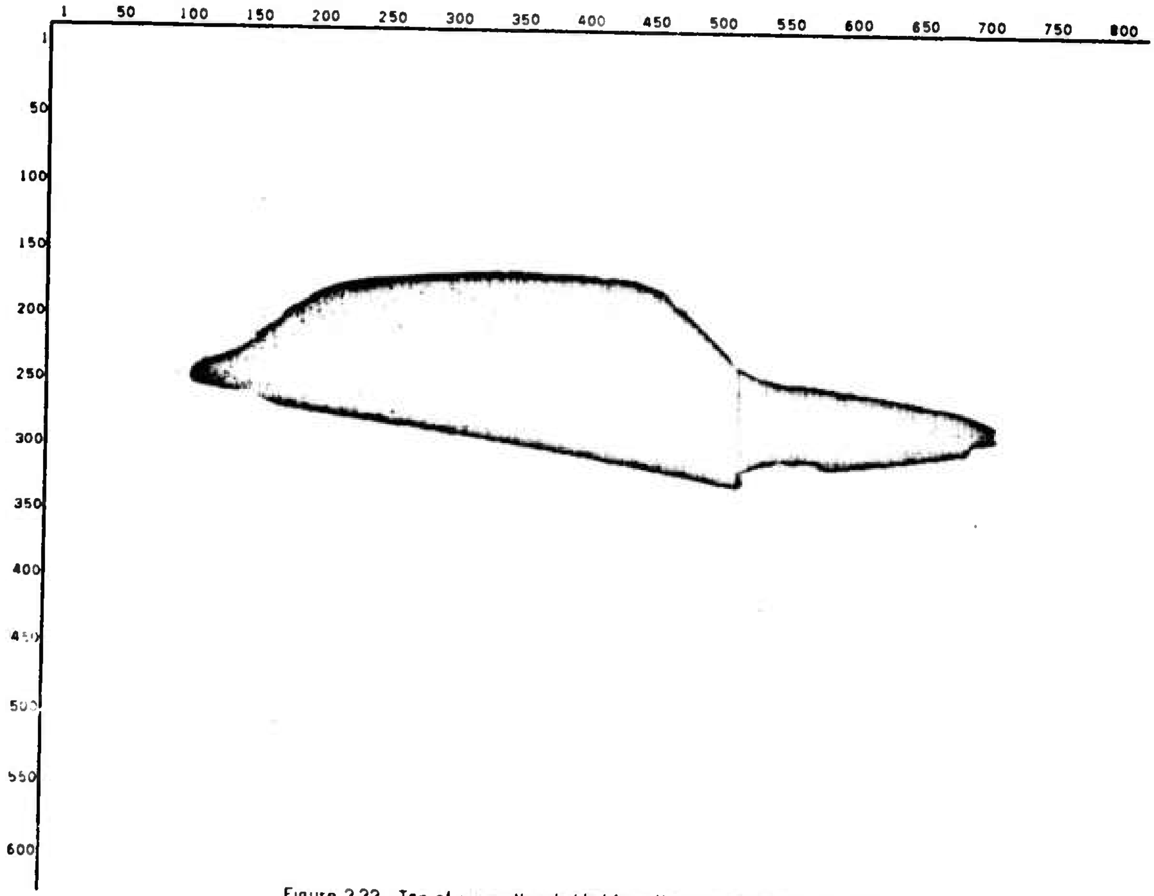


Figure 3.33. Top of car as thresholded from the analysis of the car scene.

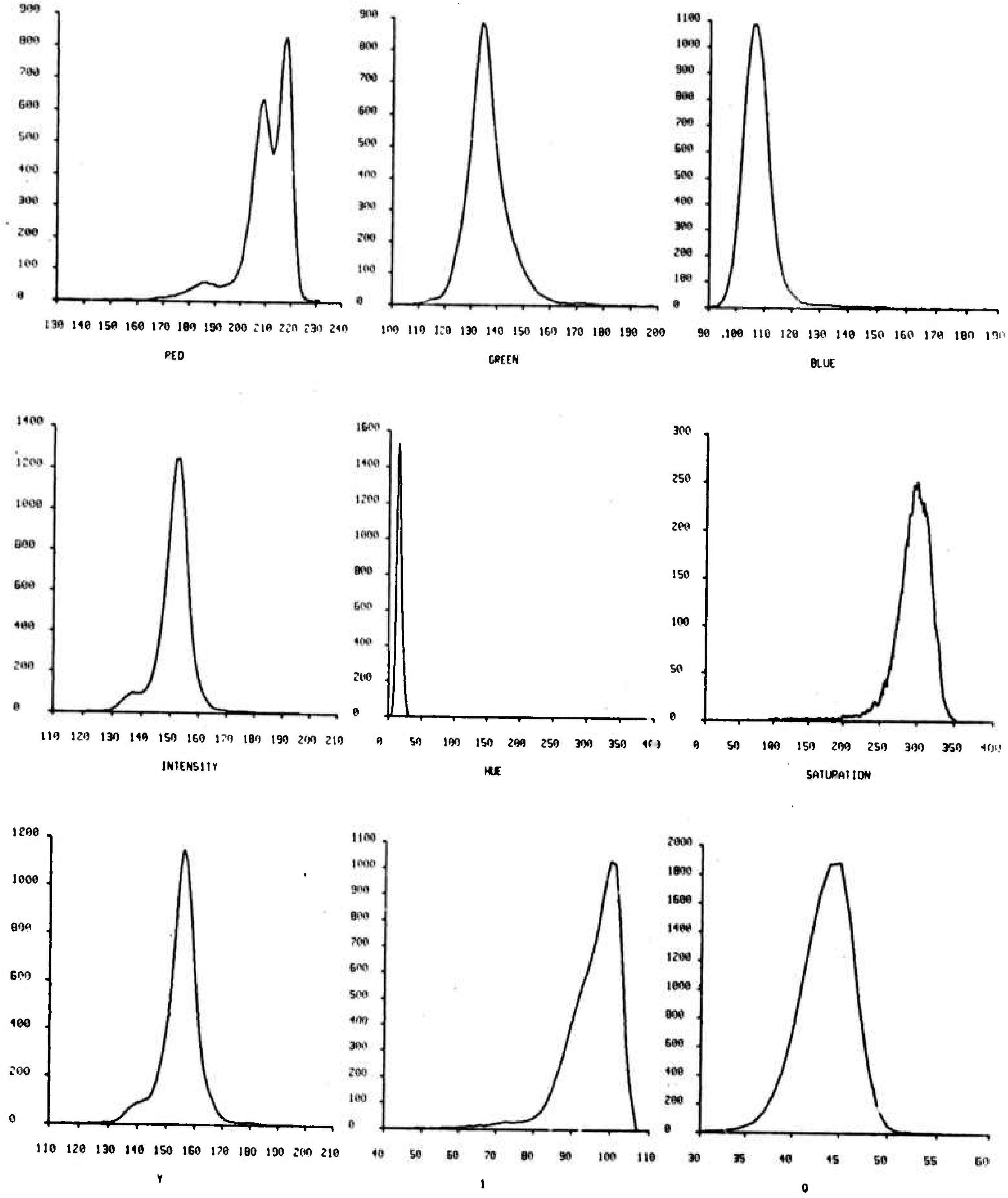


Figure 3.32. Nine parameter histograms for the chair.

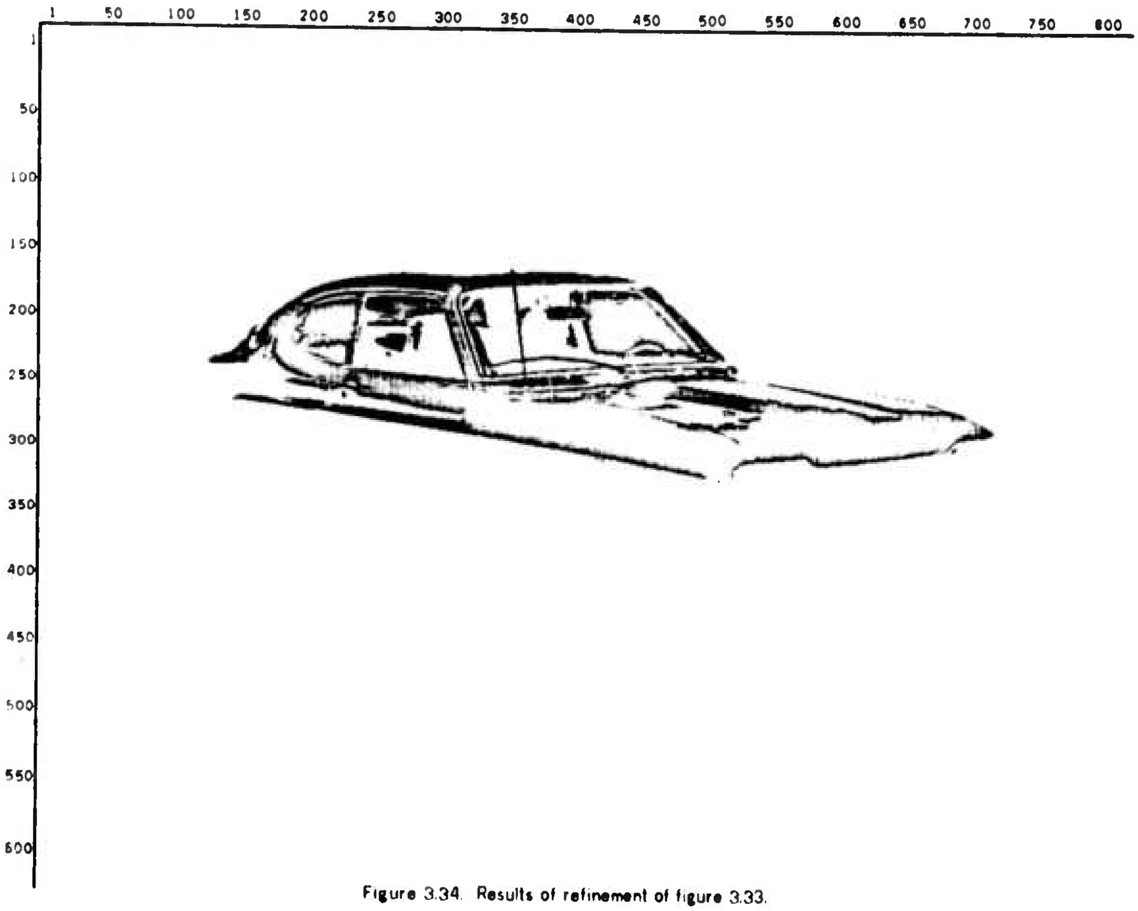


Figure 3.34. Results of refinement of figure 3.33.

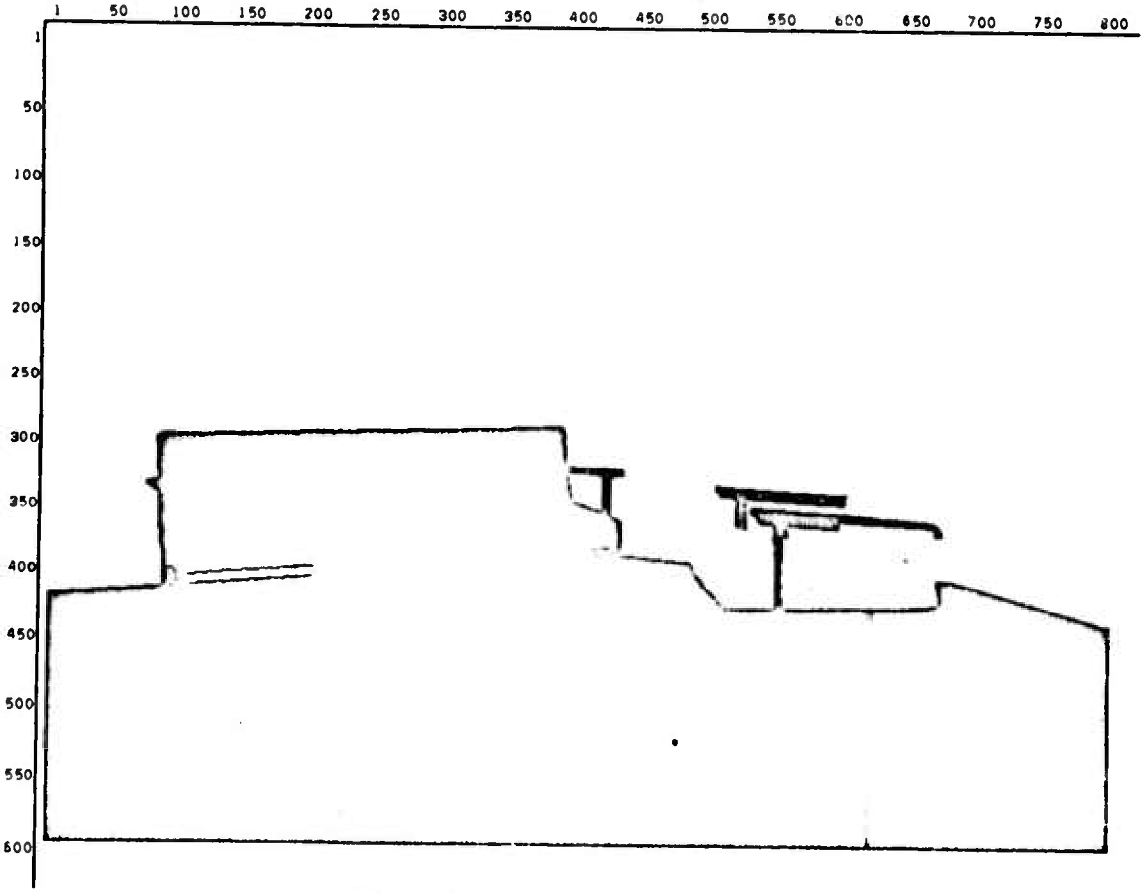


Figure 3.35. Result of removing chair from the template.

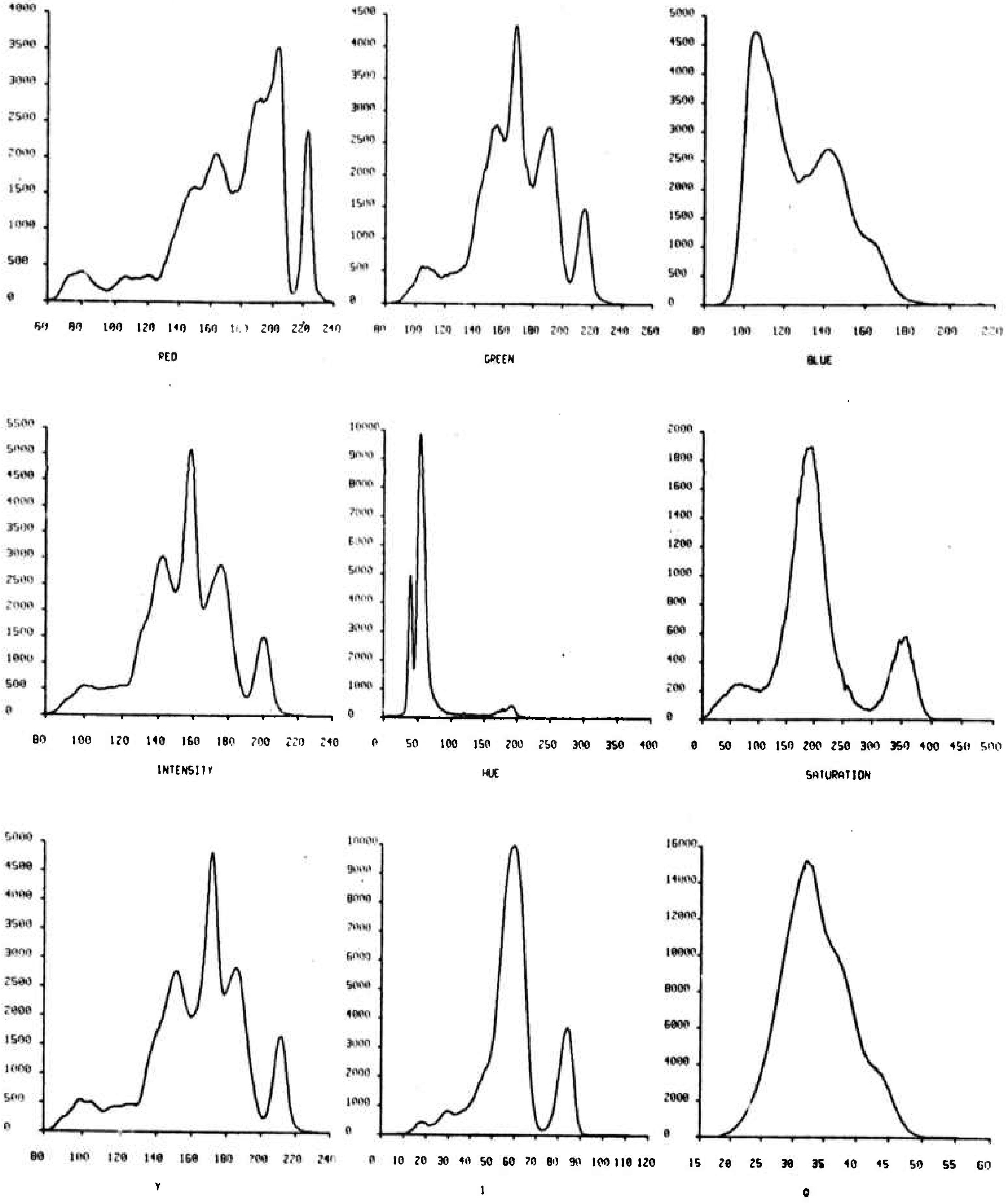


Figure 3.36. Nine parameter histograms of template from figure 3.36.

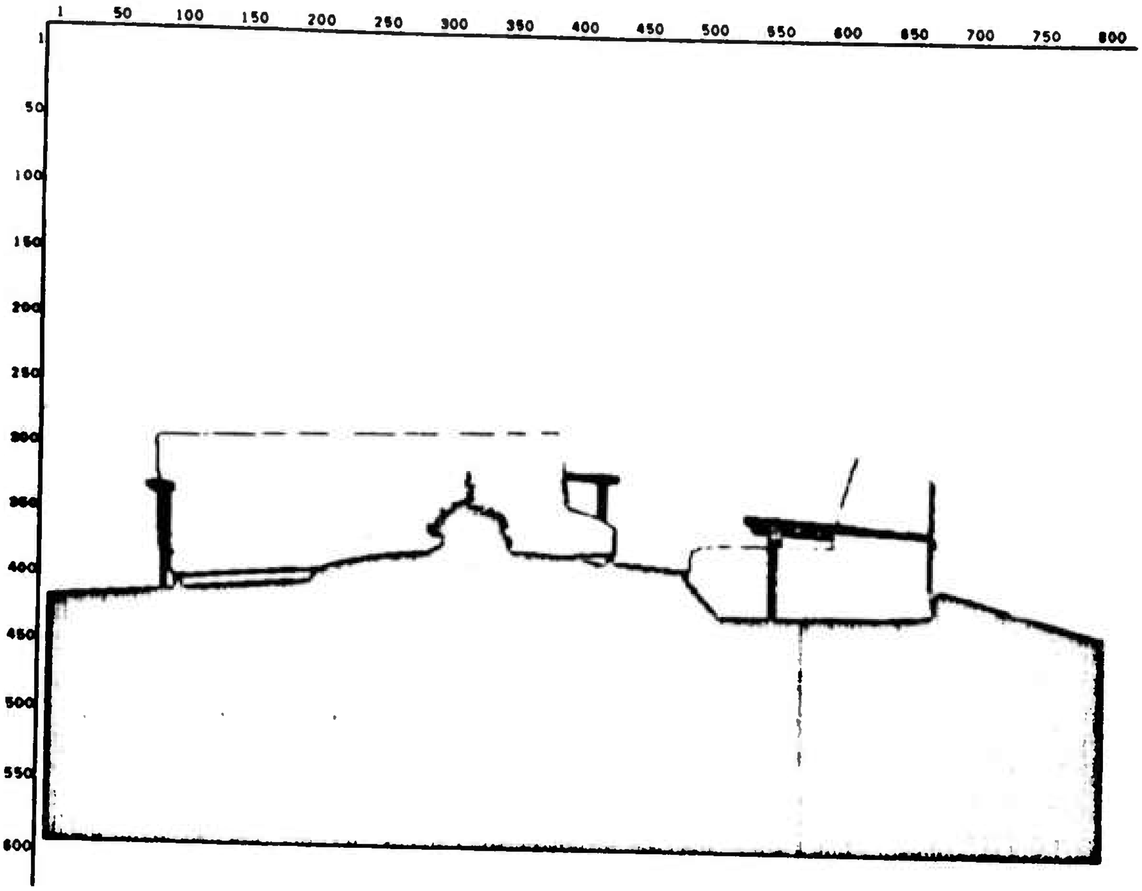


Figure 3.37. Result of masking out the sofa.

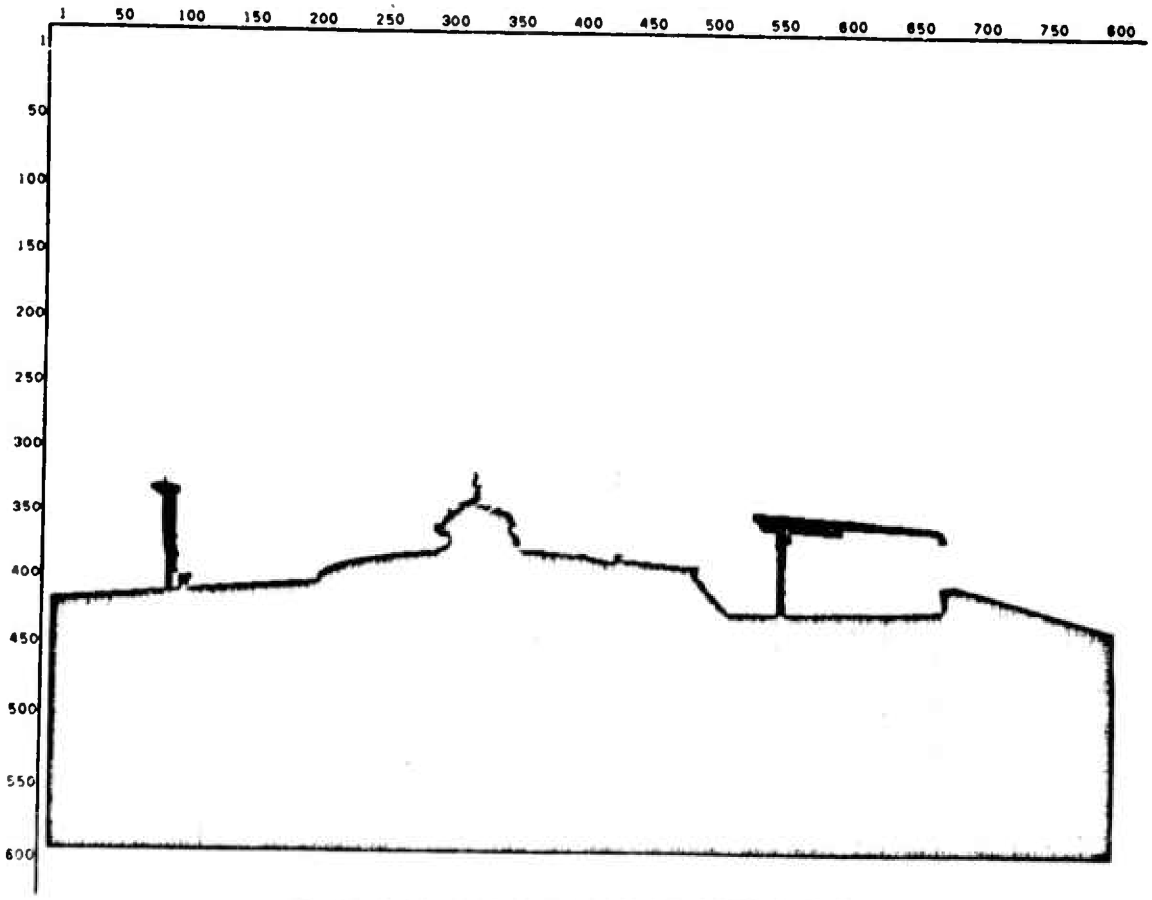


Figure 3.38. Result of application of contraction expansion operators.

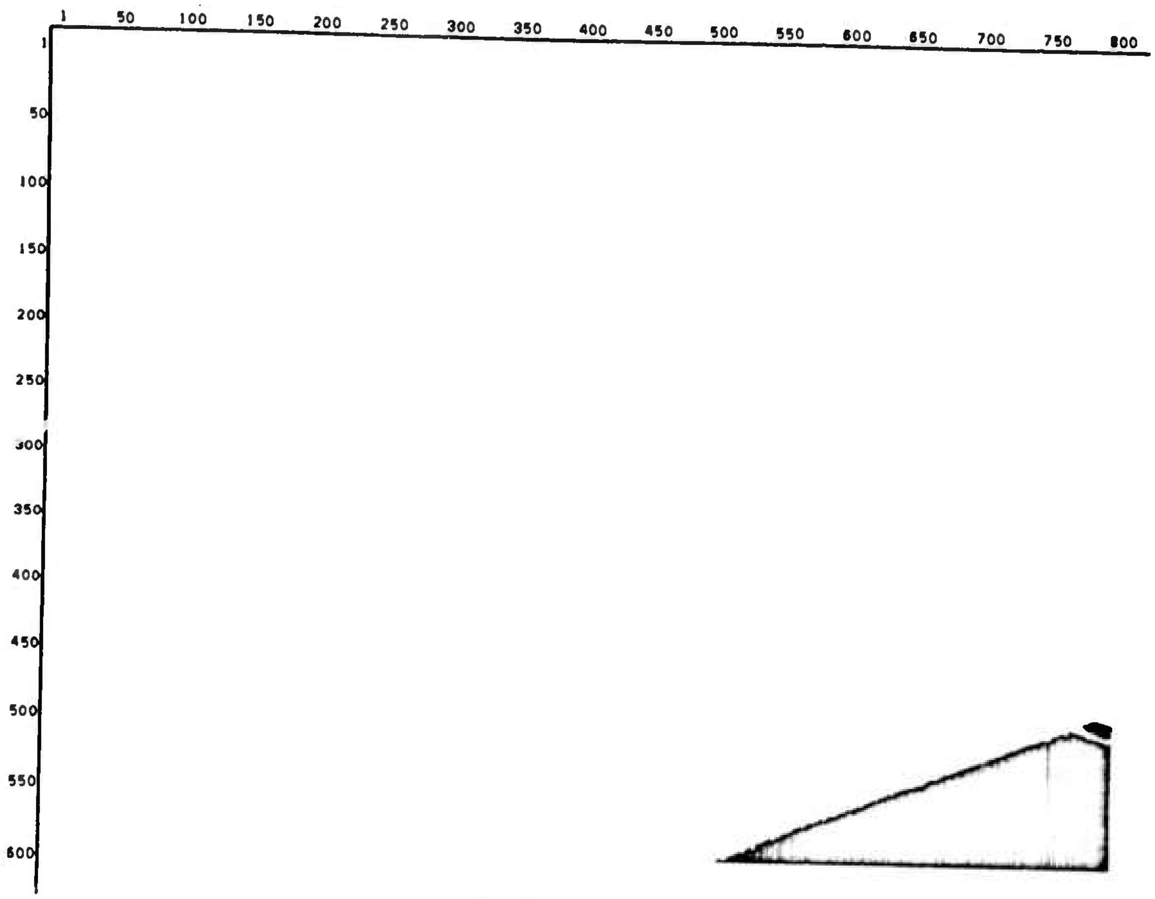
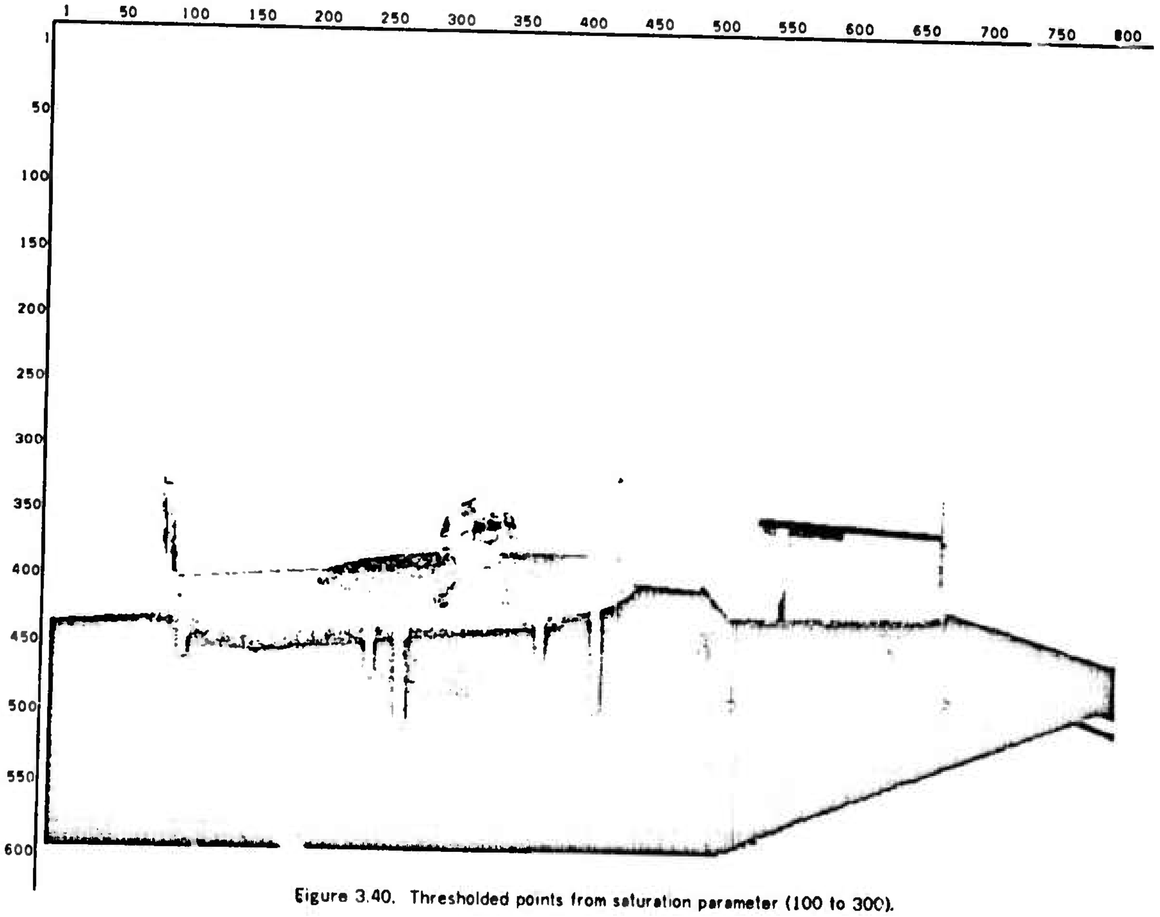


Figure 3.39. Extraction of highlighted rug.



Eigure 3.40. Thresholded points from saturation parameter (100 to 300).

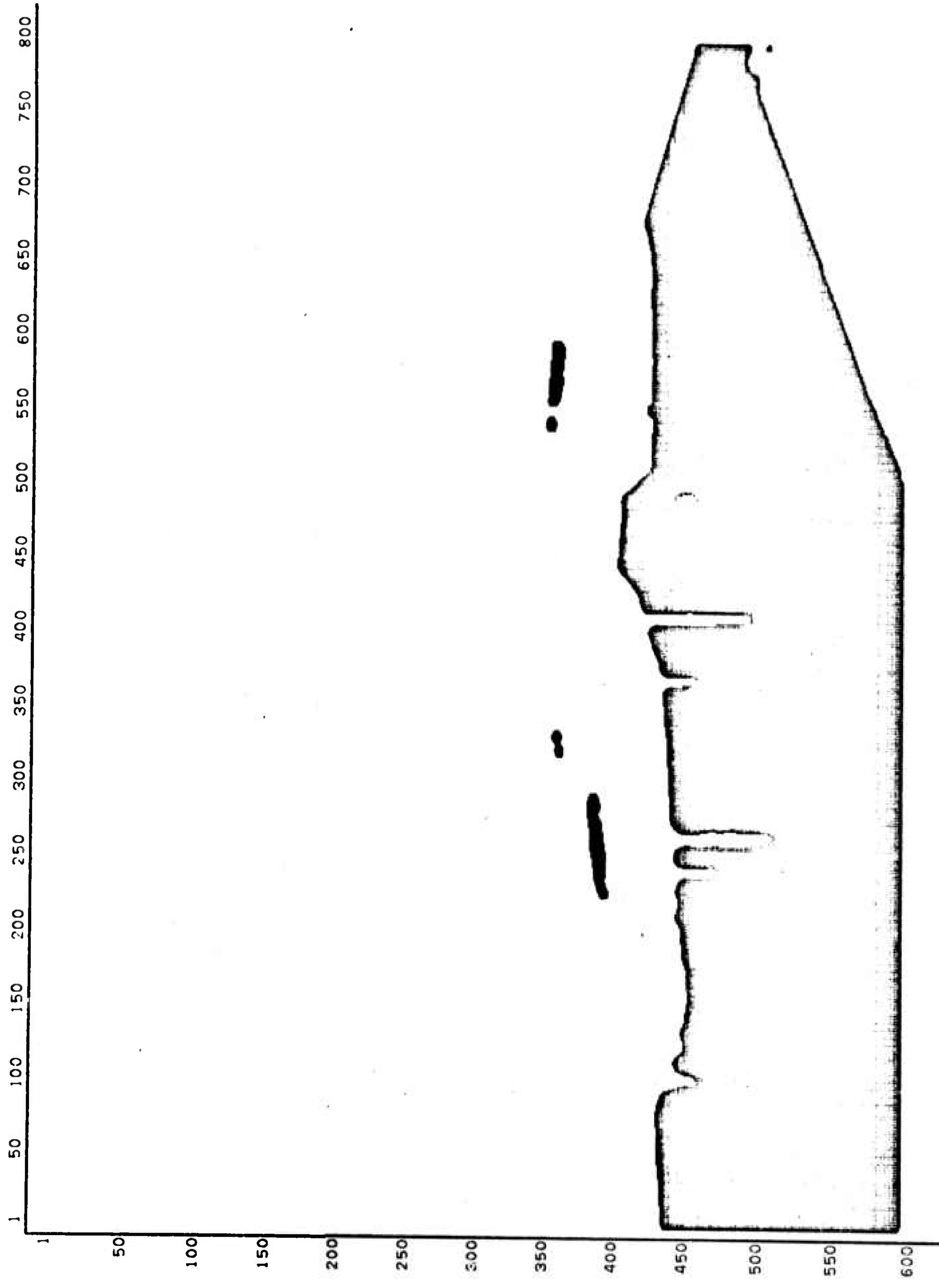


Figure 3.41. Result after contraction-expansion.

figure 3.29. Additional smoothing gives the results of figure 3.30. Eliminating those areas which are in the busy area gives figure 3.31.

Histograms produced at step 14 for the area which is uniform in the hue dimension are shown in figure 3.32. If the same criteria were invoked at this phase of the operation as were used at step 11, a priority 5 peak for the red parameter between 196 and 226 would be found. It might be argued that this is exactly what should be done; i.e., all regions should be refined until they are absolutely uniform in all parameters. If we perform an additional thresholding, we find the points eliminated were the cracks and seams of the upholstery. While useful regions could still be extracted from the further refinement, why make a difficult problem harder by additionally fragmenting areas which will have to be assembled by the identification module. The situation can become much worse for objects which contain transparent sections and reflections. The analysis of the car scene produces a segment (figure 3.33), uniform in the blue dimension, which is certainly much easier to identify than the result obtained from an additional refinement (figure 3.34). Such examples argue that it is best to take the conservative view in regards to additional processing of regions already found to be uniform along one dimension. We can note indications of further divisions and investigate a further refinement if higher level knowledge directs a search for subcomposition. We take this approach and modify the conditions in table 3.1. determination of "uniformity" at this point of the algorithm. The more stringent criteria allow further reductions of uniform regions but require stronger evidence of clean separation of data. If it should be the case that additional refinement is indicated, steps 16 through 22 provide a mechanism to isolate each region and process it separately. This will ensure that the division first noted is not a composite indication contributed by a number of uniform areas. As an added precaution we only accept the largest segment resulting from complete refinement of each region. Any additional point clusters which have been thresholded are thrown back in the "pot".

Removing the chair from the template at the top of the control stack and smoothing it yields the result of figure 3.35. Again we extract the largest region and push a new template (with associated mask) on top of the control stack. The histograms produced from the newest template indicate the best peak to be for the "I" parameter (figure 3.36). Thresholding on this parameter yields the cushions of the sofa which are masked out of the current template. Close examination of this template (figure 3.37) brings out an interesting phenomenon. Besides the mixed pixels which lightly outline the bodies of the sofa and chair, some heavier lines connecting the right sofa arm and floor areas can be observed. They may have been caused by a shadow effect on that edge of the couch. The point is that regular smoothing will not eliminate all of the superfluous pixels; i.e., we will not be able to obtain the nice clean borders that we would like. In many cases such "debris" would prevent us from getting a separation of regions when one is clearly indicated. If we use the contraction-expansion process described earlier, however, we can get rid of many of these unwanted picture points. An application of the operators has the desired effect, as can be seen in figure 3.38.

We can now execute step 5 of the algorithm and proceed with another level of recursion. The selection of cutoff values for the new templates is quite straightforward and results in the extraction of the highlighted portion of the rug shown in figure 3.39. Subsequent execution of a new level of recursion produces the thresholded points (after smoothing) shown in figure 3.40. The result persuades us that we are going to need the services of the contraction-expansion operation at step 15.2 of the algorithm also. Even then, some undesireable regions which are over the minimum size requirement are left (figure 3.41). They can be eliminated by introducing the following heuristic: unless there are a relatively large number of regions resulting from the threshold operation (more than 5), each candidate must be at least 20% the size of the largest. What we are saying is that we do not want to accept regions very much smaller than the largest one unless they appear to be the result of some larger texture pattern. This is just a reinforcement of the conservative policy concerning the acceptance of regions produced by thresholding.

Masking out the rug leaves the template shown in figure 3.42. Processing the largest portion results in the series of extractions given in figures 3.43 and 3.44. Each grouping shows the object derived from the application of one thresholding. Notice that the thresholding which gives the cluster of points shown in figure 3.44.a produces an overlap into the busy region. Masking out the heavy texture pixels yields the result shown in figure 3.44.b. We should remark at this time that it is possible that a busy region could be <u>completely</u> <u>enclosed</u> by a uniform region resulting from the thresholding operation. If this happens there is no guarantee that the extracted region will be completely separated from the heavy texture area; i.e., it could have been isolated along some dimension that is common to the busy area. Yet, we do not want to simply mask out the busy area. This approach suffers from the objections raised before concerning the imprecise nature of the boundaries of the busy estimation. To insure the best result the heavily textured region is actively extracted. The steps we take are illustrated in figure 3.45 and are an expansion of step 15.2 of the basic algorithm. The appropriate areas of the parameter matrices are averaged to smooth the texture effect. A threshold is then used to obtain the largest resulting cluster. Only the one region is accepted because the picture is basically homogeneous and should not have many busy areas.

At this point in the processing the template given in figure 3.46 is on top of the control stack. The process has succeeded in isolating the high texture portion of the picture. The vase and sofa arm come out as the algorithm proceeds through steps 11 through 11.3. Note that the test for monomodality must be able to recognize a busy region to permit its acceptance. We are finally left with a template from which nothing useful can be extracted (figure 3.47). After popping the stack (step 6.1) the associated mask (figure 3.48) is subtracted from the template now on top of the stack (figure 3.42). The result is shown in figure 3.49.

The procedure continues by extracting the baseboard as the largest remaining region and then the chair arm. Recursion will unwind, each time removing additional objects from the scene. At some point the phase is reached where all the processed segments accumulated from the lower half of the picture (figure 3.50) are masked from the original template put on the stack. This leaves the design as the only major region to process (figure 3.51). Good color separation makes this an easy task. Final removal of the design and plug on the wall will yield a template empty of all processable regions (figure 3.52). This remainder is saved (step 9.3) in case any interpretable region has been overlooked because of size. Higher level knowledge would have to

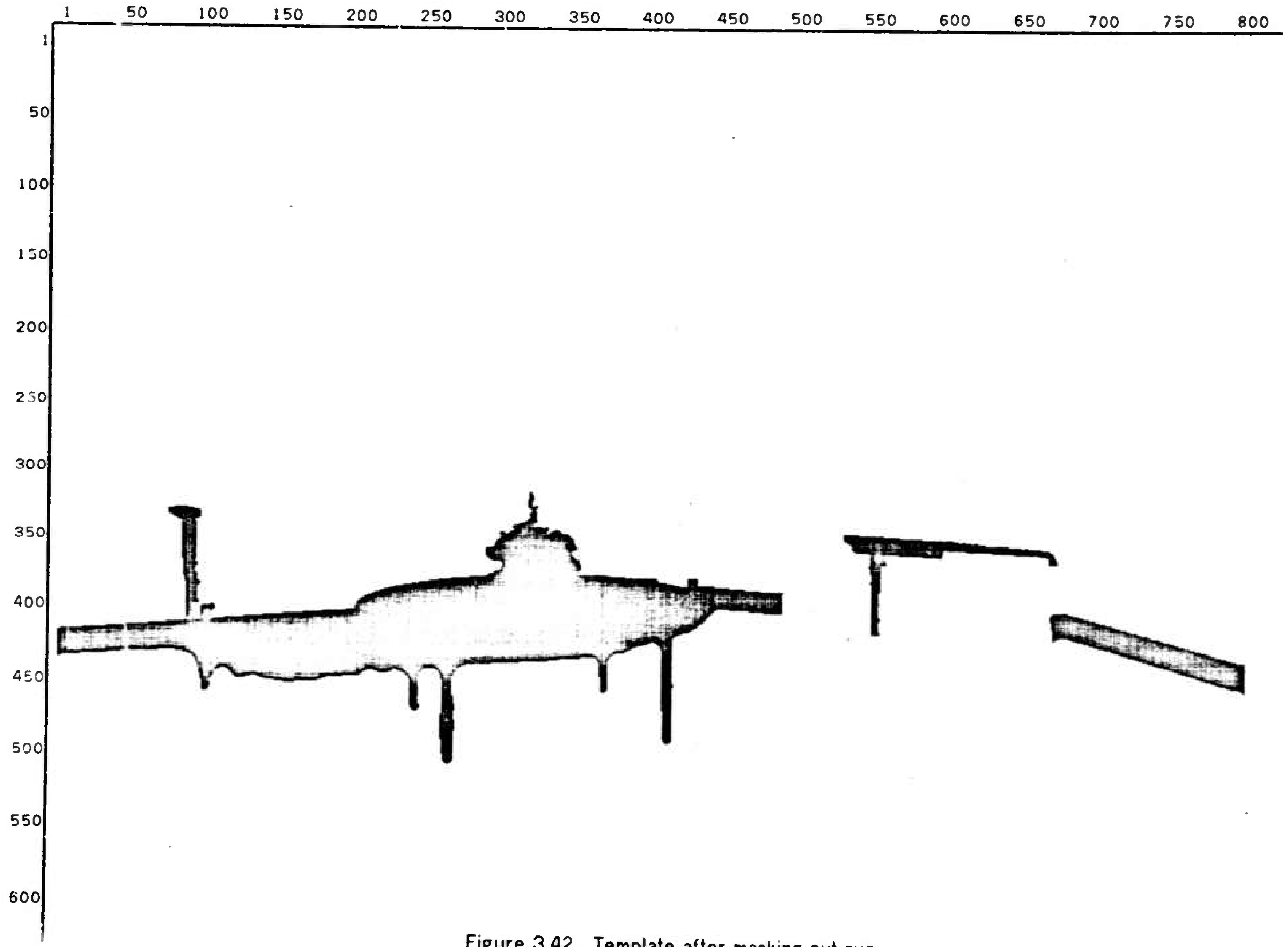


Figure 3.42. Template after masking out rug.

Figure 3.43. Extractions resulting from processing of largest segment in figure 3.42.

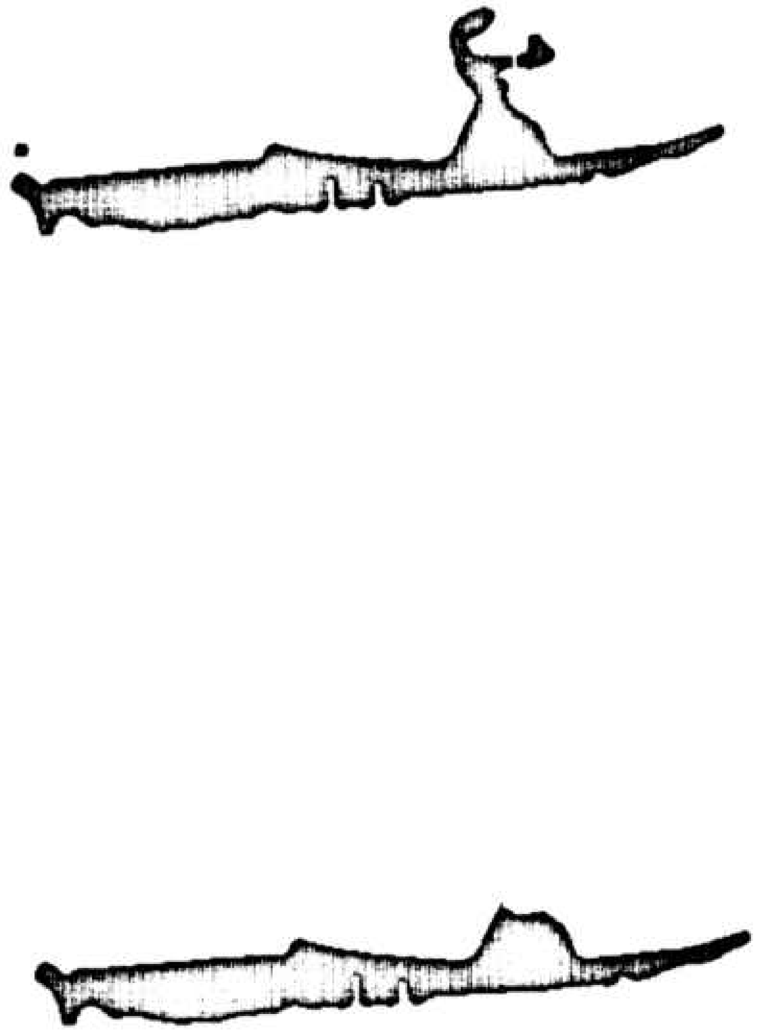


Figure 3.44. Extractions resulting from processing of largest segment in figure 3.42.

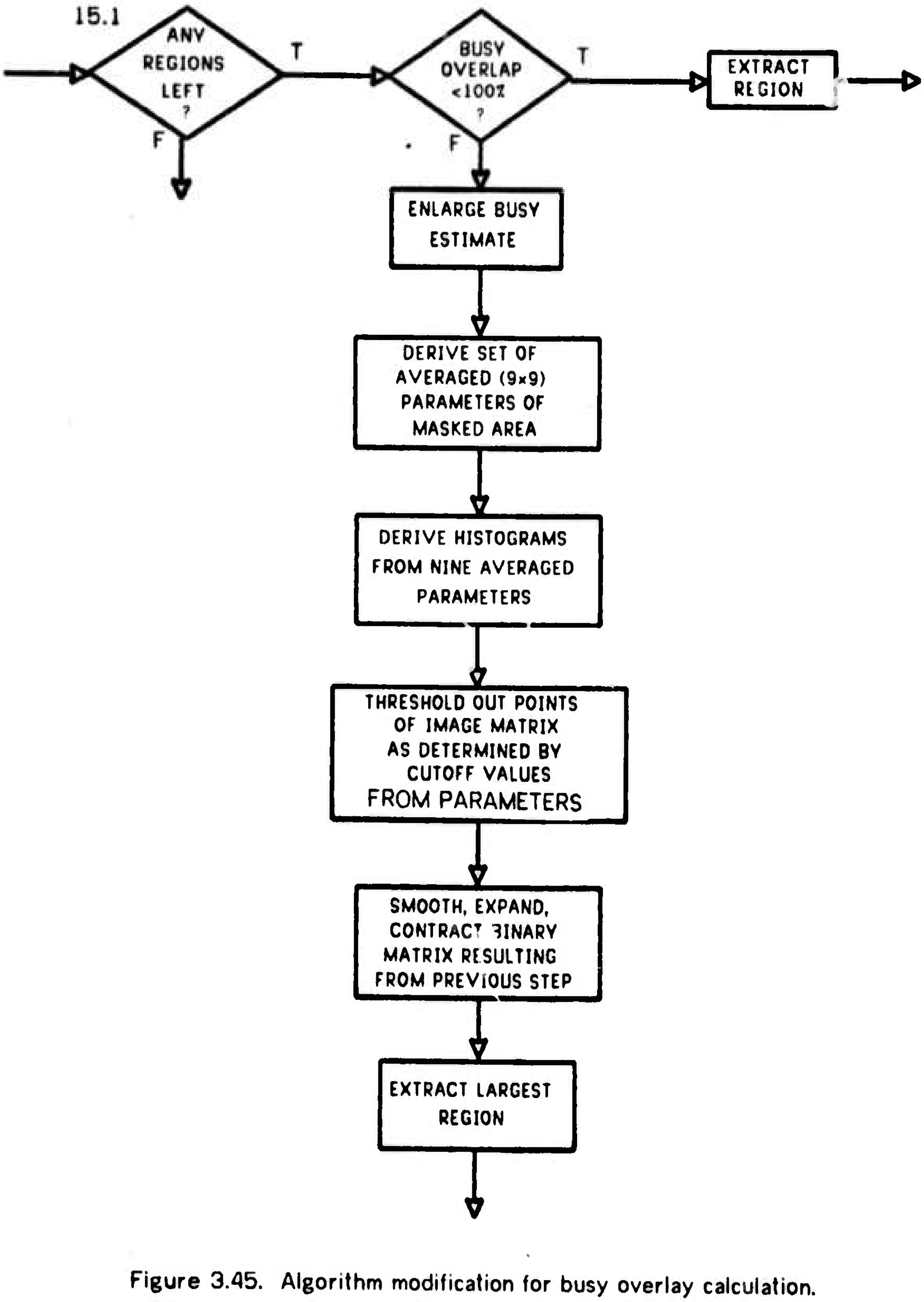

15.1
ANY REGIONS LEFT ?
T
F
BUSY OVERLAP <100% ?
T
F
EXTRACT REGION
ENLARGE BUSY ESTIMATE
DERIVE SET OF AVERAGED (9×9) PARAMETERS OF MASKED AREA
DERIVE HISTOGRAMS FROM NINE AVERAGED PARAMETERS
THRESHOLD OUT POINTS OF IMAGE MATRIX AS DETERMINED BY CUTOFF VALUES FROM PARAMETERS
SMOOTH, EXPAND, CONTRAC⁊ ƎINARY MATRIX RESULTING FROM PREVIOUS STEP
EXTRACT LARGEST REGION

Figure 3.45. Algorithm modification for busy overlay calculation.

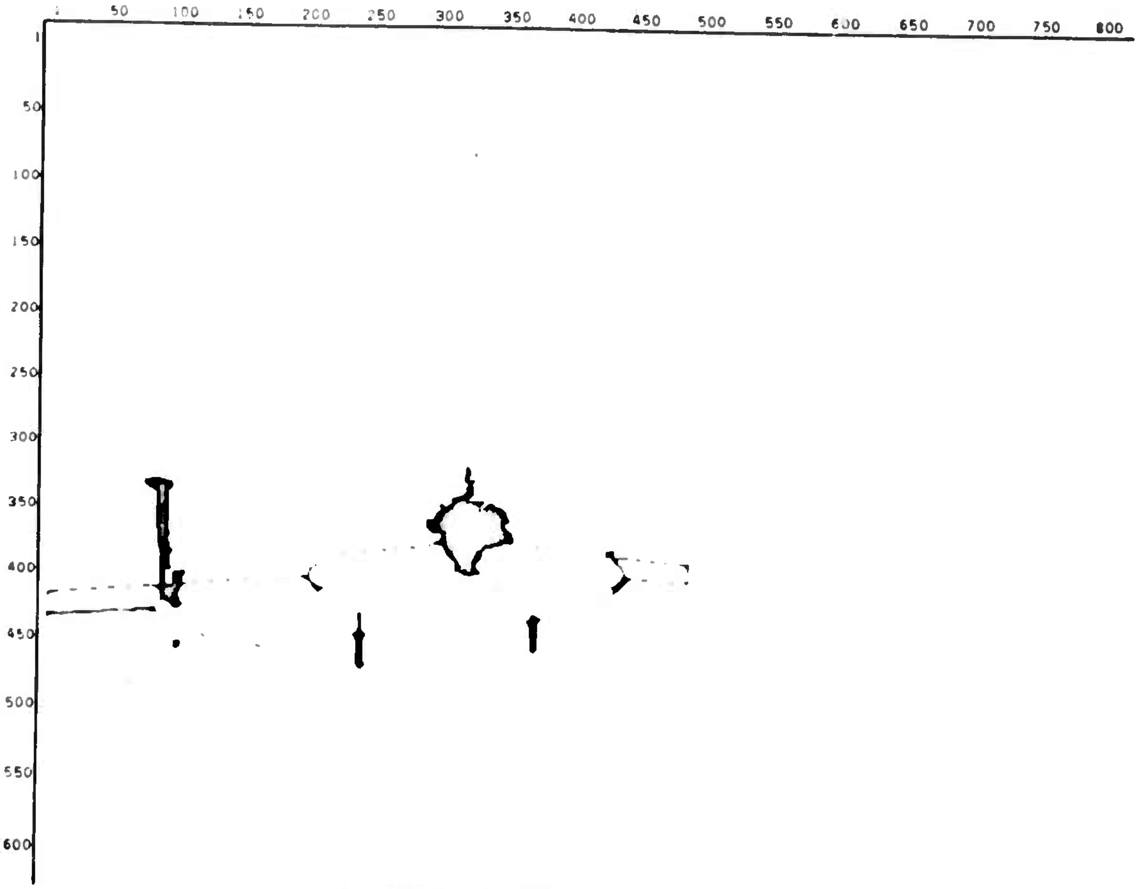


Figure 3.46. Template for a given level of the process.

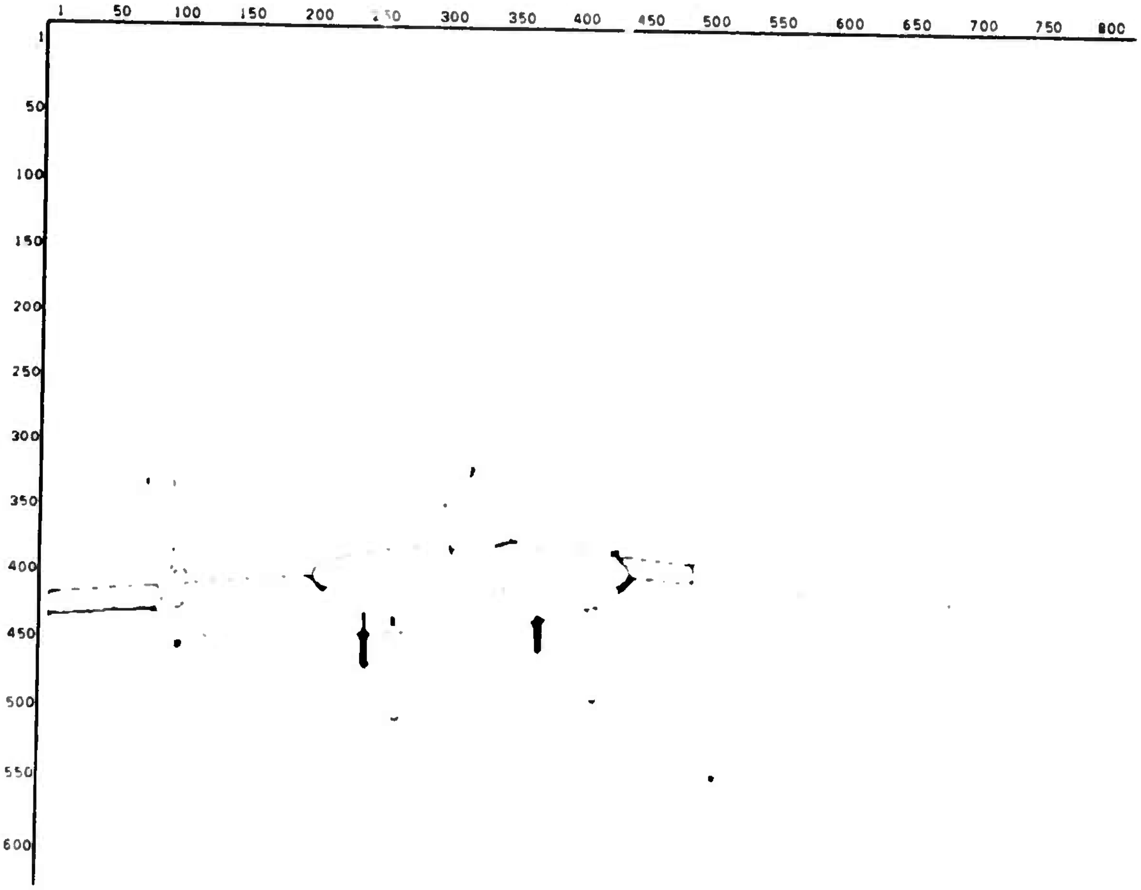


Figure 3.47. Template from figure 3.46 after masking out vase and arm.

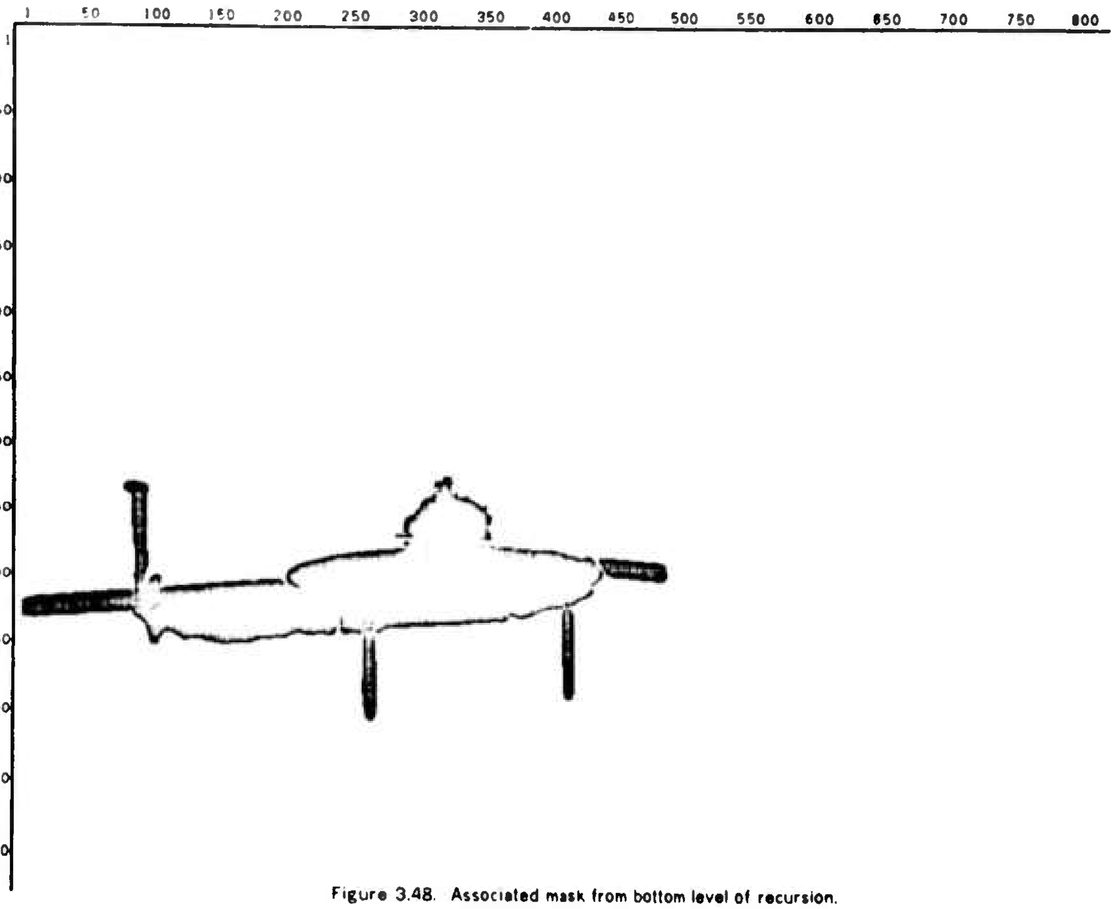


Figure 3.48. Associated mask from bottom level of recursion.

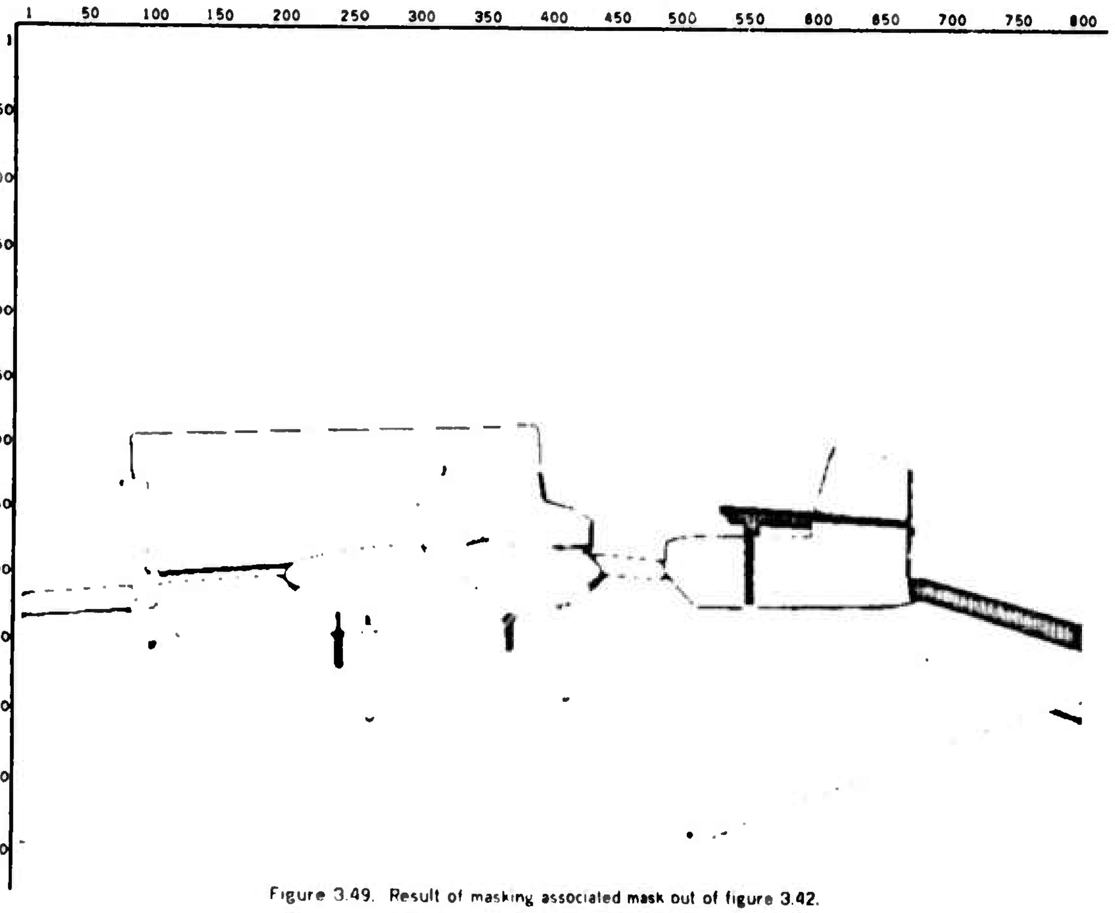


Figure 3.49. Result of masking associated mask out of figure 3.42.

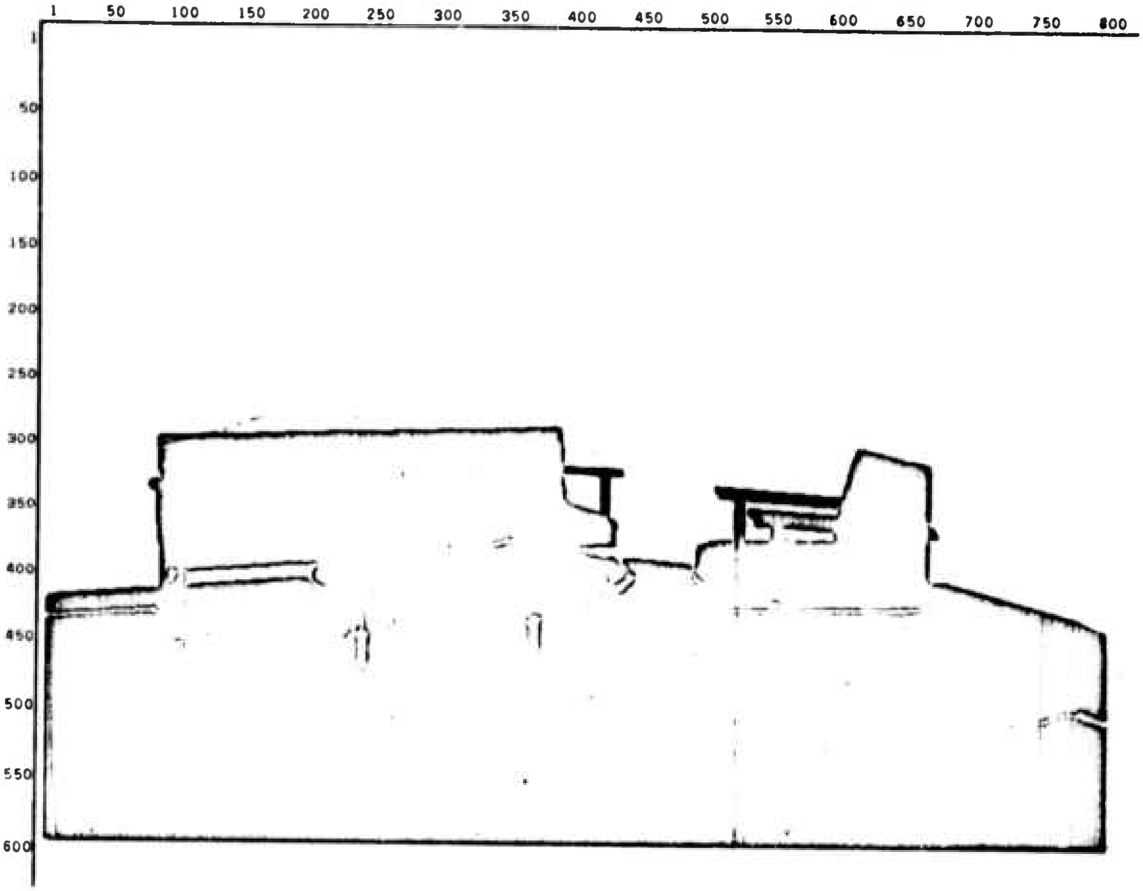


Figure 3.50. Associated mask after processing lower half of the picture.

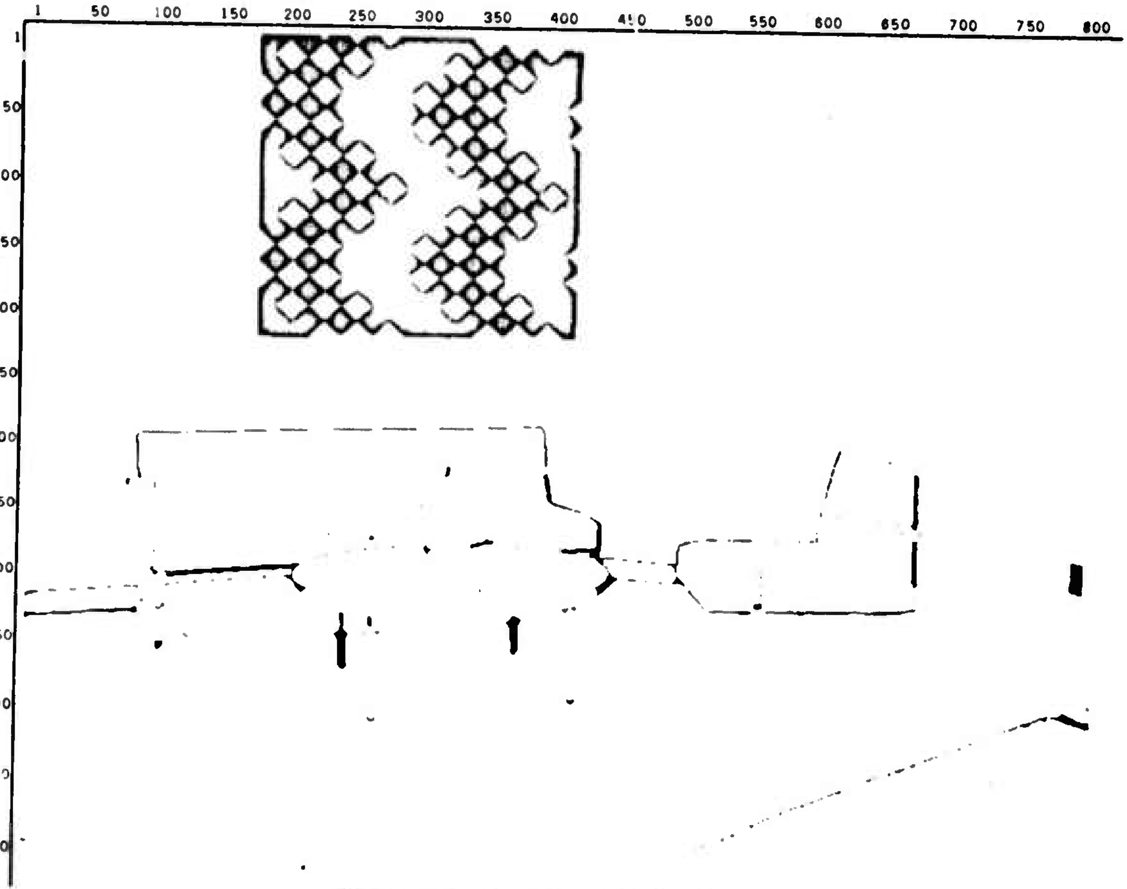


Figure 3.51. Result of using mask in figure 3.50.

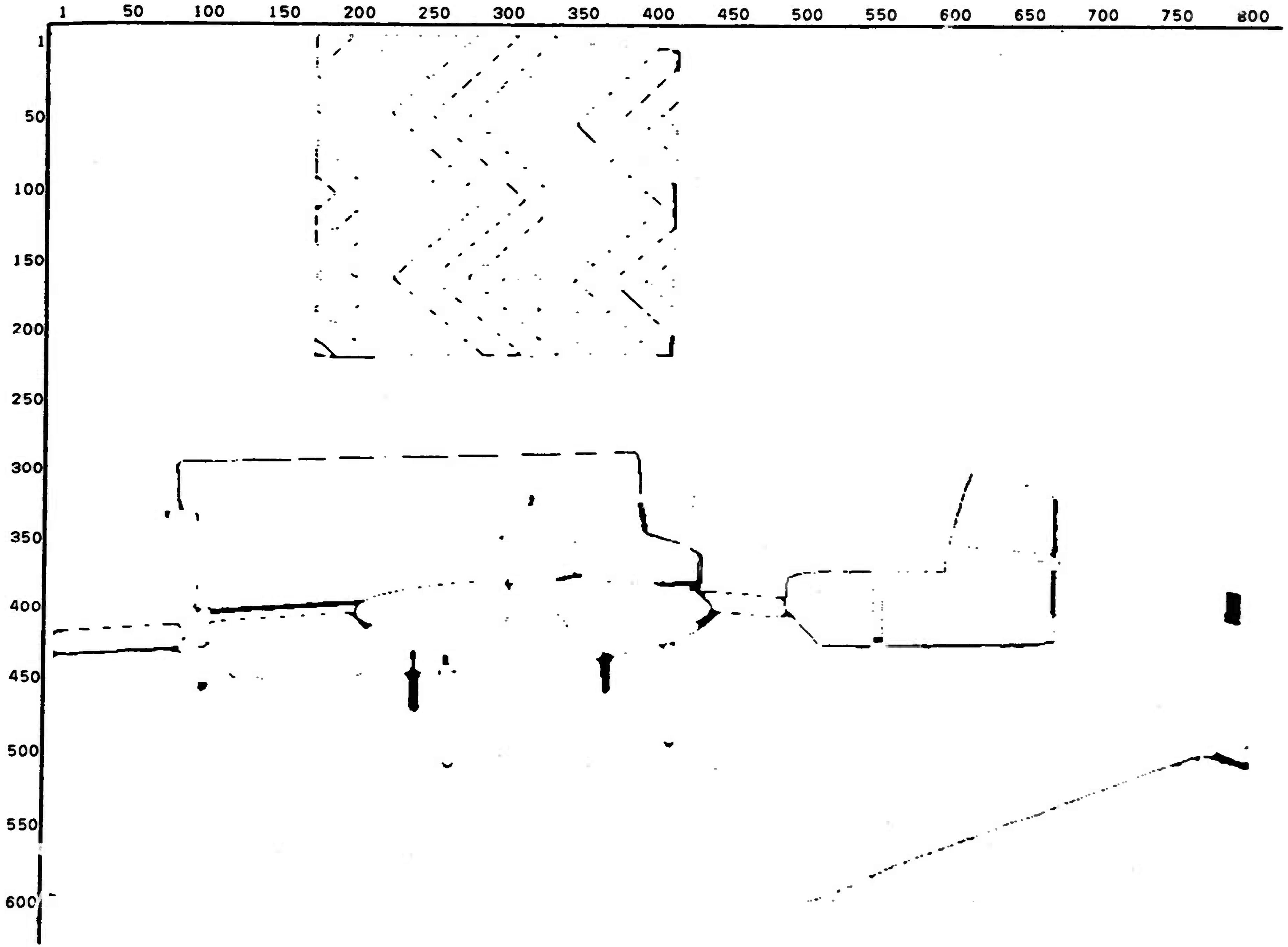


Figure 3.52. Final template of the process.

direct a search in this area. The final partitioning of the scene is shown the results section of this chapter.

We have taken pains to present the operation of the basic algorithm in some detail. Many figures and diagrams have been used to illustrate the steps and some of the modifications required to induce the kind of results that we want and believe possible. We now want to apply the existing algorithm to the house scene (figure 2.4.b) which possesses richness of color but also contains a great deal more texture than the preceding image. Proceeding as before in the preprocessing phase, a heavy texture (busy) matrix is extracted (figure 3.53). This matrix cannot be utilized in the simple masking routine that was employed above because of the error in boundaries which must result. For example, the lower right window frame and front center window are shown as busy. They can, however, be thresholded out quite nicely. The types of errors that have occurred in the current busy estimation are the same as for the previous scene but they occur to a much greater extent. Narrow regions can no longer be eliminated effectively; they are connected to larger ones. A density test can no longer be employed because the heavy texture is spread throughout the scene. An alternate scheme must be utilized to direct the progress of analysis.

The approach we have chosen involves analysis of the parameters of the busy and non-busy (the complement of the busy matrix) portions of the picture. Histograms for both these areas are shown in figures 3.54 and 3.55 (figure 3.56 shows the non-busy area with the sky removed so that the remaining peaks can be better observed). Note the similar peaks for both of the hue parameters in the 50 to 140 range. As noted earlier, these points cover the grass (homogeneous) and the shrubs (textured) areas of the image. These histograms will be used to direct later analysis. In order for a peak to qualify for thresholding at step 13, a similar mode must be displayed for the same range in the non-busy histogram. If there is no corresponding high point in the busy histograms we allow the thresholding to proceed in the normal way. If there is a corresponding high point this means some feature is common to non-busy and busy portions of the picture. As a consequence we will be less critical in permitting further refinement of a thresholded matrix. Any sign of further discontinuity might serve to eliminate texture clusters. As a final precaution any candidate region for extraction must meet a more stringent size requirement and no more than 20% of its area can overlap any busy area. These requirements do not apply to small white or black areas of the picture. There is one additional way in which we use the busy mask to refine the textured area of the picture; this will be described later.

We make no attempt to describe in detail the entire decomposition of the house scene, but will illustrate some interesting features of its analysis. The first thing that is noticed in the histogram for the entire scene (figure 3.57) is the indication of small white areas (intensity, 210 to 240). Thresholding on this parameter gives the regions shown in figure 3.58. They segment out quite nicely and prove to be of further use in partially separating some large areas at a later stage of analysis (figure 3.59).

The next interesting occurrence in the processing comes when faced with the histograms of figure 3.60. The obvious choice of cutoff values is 50 and 130 for the hue. This is the peak that was so prominant in both the non-busy and busy histogram sets. Thresholding on this paramater yields the point matrix shown in figure 3.61. As

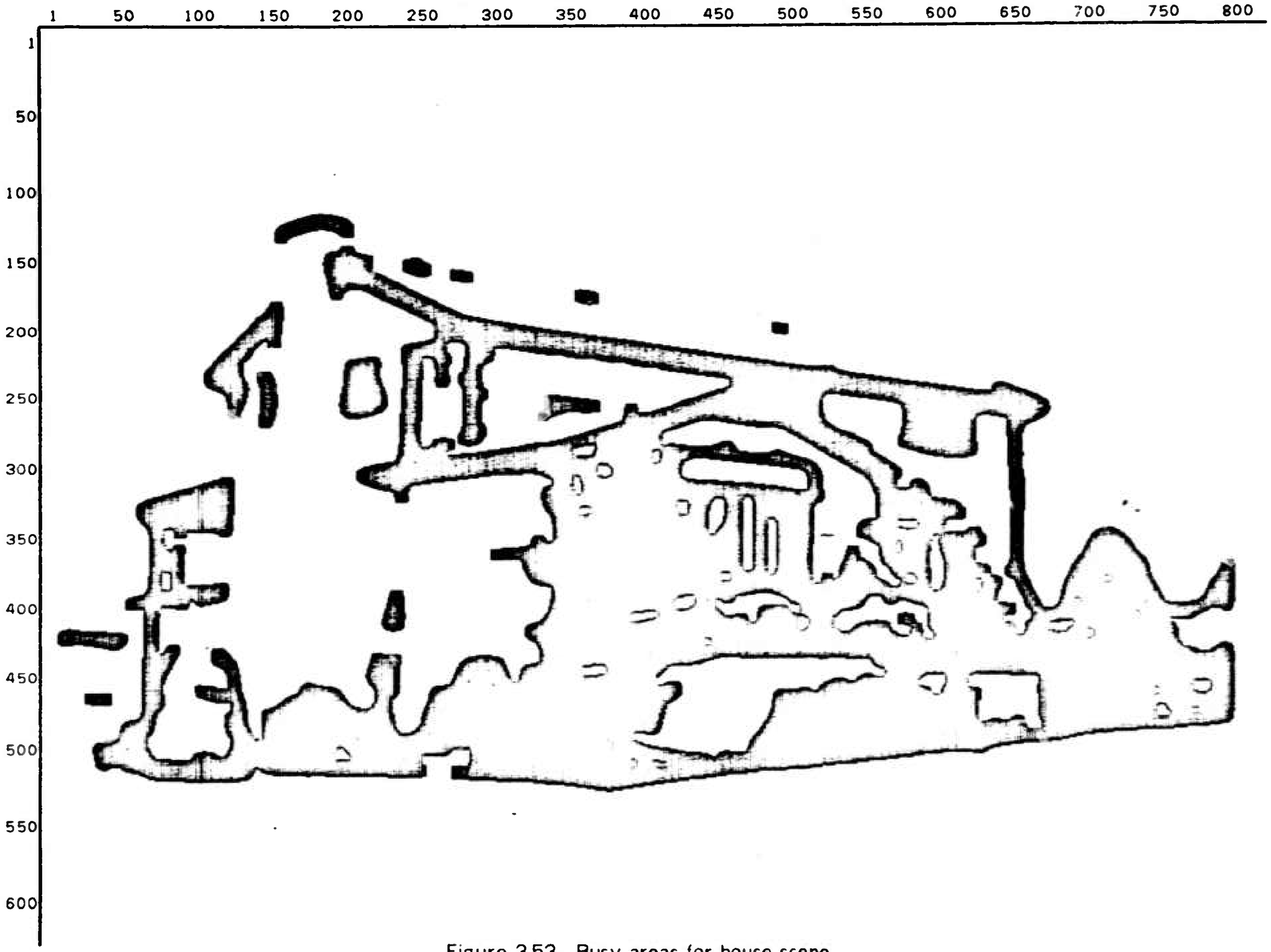


Figure 3.53. Busy areas for house scene.

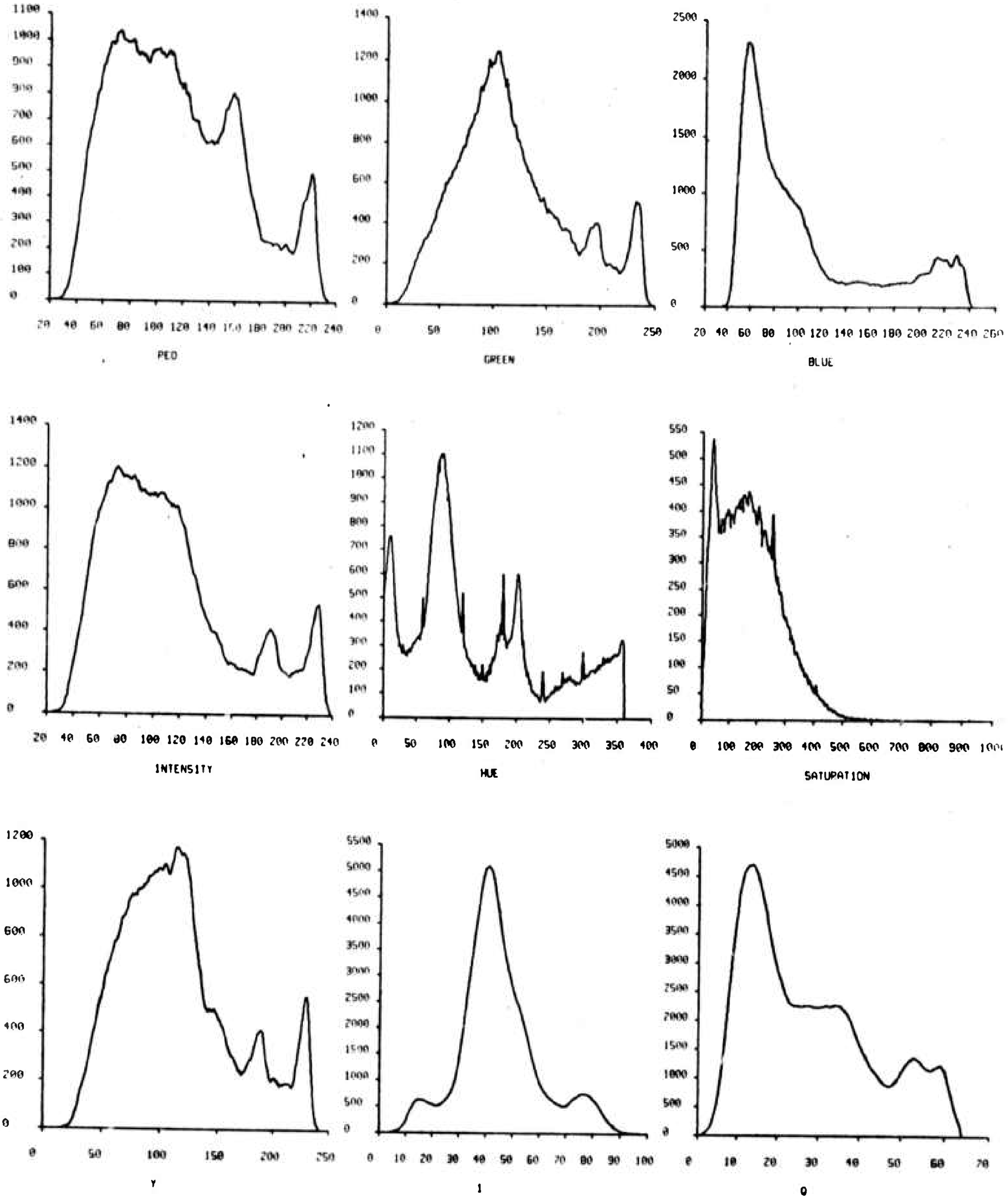


Figure 3.54. Nine parameter histograms for the busy areas of the house scene.

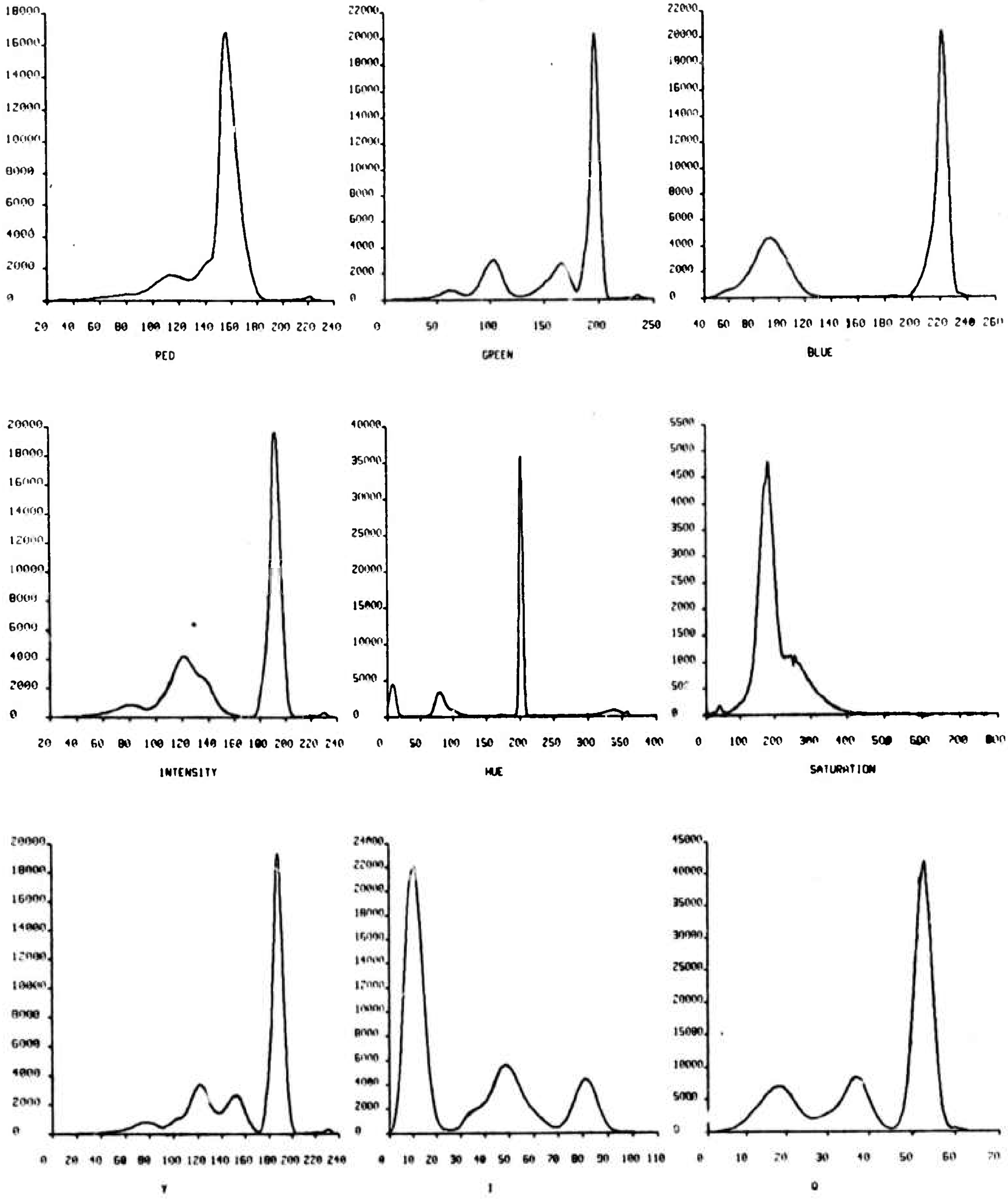


Figure 3.55. Nine parameter histograms for the non-busy areas of the house scene.

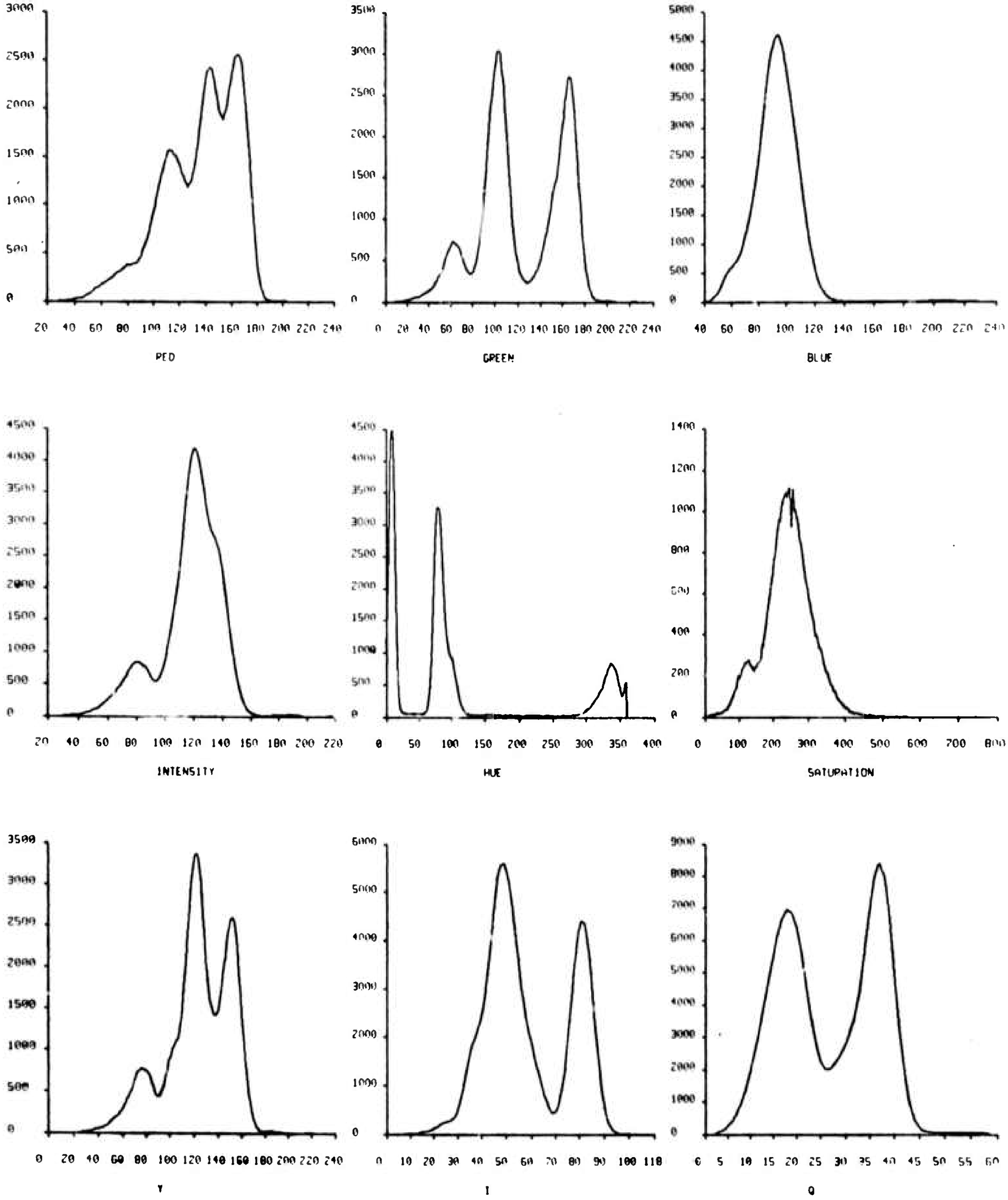


Figure 3.56. Nine parameter histograms for the non-busy areas of the house scene (sky removed).

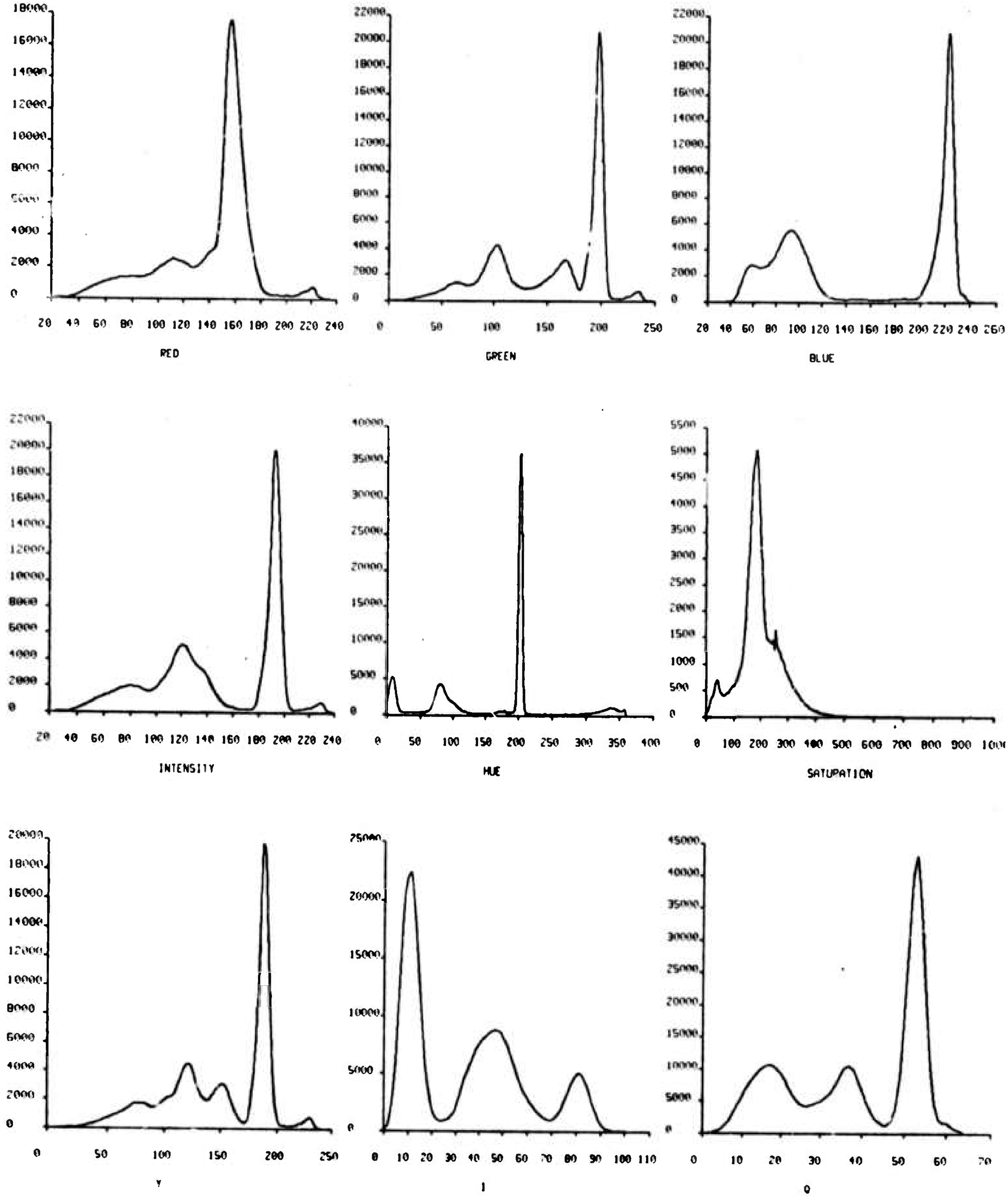


Figure 3.57. Nine parameter histograms for the house scene.

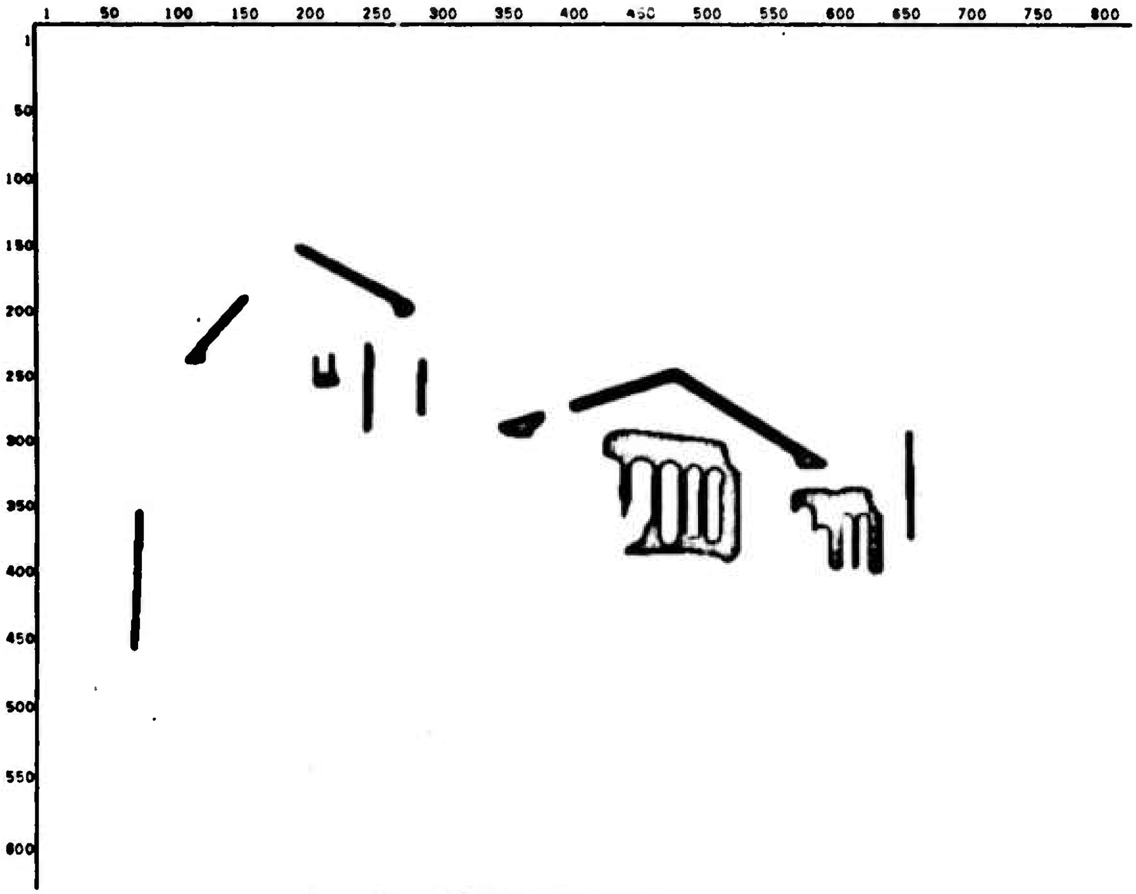


Figure 3.58. White areas of the house scene.

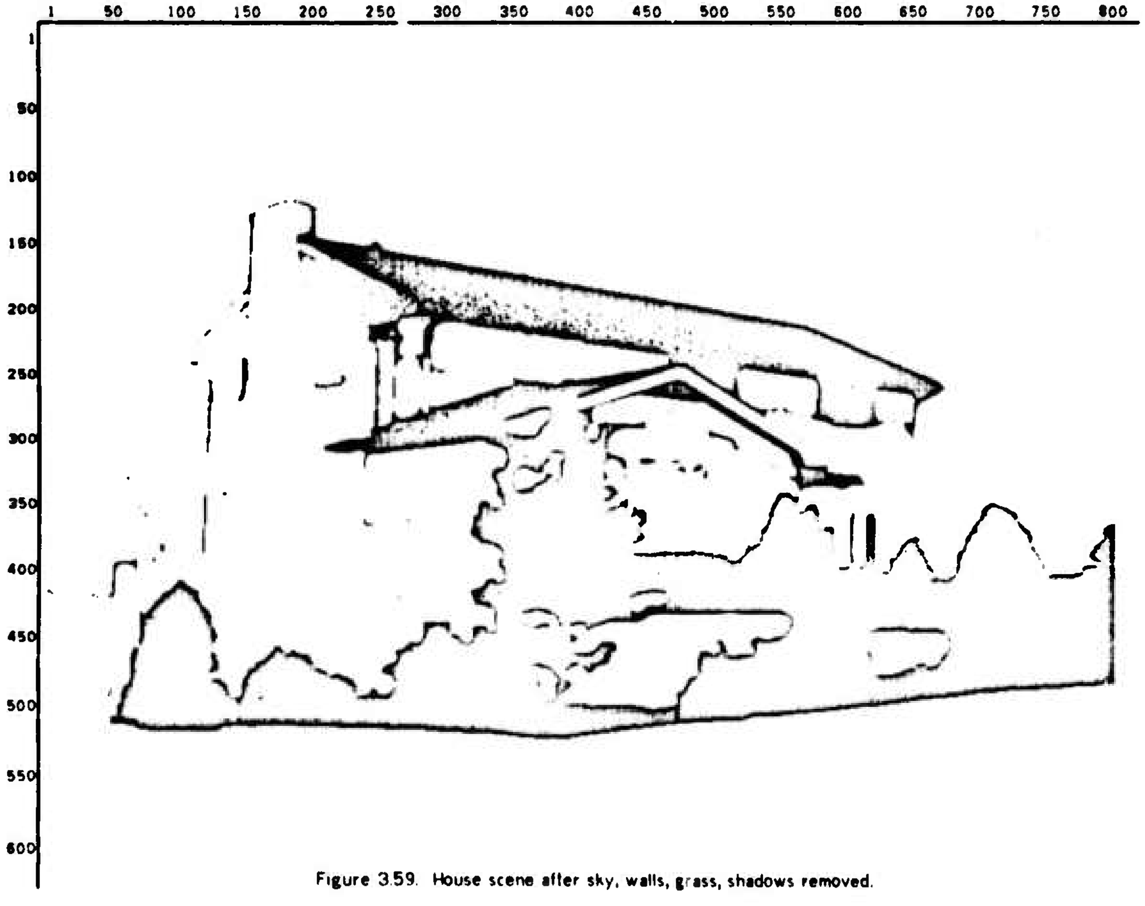


Figure 3.59. House scene after sky, walls, grass, shadows removed.

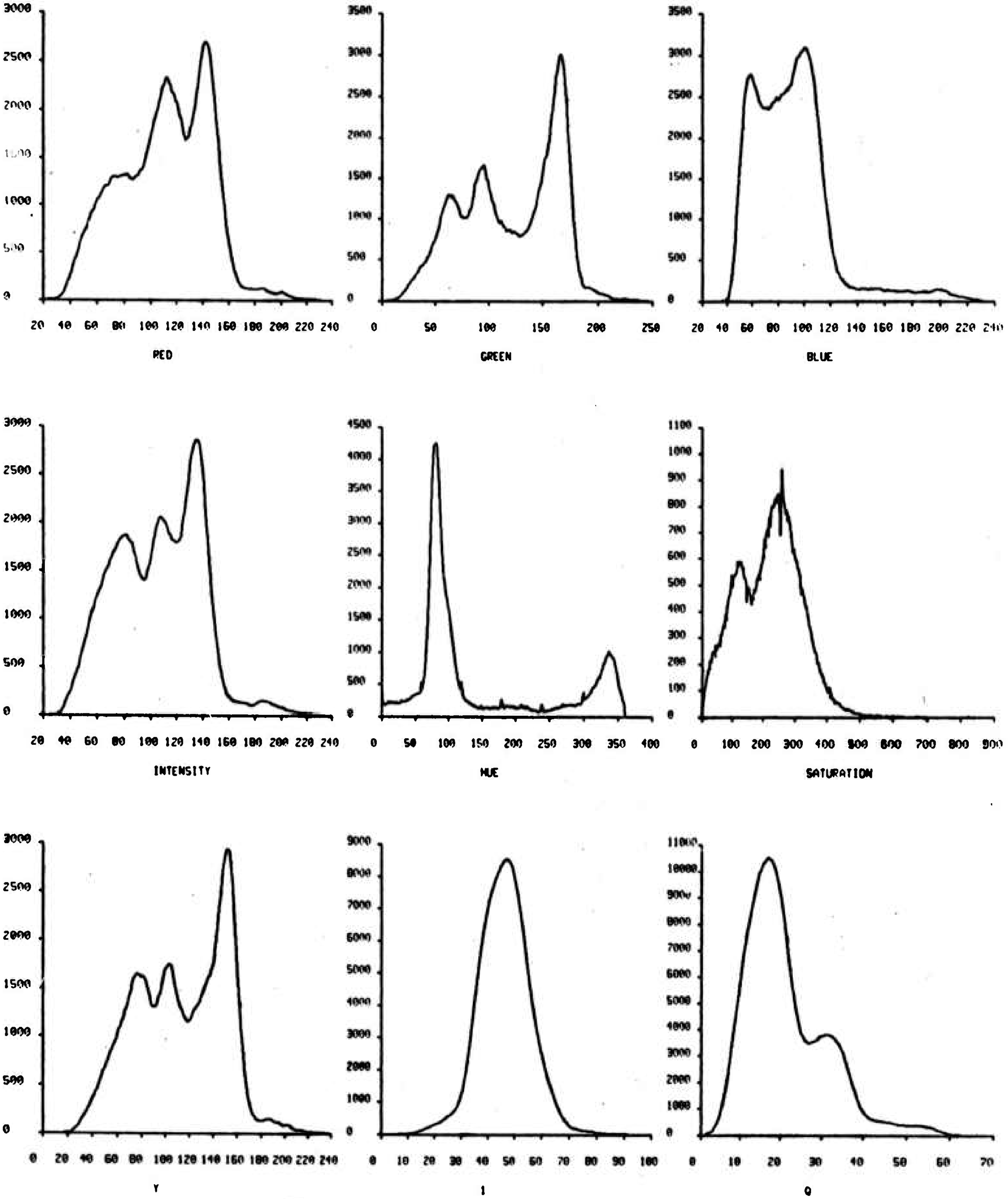


Figure 3.60. Nine parameter histograms for some level of the process.

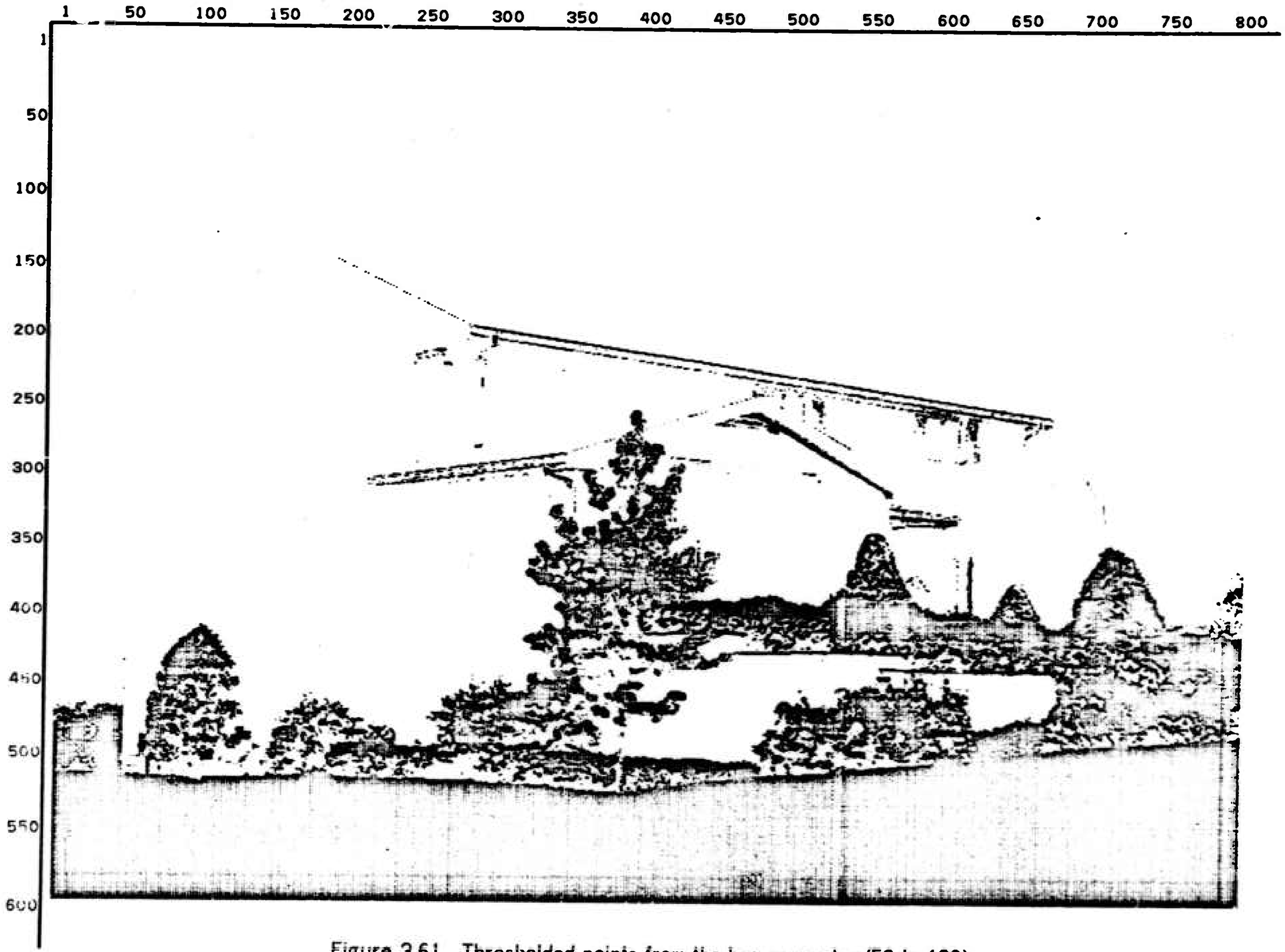


Figure 3.61. Thresholded points from the hue parameter (50 to 130).

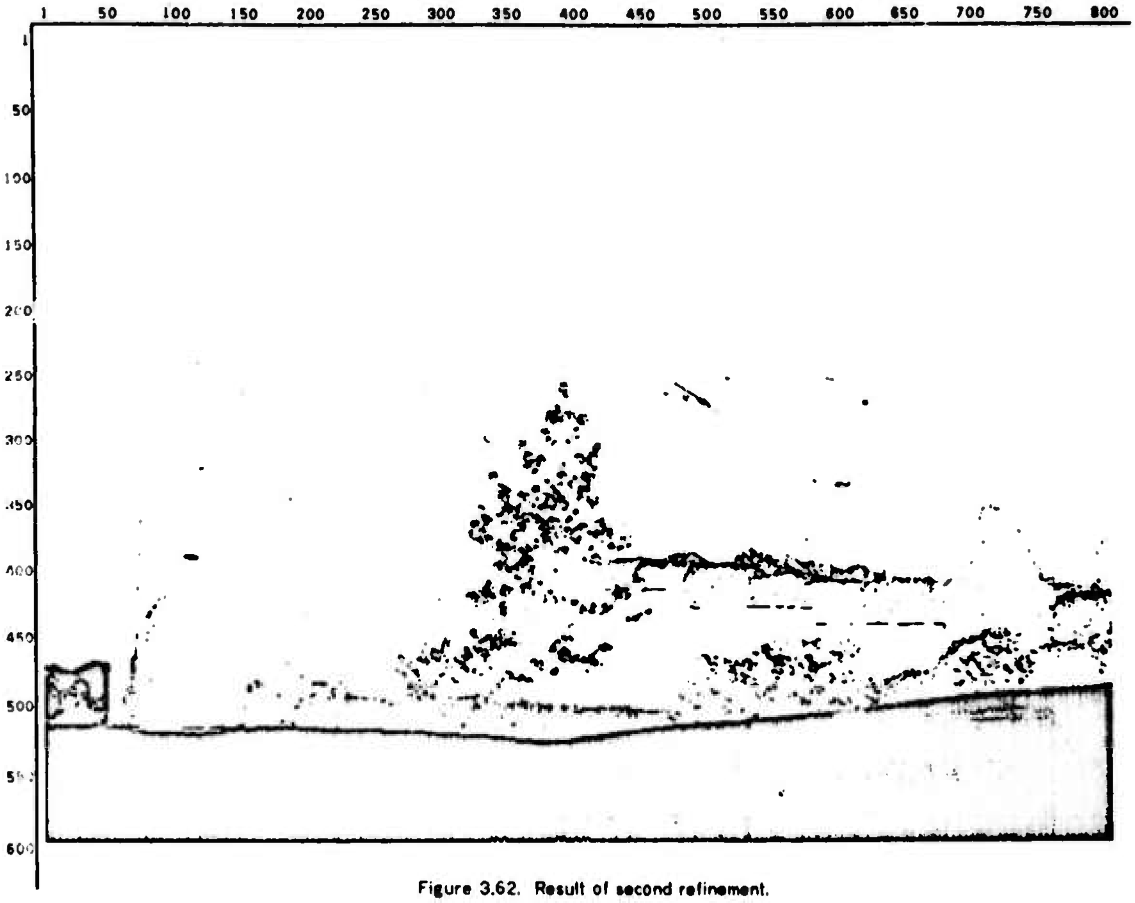


Figure 3.62. Result of second refinement.

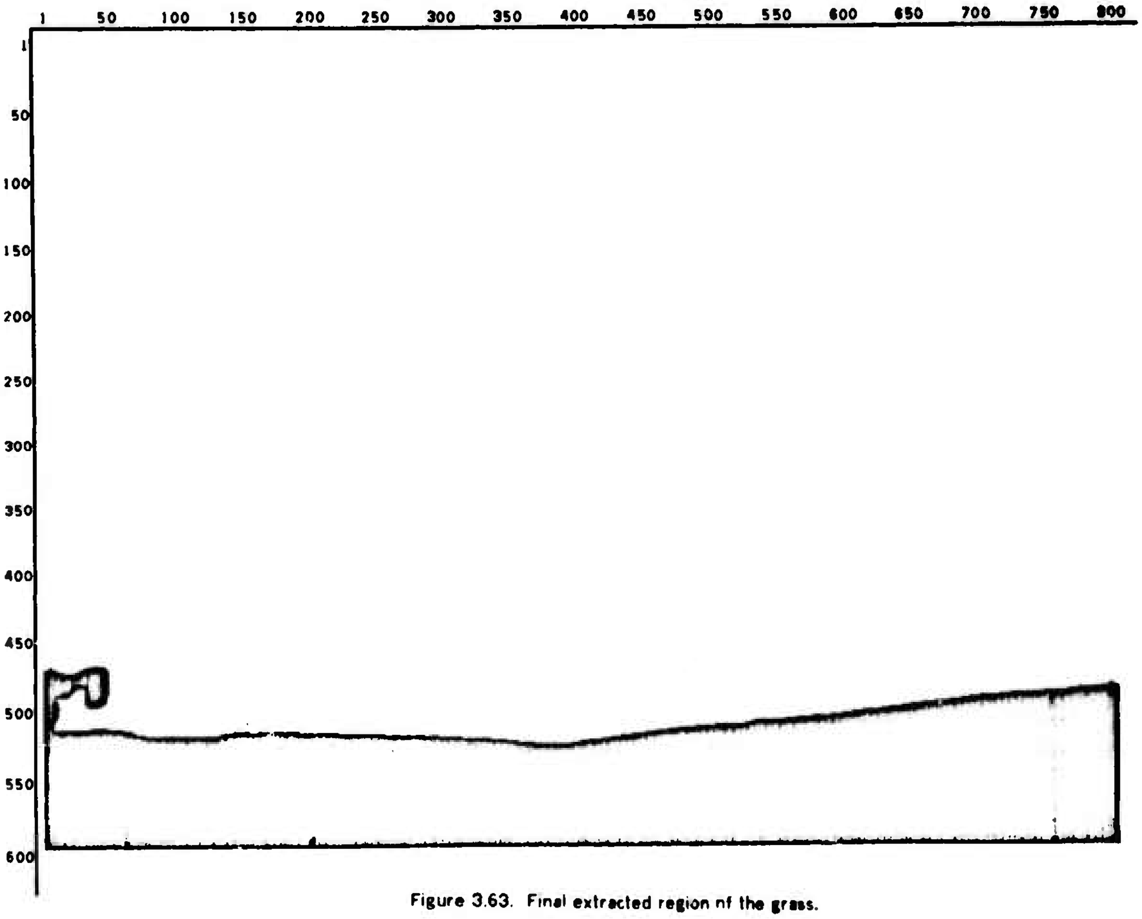


Figure 3.63. Final extracted region nf the grass.

expected a large portion of the heavy texture area has surfaced. An examination of a second histogram derivation shows that, although the matrix is uniform in hue, there exists a discontinuity in the green parameter histogram at about 125. A refinement based on this evidence produces the matrix of figure 3.62. A contraction-expansion application and extraction finally gives the results shown in figure 3.63. Checking with the business matrix proves it to be non-busy.

The next occurrence of special note takes place when the procedure has reached the stage of analysis which has led to the template shown in figure 3.64 being on the top of the control stack. The histograms for the template (figure 3.65) show no decisive peaks for cutoff. The peak to the left in the hue would ordinarily be a candidate for thesholding, but it is in that troublesome area of the hue busy parameter. We have obtained one extraction along this dimension in the indicated range so no more can be expected. Any other possible portions of the curve that might suggest a cutoff are suppressed by the comparatively overwhelming amount of data provided by the heavily textured area. If the area under current consideration were non-busy the normal course of events would remove a large part of it by thresholding on the basis of cutoffs provided by the histograms. If we can eliminate some of the interference we may yet be able to obtain useful information. Proceeding on this assumption the current template is masked with the busy matrix to get the temporary binary picture shown in figure 3.66. Resulting point clusters are not very accurate borderwise but the matrix does yield histograms of some utility (figure 3.67). Now we can observe a reasonable peak in the graphs to the extreme right of the hue graph. We use the limits thus obtained to threshold the data of the original template (figure 3.64). The final result is worth the effort (figure 3.68).

The last point that we have to make concerning this scene concerns the template shown in figure 3.69 (the result of masking out the processed segments obtained in the last paragraph). There is an eave which "hangs" off the tree to its left. We would like to separate this object, and it is only by sheer chance that we do so. The contraction-expansion operation accomplishes the desired split. If we had been required to depend on histograms providing useful information we would have been out of luck (figure 3.70). No useful peaks are indicated at all in this set of histograms. The masking ploy used last time will not work again because the region in question is also textured. One might not accept this at first, because eaves are thought of as white homogeneous surfaces. Close examination however, will show that there are a number of longitudinal lines caused by moldings and shadow effects. Since the eave is in shadow, intensity is not high enough to provide values in the white range of the histograms. There is another difficulty caused by shadows in this case. Since the object is not receiving direct sunlight and a white surface is highly reflective, the eave takes on some of the properties of the surrounding greenery and structure. The hue expecially ranges from green to red in color for various portions of the surface. Averaging operations and further threshold application might be useful, but we would be reluctant to employ them without some cause. If the eaves had not been separated by chance, the resulting region would still have been one that conformed to the busy area of the picture. It would have been obtained by elimination of the surrounding homogeneous regions. Any decision for further refinement would have to be based on contextual information and come from higher level knowledge sources. As low level processing is concerned we choose to stop after business indicates complete isolation.

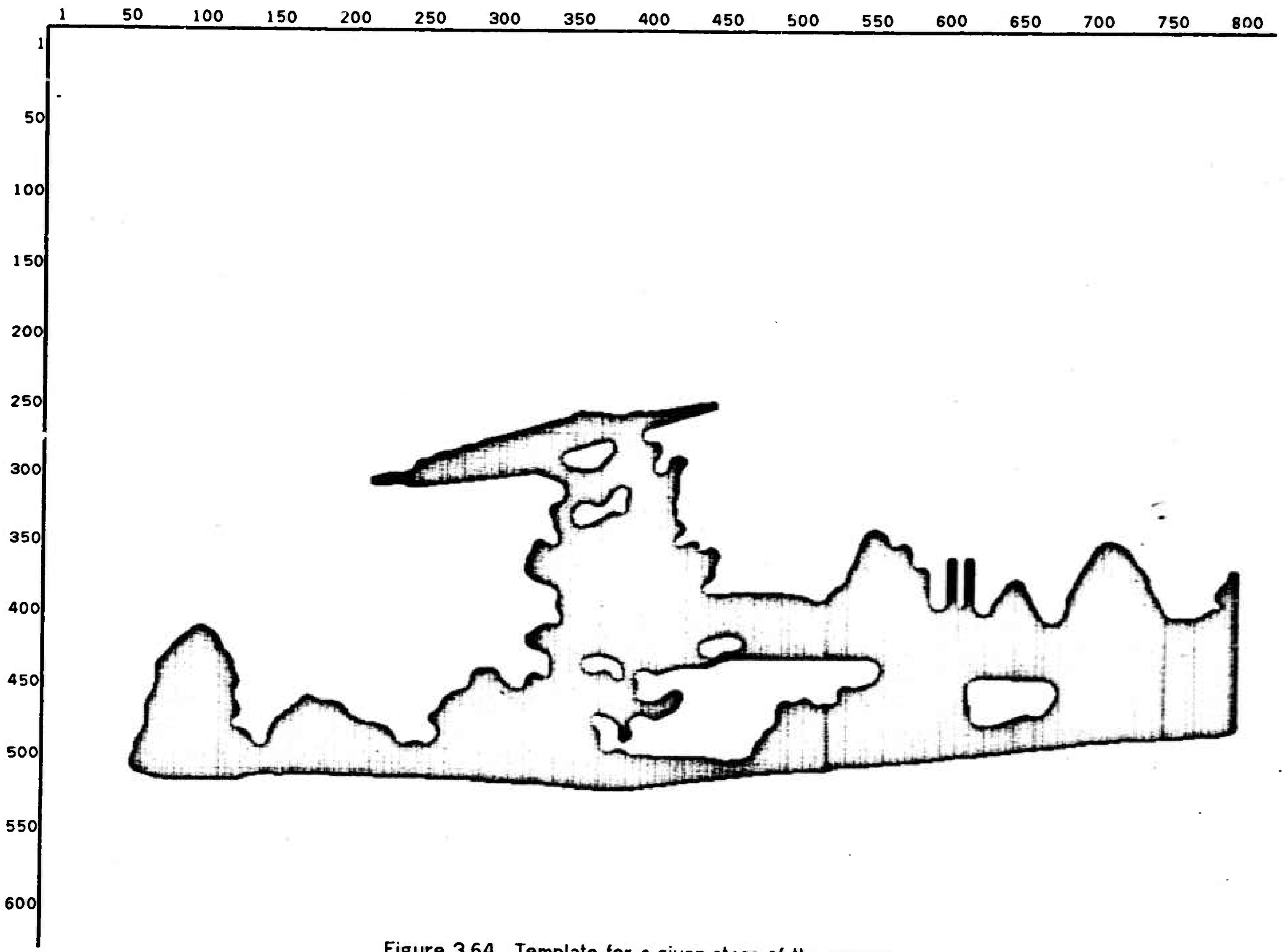


Figure 3.64. Template for a given stage of the process.

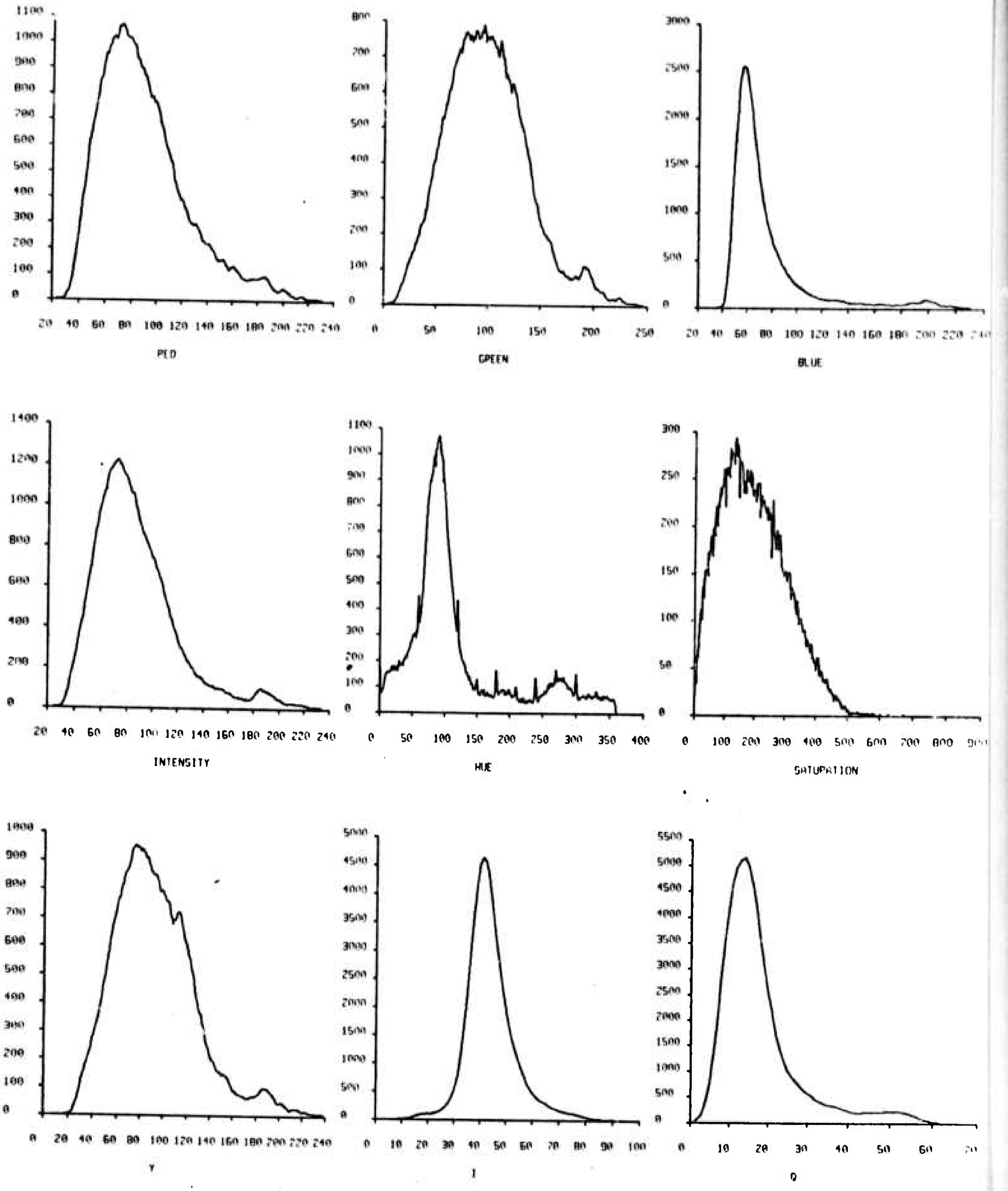


Figure 3.65. Nine parameter histograms for template of figure 3.64.

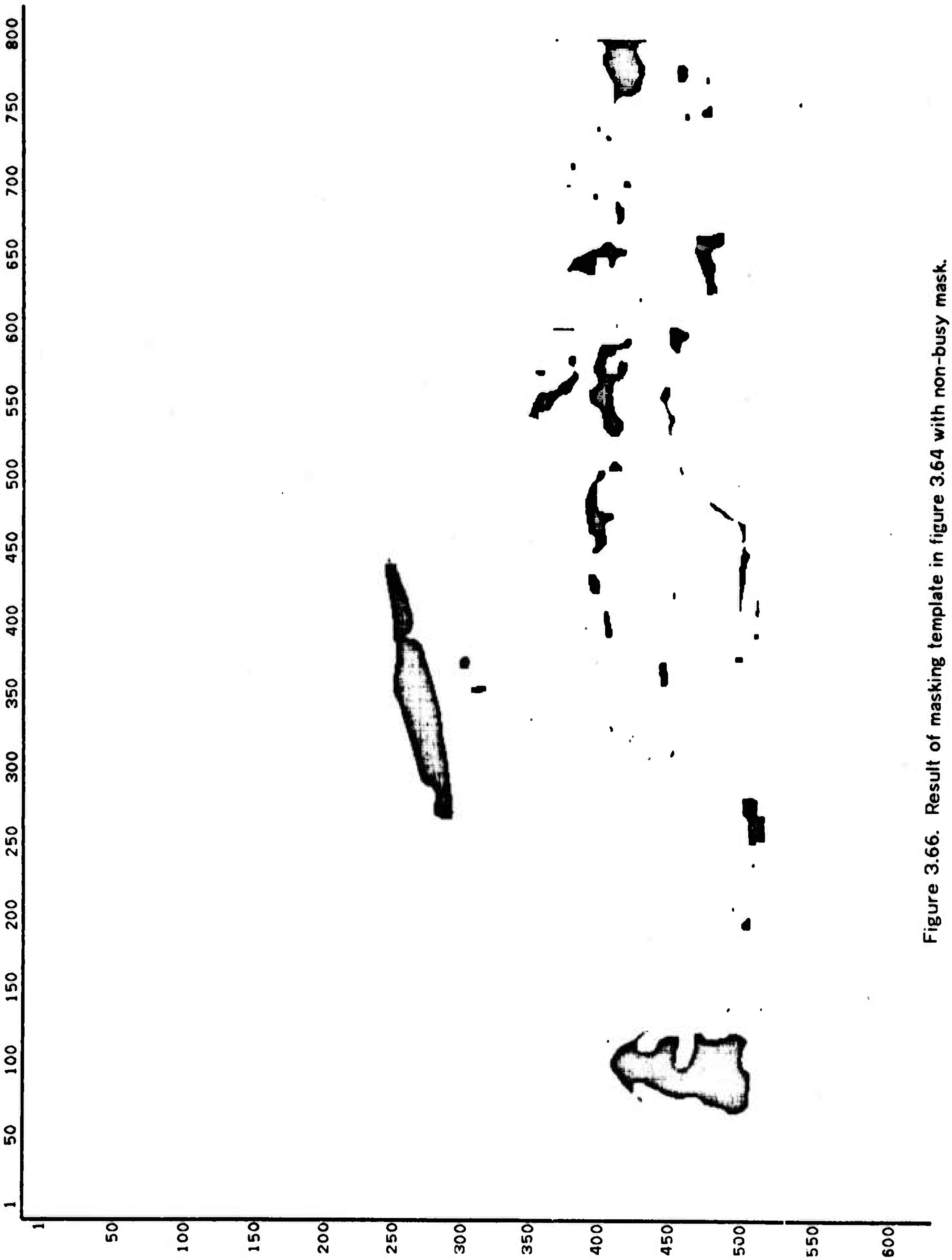


Figure 3.66. Result of masking template in figure 3.64 with non-busy mask.

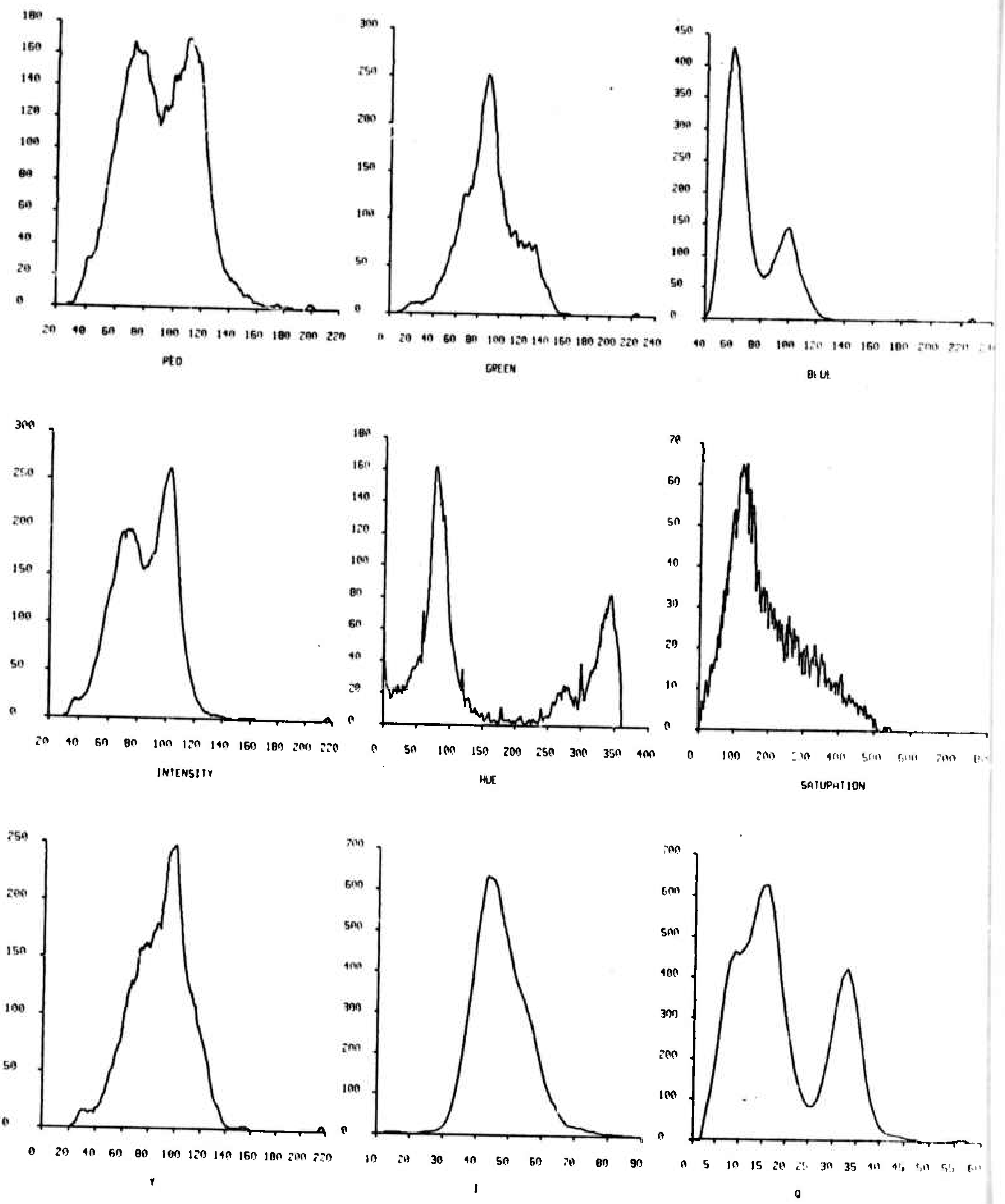


Figure 3.67. Nine parameter histograms for mask of figure 3.66.

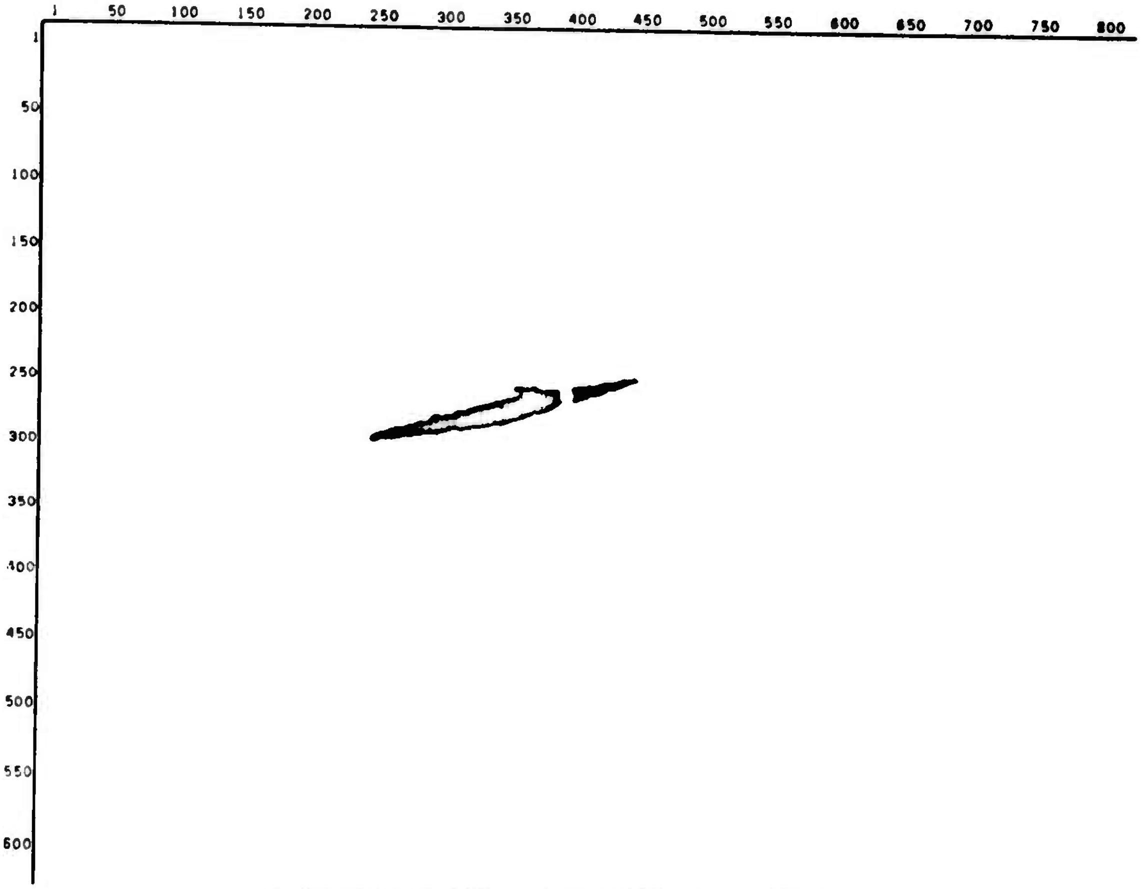


Figure 3.68 Result of extracting on nue limits provided by figure 3.67.

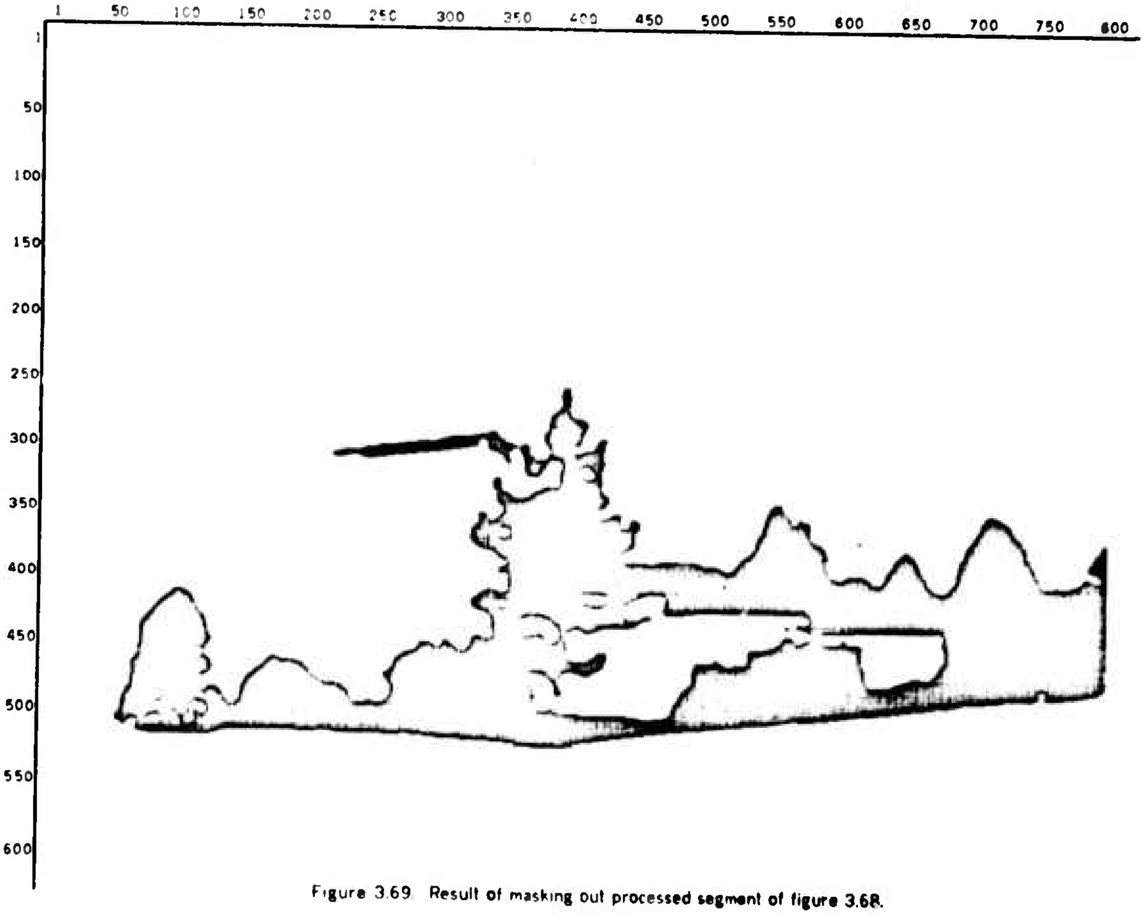


Figure 3.69 Result of masking out processed segment of figure 3.68.

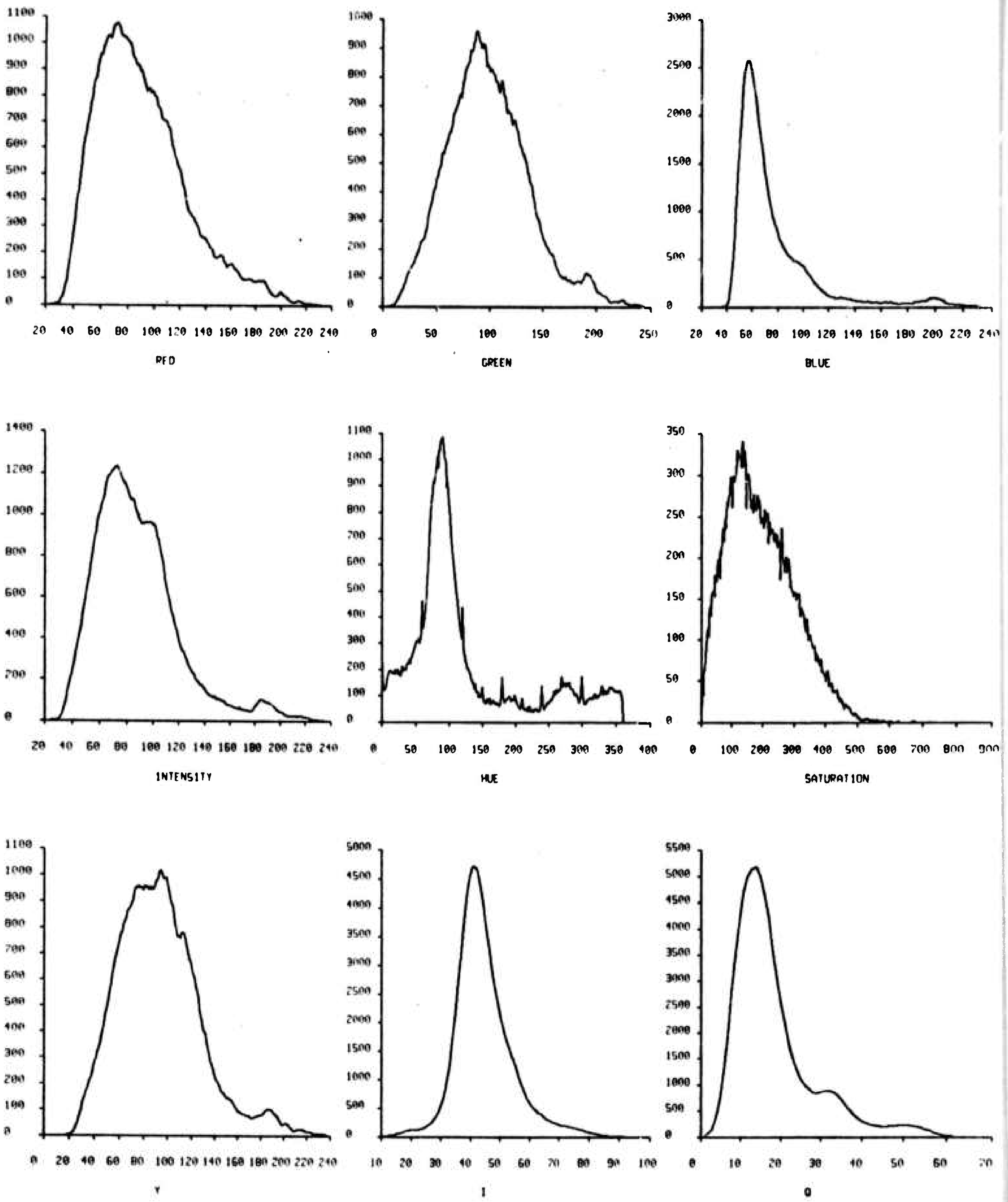


Figure 3.70. Nine parameter histograms of figure 3.69.

The final picture to provide input to the formation of the algorithm was the skyline scene (figure 2.4.f). Again, we will not attempt to cover the complete decomposition of the scene. We will point out those issues, arising during analysis, which gave particular trouble and required additions to, or modifications of, the algorithms.

The standard preprocessing gave rise to the busy matrix shown in figure 3.71. As can be seen, the scene in question is quite heavily textured. The histograms of the scene provided little information (figure 3.72). There is one well-defined mode, but no indications of anything else that might prove useful later. A first thresholding operation extracts the sky and leaves the template shown in figure 3.73 for further processing. A subsequent derivation of histograms for this area (figure 3.74) supports our fears of lack of feature indications. The only curve which shows some signs of separation is the one for the blue parameter. The conditions for its acceptance are way below standard and are not seriously considered. On the other hand, we cannot possibly settle for the current partition without admitting defeat for the procedure, at least for the given scene. What must be done at this point is force a subdivision of the picture into sections that can be handled by the thresholding operation. This must be accomplished in such a way that some integrity of structure is retained in the process.

We already know one way to remove heavily textured parts of the scene. Perhaps the business matrix can be utilized to get additional segmentation. We might be able to locate and isolate the homogeneous areas of the scene. To accomplish this end we return to a consideration of the business matrix obtained in the preprocessing phase. A very small area of the scene is classed as homogeneous if no more than one edge runs through it. Thus, an upper threshold limit of 10 is selected for a 9x9 window. Applying the thresholding operator to the business matrix with this limit produces the result shown in figure 3.75, after smoothing and contraction. There are a number of areas which warrant further consideration. We set out to extract the regions of appreciable size (1% or more of the scene). Once they are isolated they are reexpanded. Notice that the sky and hill portion is isolated as one piece even though the sky has already been processed (figure 3.76). This is an intermediate step that is performed to fill in some of the holes caused by the edges that separate hills and sky. If the processed segment were masked out first, an imprecise boundary determination would result that could not be completely adjusted by the expansion operation. Once the combined regions are extracted and filled the sky is masked out to give the result shown in figure 3.77 The other regions that meet the size requirement are isloated in the same manner (figure 3.78).

Once the homogeneous portions of the picture are isolated they are processed in accordance with the basic algorithm. There are, however, some modifications that need to be made to compensate for the manner in which the regions were extracted. The changes that we feel are necessary are shown in figures 3.79 and 3.80. The flow chart shown in figure 3.79 depicts a subroutine which is to be inserted between steps 11.1 and 11.2 of the original algorithm (figure 3.12). The constructs shown in figure 3.83 are pieces of flow diagram that replace the designated steps in the chart of figure 3.12.

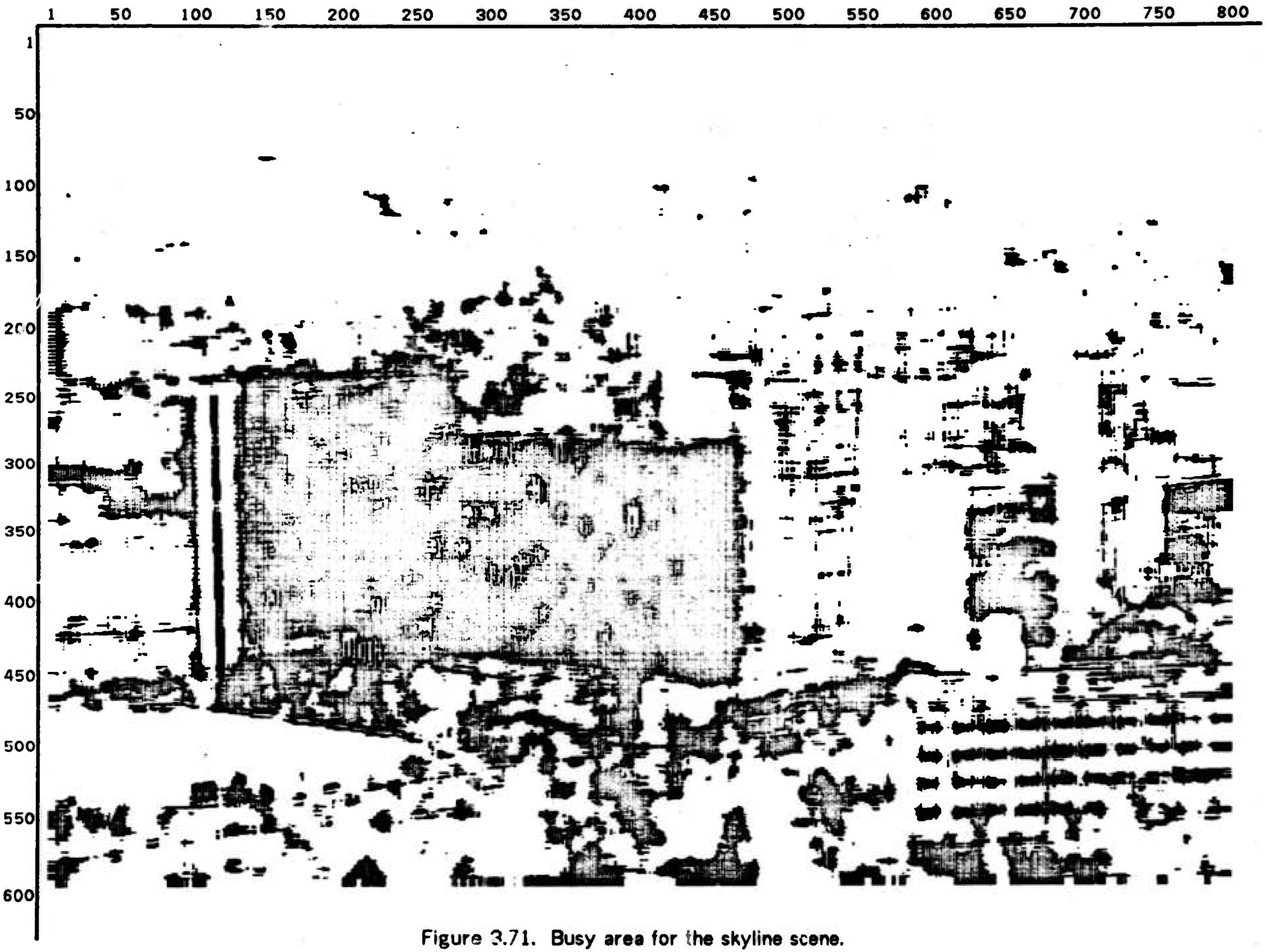


Figure 3.71. Busy area for the skyline scene.

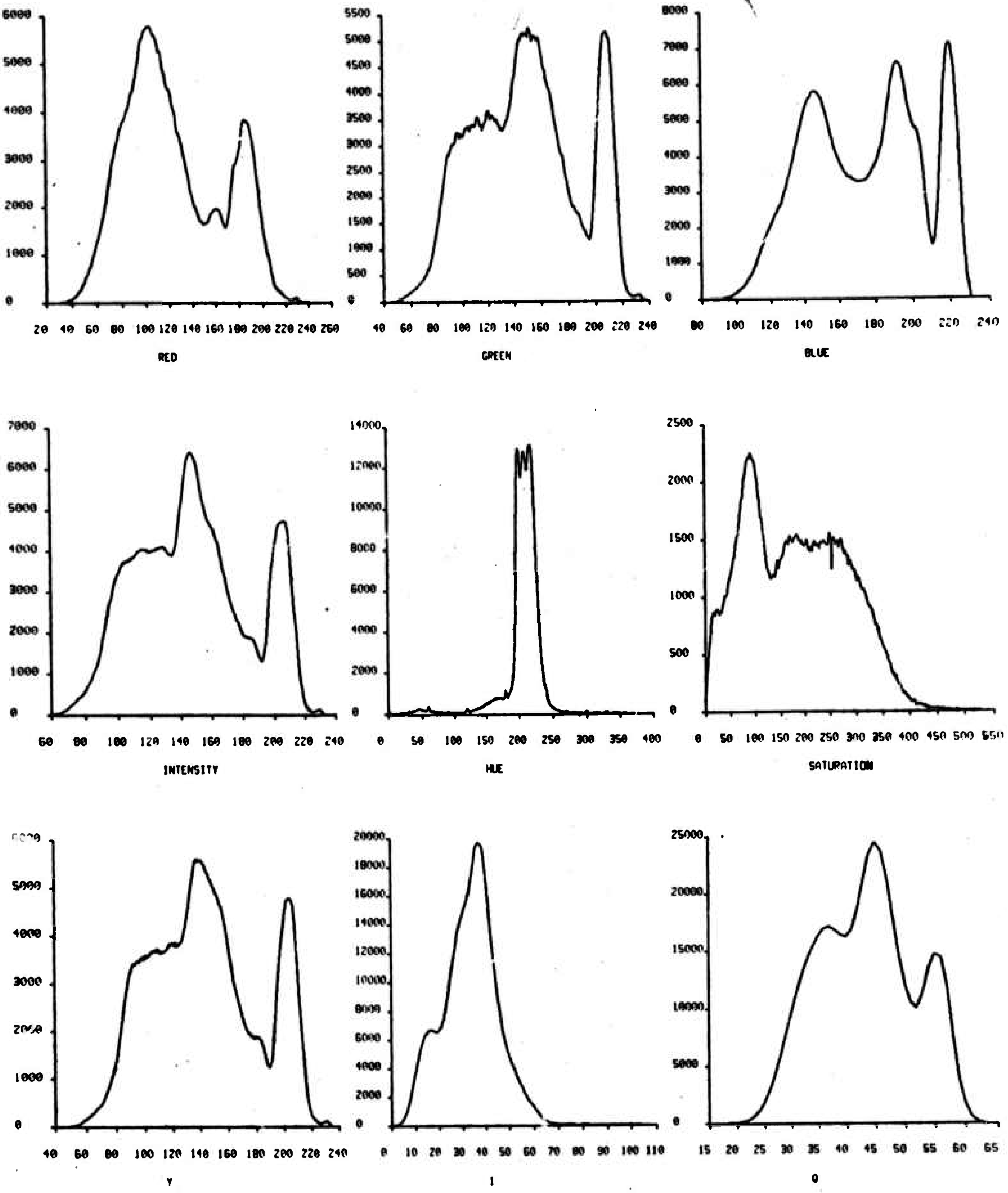


Figure 3.72. Nine parameter histograms of skyline scene.

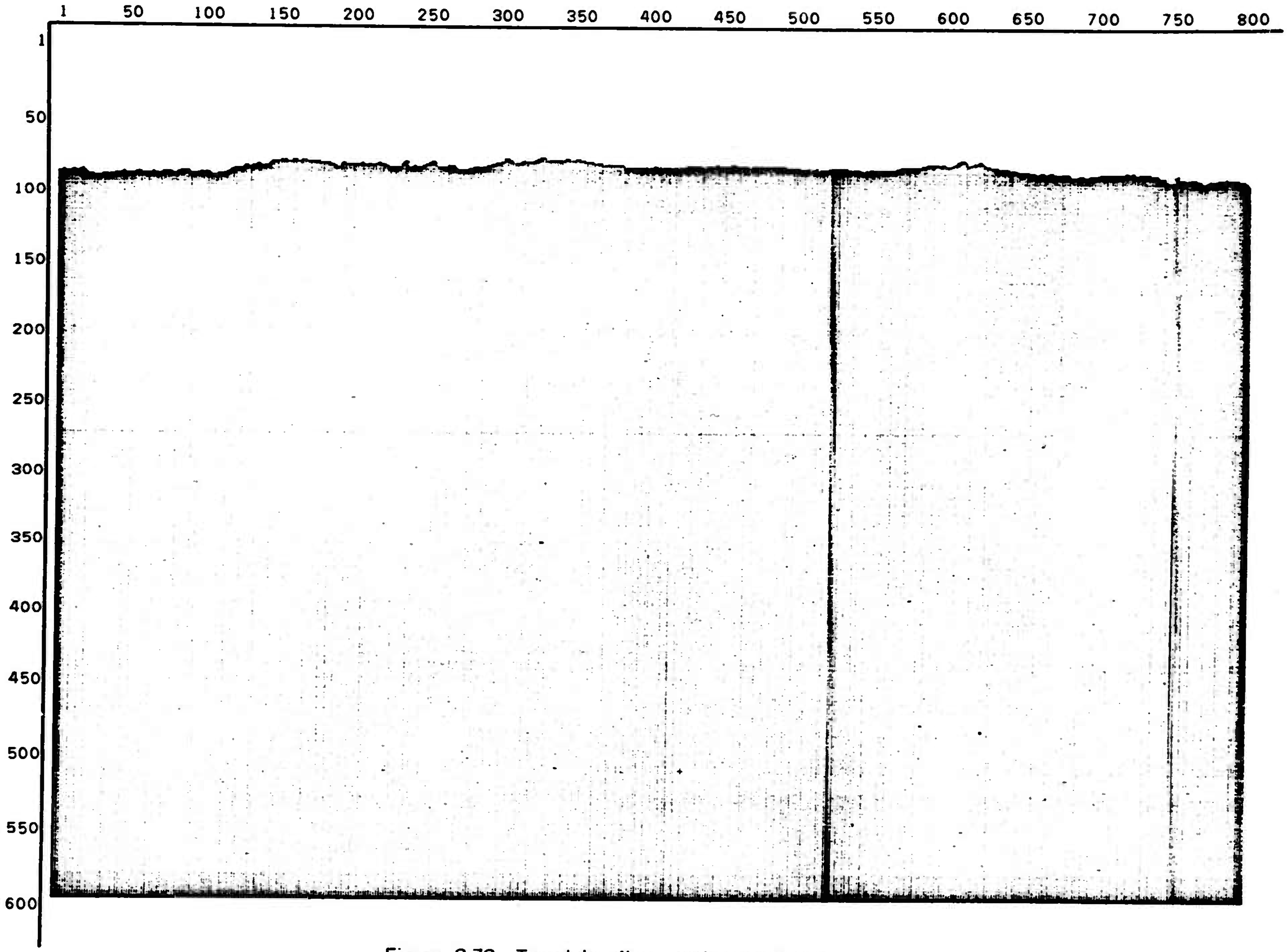


Figure 3.73. Template after masking out the sky.

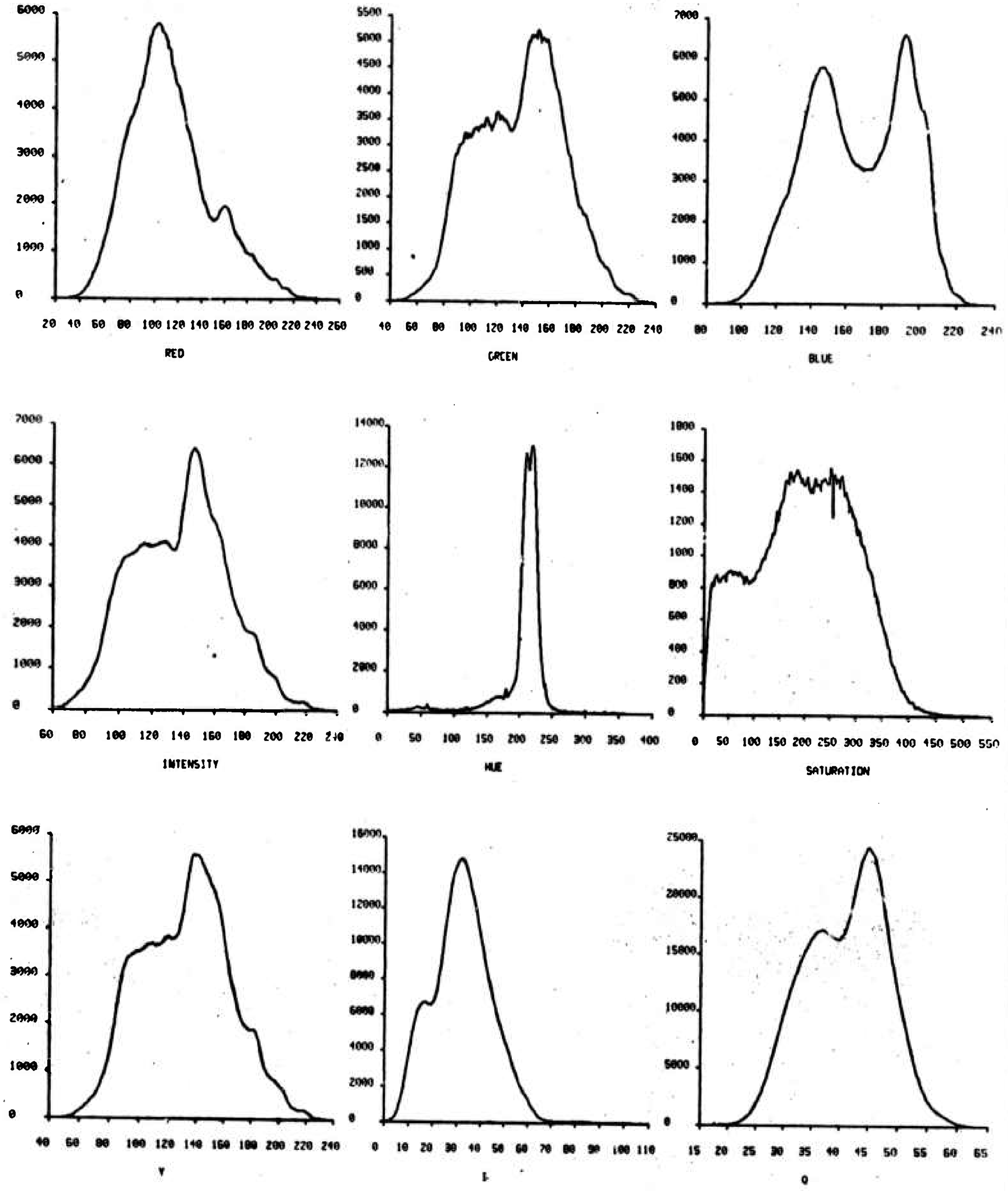


Figure 3.74. Nine parmeter histograms of figure 3.73.

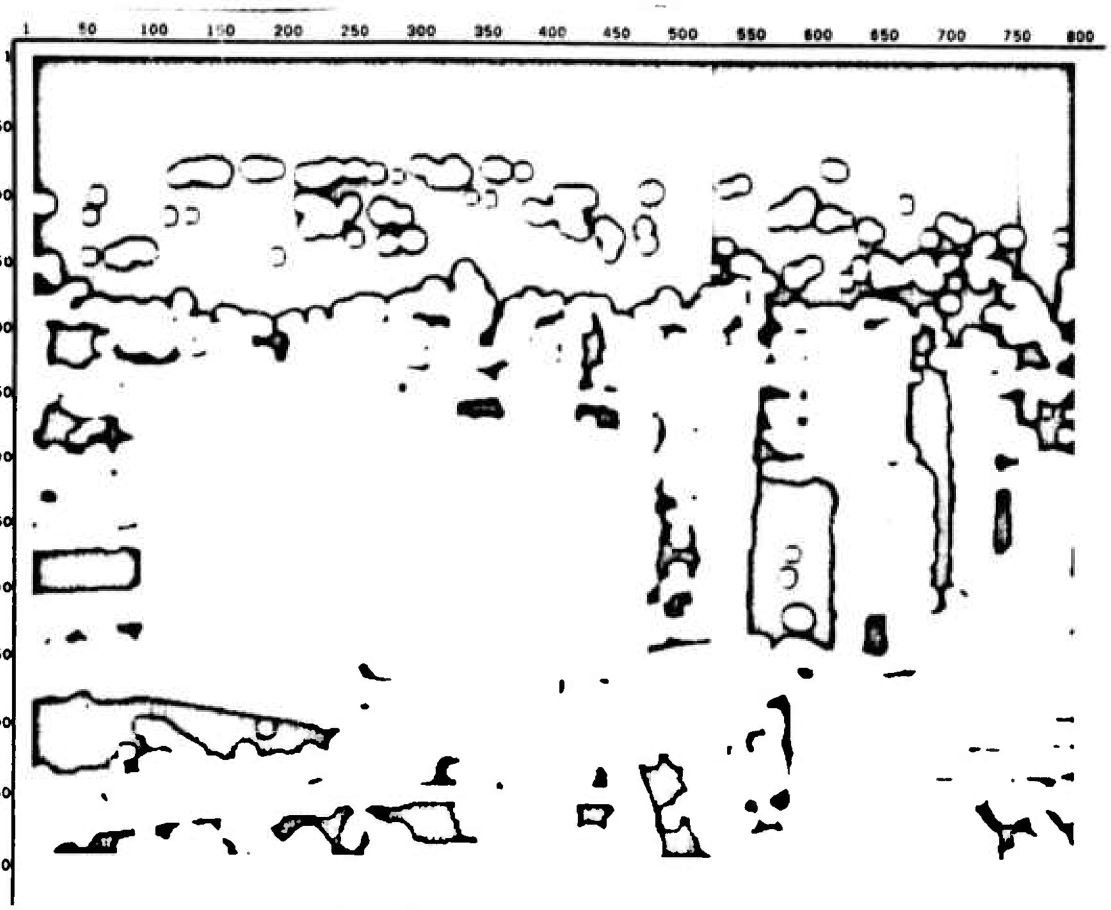


Figure 3.75. Light texture area of the skyline scene.

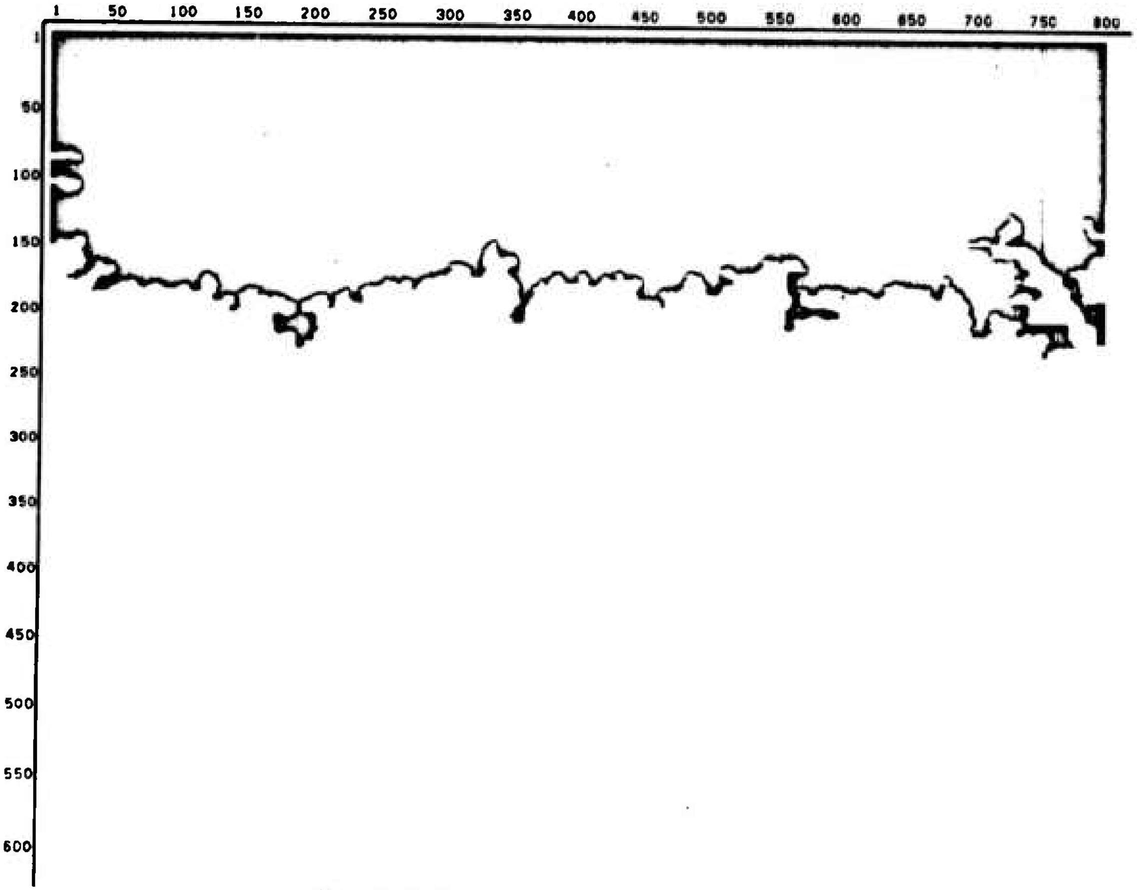


Figure 3.76. Sky and hill portion of the skyline scene

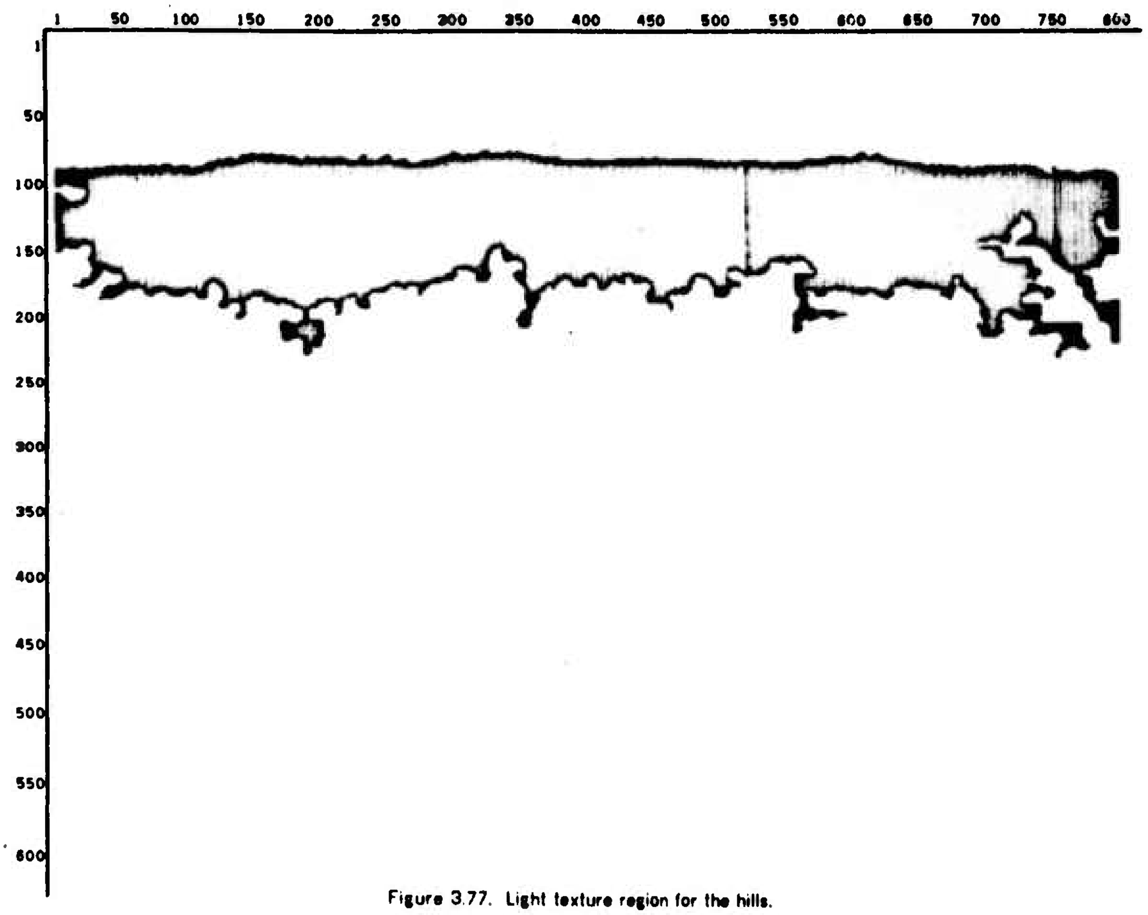


Figure 3.77. Light texture region for the hills.

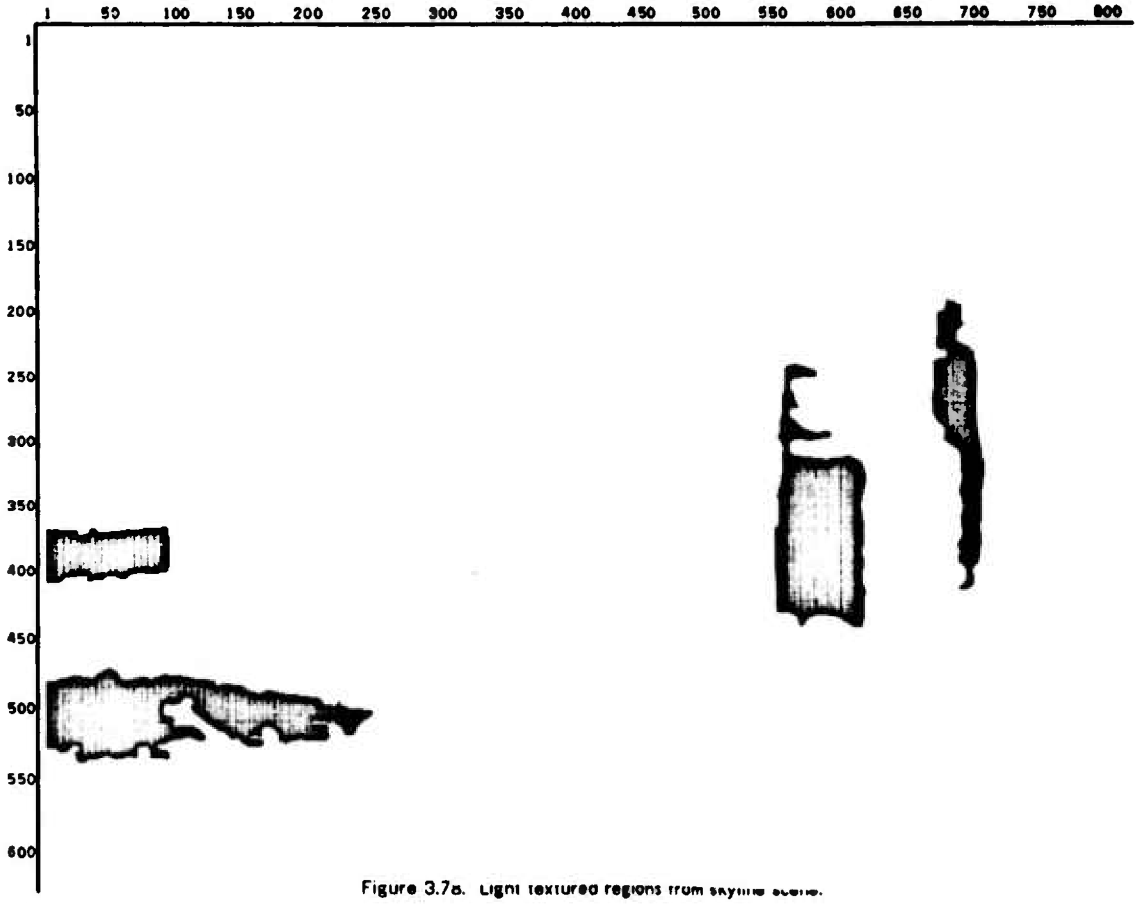


Figure 3.78. Light textured regions from skyline scene.

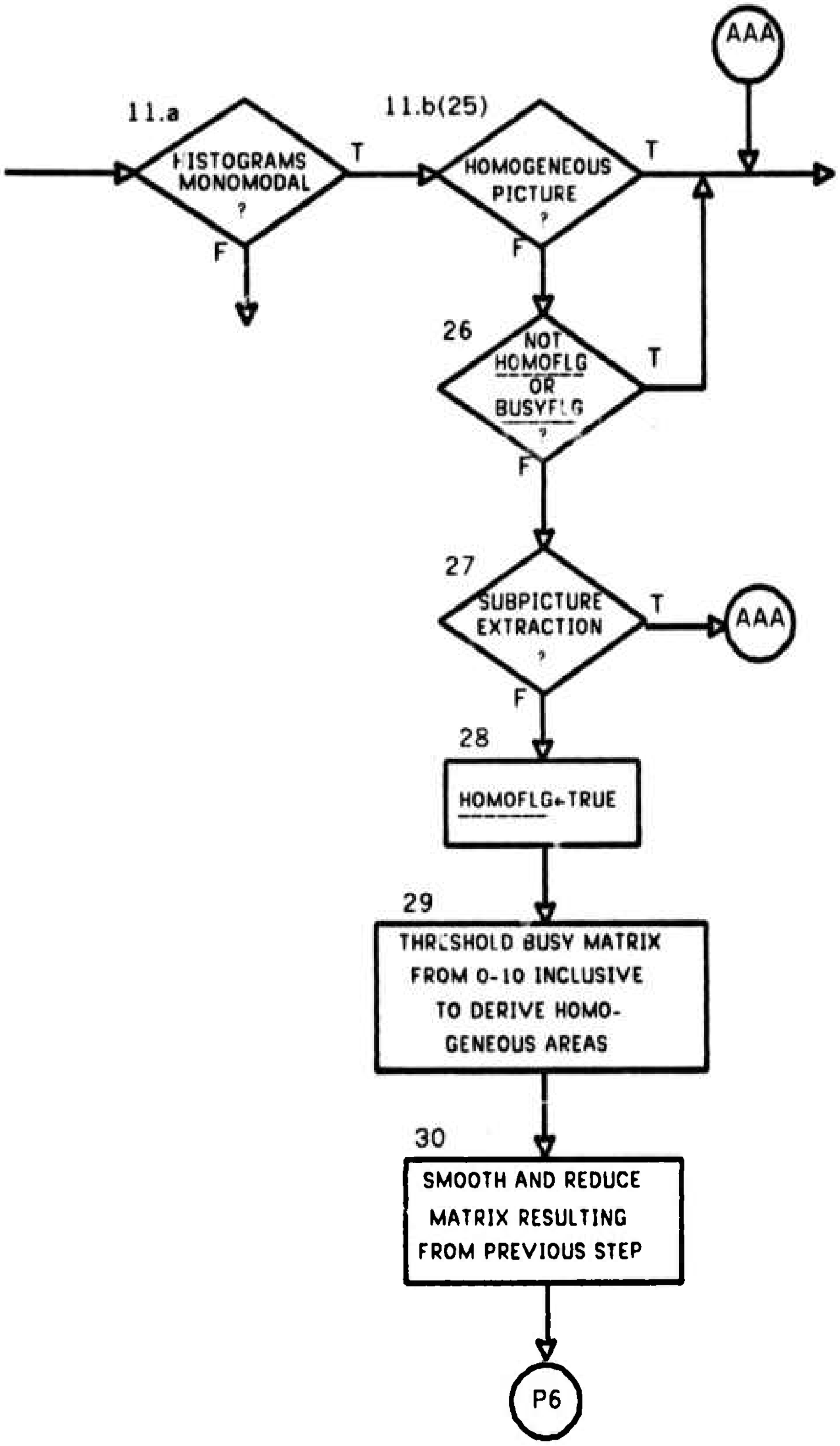


Figure 3.79. Modification to the algorithm for forced isolation.

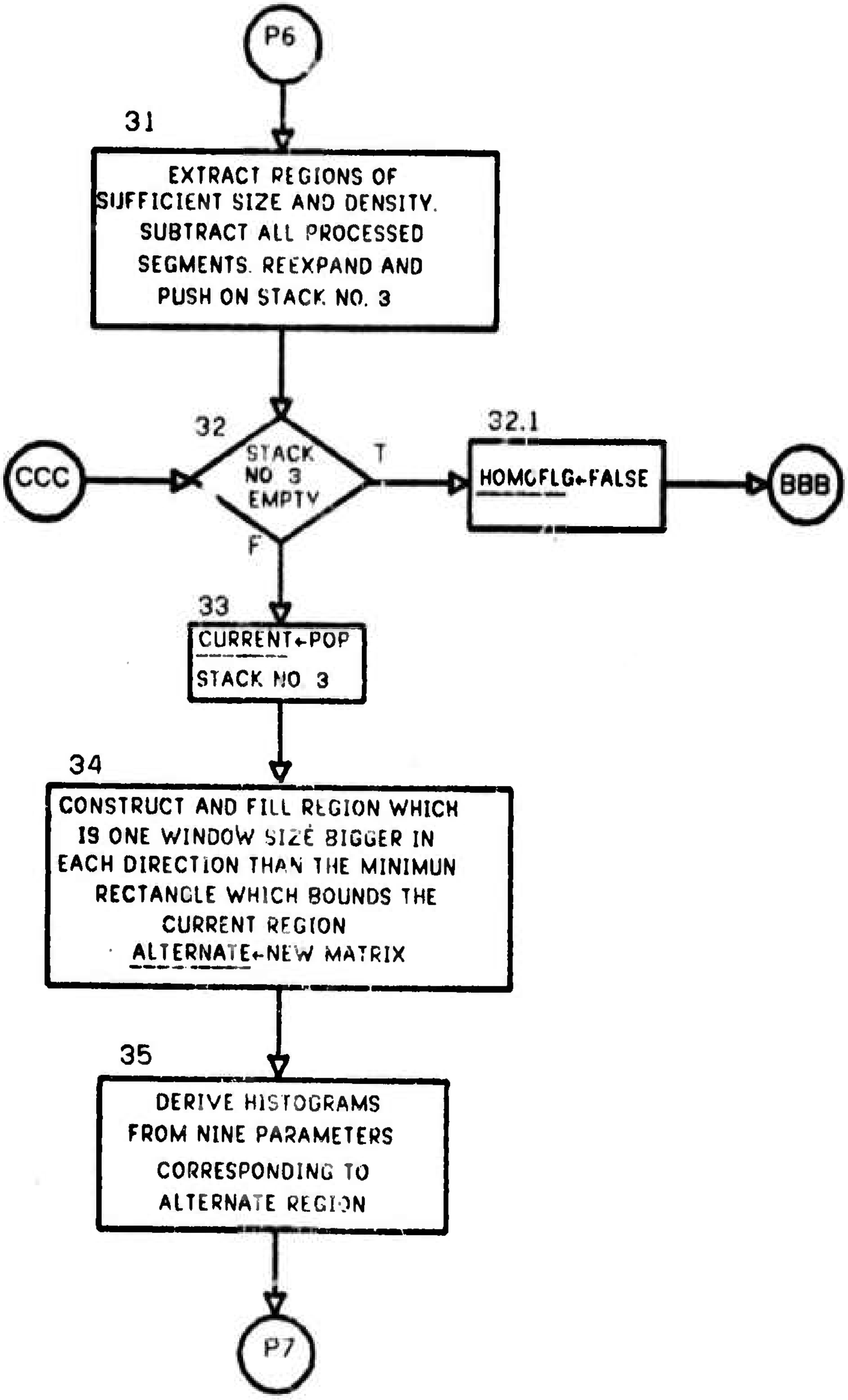

P6
31
EXTRACT REGIONS OF SUFFICIENT SIZE AND DENSITY. SUBTRACT ALL PROCESSED SEGMENTS. REEXPAND AND PUSH ON STACK NO. 3
32
STACK NO 3 EMPTY
T
F
CCC
32.1
HOMOFLG←FALSE
BBB
33
CURRENT←POP STACK NO 3
34
CONSTRUCT AND FILL REGION WHICH IS ONE WINDOW SIZE BIGGER IN EACH DIRECTION THAN THE MINIMUN RECTANGLE WHICH BOUNDS THE CURRENT REGION
ALTERNATE←NEW MATRIX
35
DERIVE HISTOGRAMS FROM NINE PARAMETERS CORRESPONDING TO ALTERNATE REGION
P7

Figure 3.79 (continued). Modification to the algorithm for forced isolation.

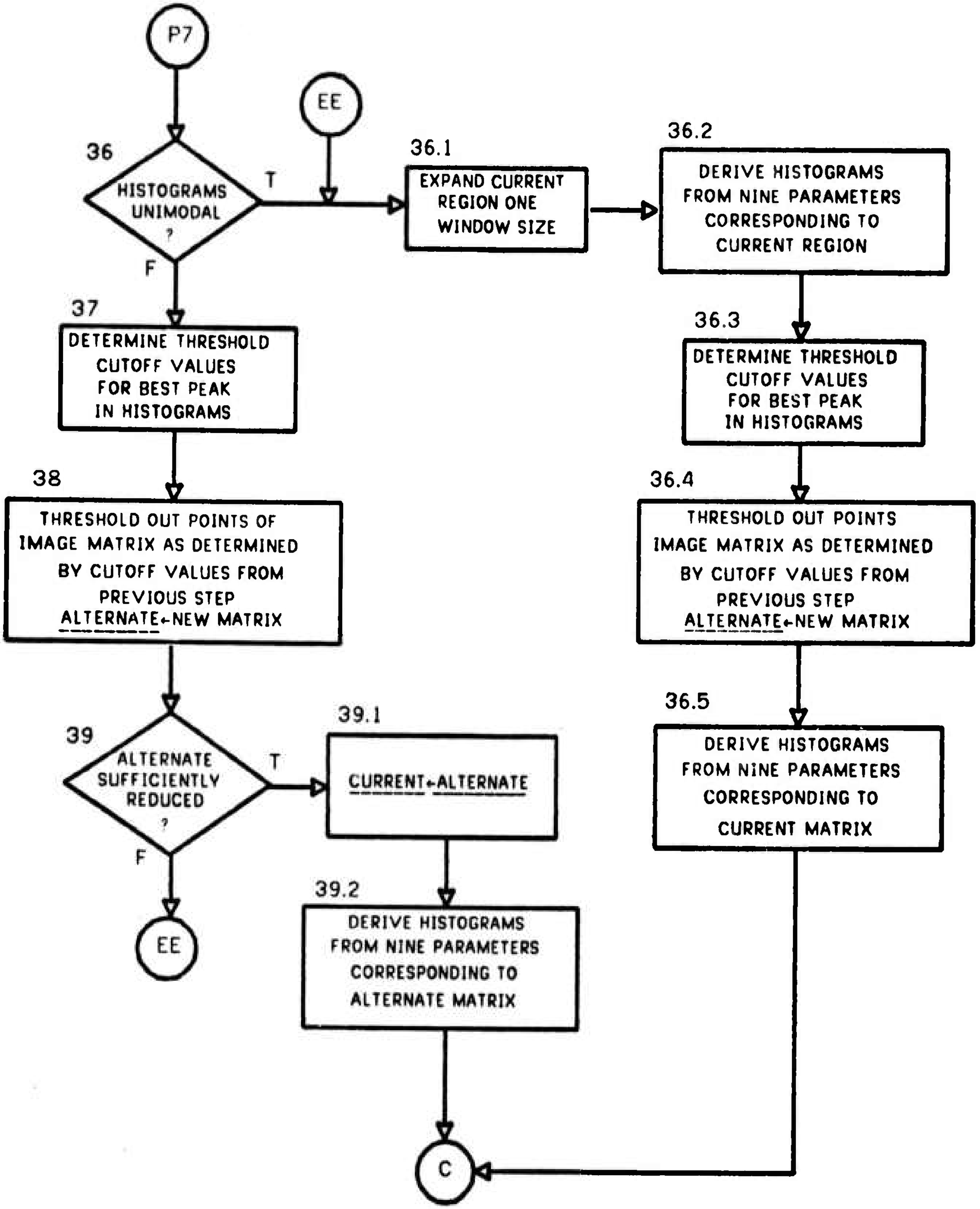

P7
EE
36
HISTOGRAMS UNIMODAL ?
T
F
36.1
EXPAND CURRENT REGION ONE WINDOW SIZE
36.2
DERIVE HISTOGRAMS FROM NINE PARAMETERS CORRESPONDING TO CURRENT REGION
36.3
DETERMINE THRESHOLD CUTOFF VALUES FOR BEST PEAK IN HISTOGRAMS
36.4
THRESHOLD OUT POINTS IMAGE MATRIX AS DETERMINED BY CUTOFF VALUES FROM PREVIOUS STEP ALTERNATE←NEW MATRIX
36.5
DERIVE HISTOGRAMS FROM NINE PARAMETERS CORRESPONDING TO CURRENT MATRIX
37
DETERMINE THRESHOLD CUTOFF VALUES FOR BEST PEAK IN HISTOGRAMS
38
THRESHOLD OUT POINTS OF IMAGE MATRIX AS DETERMINED BY CUTOFF VALUES FROM PREVIOUS STEP ALTERNATE←NEW MATRIX
39
ALTERNATE SUFFICIENTLY REDUCED ?
T
F
EE
39.1
CURRENT←ALTERNATE
39.2
DERIVE HISTOGRAMS FROM NINE PARAMETERS CORRESPONDING TO ALTERNATE MATRIX
C

Figure 3.79 (continued). Modification to the algorithm for forced isolation.

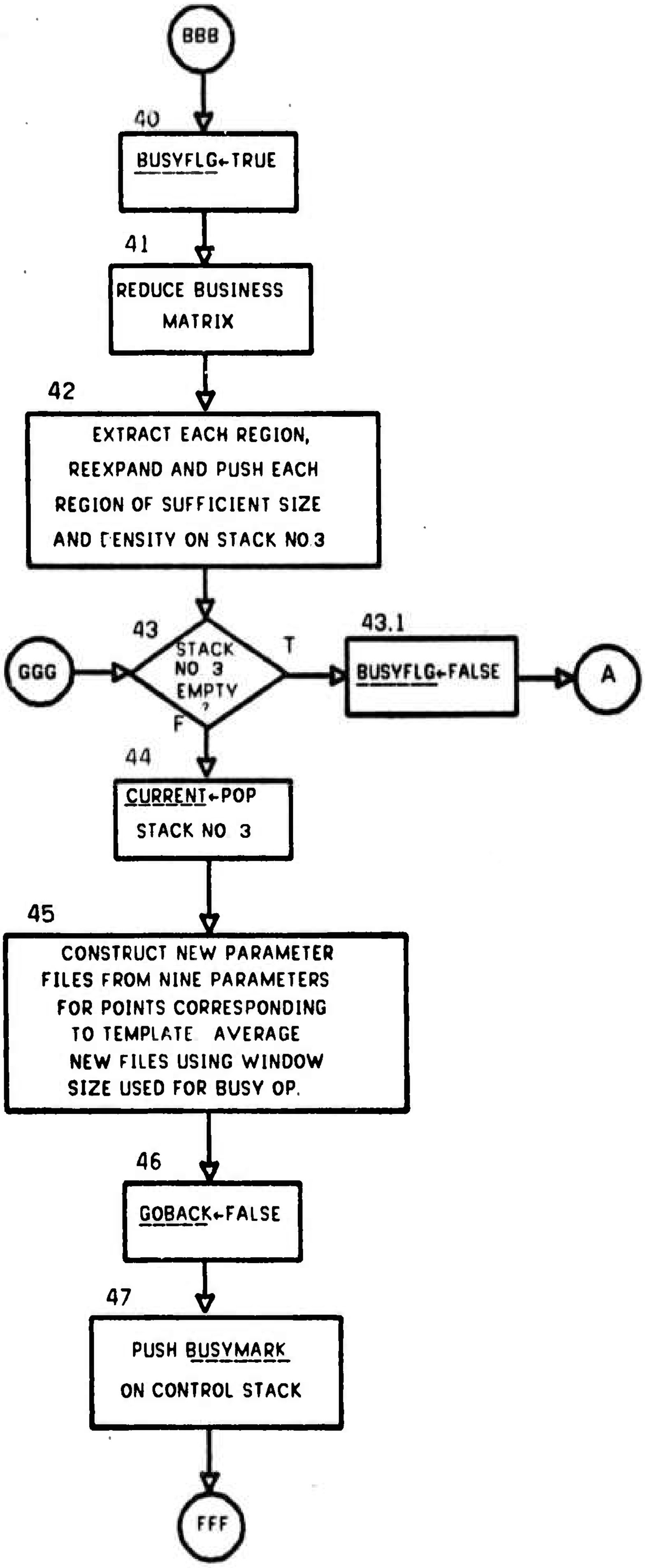

BBB
40
BUSYFLG←TRUE
41
REDUCE BUSINESS MATRIX
42
EXTRACT EACH REGION, REEXPAND AND PUSH EACH REGION OF SUFFICIENT SIZE AND DENSITY ON STACK NO 3
43
STACK NO 3 EMPTY ?
T
F
GGG
43.1
BUSYFLG←FALSE
A
44
CURRENT←POP STACK NO 3
45
CONSTRUCT NEW PARAMETER FILES FROM NINE PARAMETERS FOR POINTS CORRESPONDING TO TEMPLATE AVERAGE NEW FILES USING WINDOW SIZE USED FOR BUSY OP.
46
GOBACK←FALSE
47
PUSH BUSYMARK ON CONTROL STACK
FFF

Figure 3.79 (continu[illegible]. Modification to the algorithm for forced isolation.

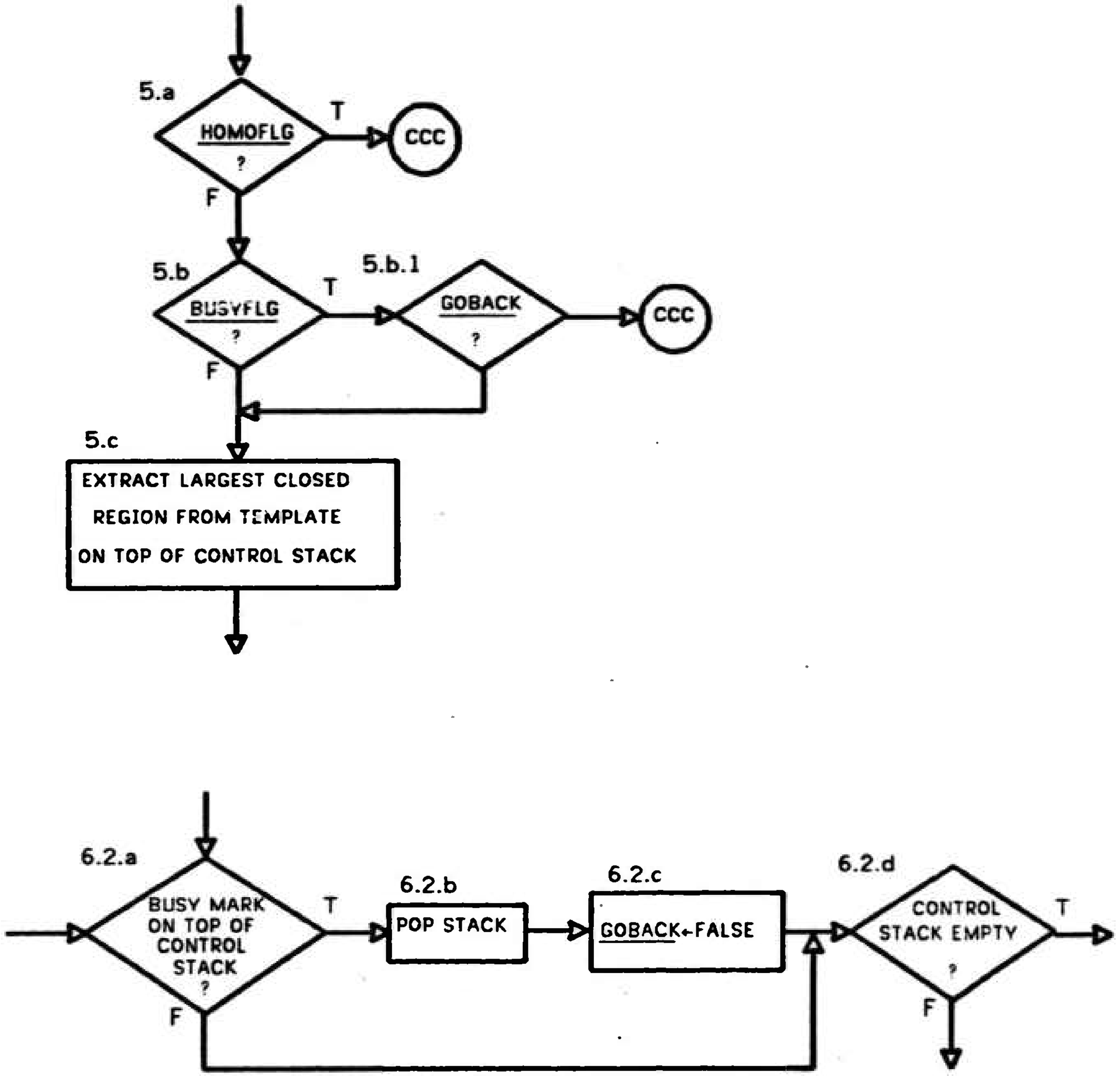


Figure 3.80. Modifications to steps in the main algorithm.

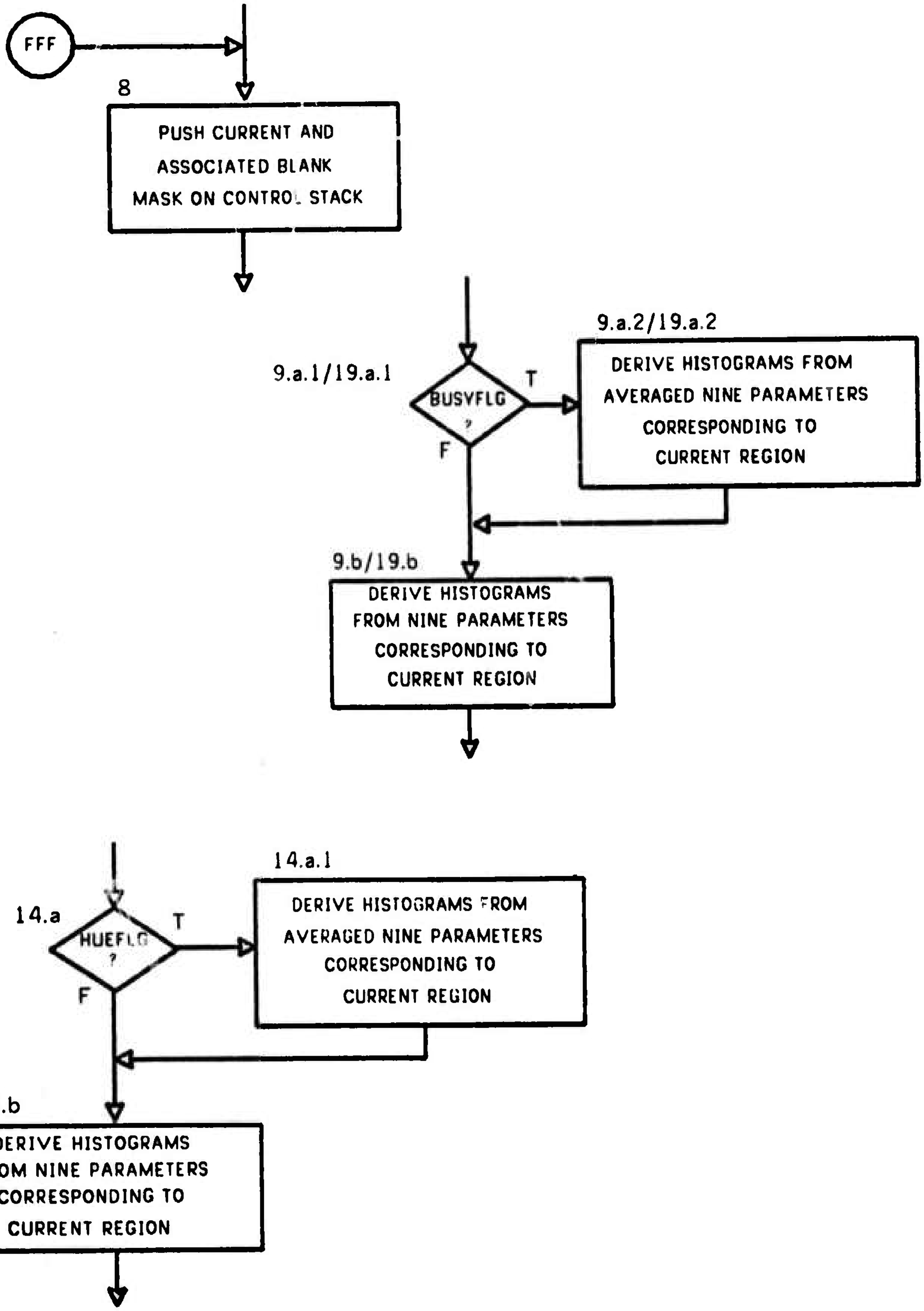

FFF
8
PUSH CURRENT AND ASSOCIATED BLANK MASK ON CONTROL STACK
9.a.1/19.a.1
BUSYFLG ?
T
F
9.a.2/19.a.2
DERIVE HISTOGRAMS FROM AVERAGED NINE PARAMETERS CORRESPONDING TO CURRENT REGION
9.b/19.b
DERIVE HISTOGRAMS FROM NINE PARAMETERS CORRESPONDING TO CURRENT REGION
14.a
HUEFLG ?
T
F
14.a.1
DERIVE HISTOGRAMS FROM AVERAGED NINE PARAMETERS CORRESPONDING TO CURRENT REGION
14.b
DERIVE HISTOGRAMS FROM NINE PARAMETERS CORRESPONDING TO CURRENT REGION

Figure 3.80.(continued) Modifications to steps in the main algorithm.

Segmentation

The steps 35 through 39 shown in figure 3.79 are designed to compensate for the inaccurate delimitation of the homogeneous regions that have been obtained. If the extracted region is surrounded by areas of dissimilar properties, the thresholding operation could successfully be applied to a portion of the scene which just encloses the desired region. The result should be a precise segmentation. Operating on this assumption, step 34 of the procedure constructs a solid template which is one window size larger, in each dimension, than the minimum bounding rectangle (MBR) that contains the original extraction. This MBR is then masked with all existing processed segments which overlap the new construction. This restores precise boundaries where possible. Figure 3.81 shows the result obtained for the hills. We then derive histograms for the newly constructed region and test for monomodality. For this phase of the operation we are willing to relax our standards concerning acceptability of a peak for cutoff. We still look for the peaks showing signs of sharp discontinuity, but if these are not available we will accept any that show straggling tails. After all, what we are trying to do is trim the "square plug" that we have formed. If the histograms are strictly monomodal, steps 36.1 through 36.5 retrieve the original homogeneous segment, expand it one window size, derive histograms, threshold the result, and then go into the refinement phase of the basic algorithm. The final result will become a processed segment. If the histograms extracted at step 35 do provide some cutoff values to act upon, a threshold operation is performed. The result is tested to see if a significant reduction was obtained (e.g. 15%). This step is necessary because the plug may constitute a substantial expansion on the original region. If surrounding portions of the picture do not allow an effective thresholding and paring of the plug we want to accept the best alternative, which is the original. If the required reduction was obtained, new histograms are derived and possibility of refinement investigated. If the reduction was not sufficient the same steps are taken as for monomodality.

Let us process some of the previously extracted homogeneous regions in terms of the operations just discussed. The histograms derived from figure 3.81 do not show much sign of discontinuity (figure 3.82). Just to be sure we threshold on the intensity parameter using the cutoff values 32 and 183 obtained by a Gaussian extension. The insufficient reduction forces us to step 36.1 of the algorithm to recover the original segment. This is expanded (figure 3.83) and new histograms are derived (figure 3.84). No further refinements can be made so thresholding, smoothing, contraction, and expansion operations are applied to produce the result of figure 3.85 as our best segmentation of the hill area.

Continuing with this phase of the processing brings the park in the lower left corner of figure 3.78 into consideration. A plug is constructed as before (figure 3.86) and histograms derived (figure 3.87). This time a number of the graphs provide adequate signs of discontinuity around a uniform area. The red parameter is especially indicative of this phenomenon. Thresholding on this parameter yields a region which provides an example of the best kind of results obtainable from this procedure. Complete processing of the homogeneous regions yields the results shown in figure 3.88. One may have noted that a progressive reduction of the extracted homogeneous regions was not performed. The tacit assumption is that <u>we have not</u> isolated two or more regions which are in immediate proximity, which are possessed of the common feature of homogeneity, and which are dissimilar in some other parameter. If the assumption is valid (which it is for this scene), nothing is to be gained by a recursive

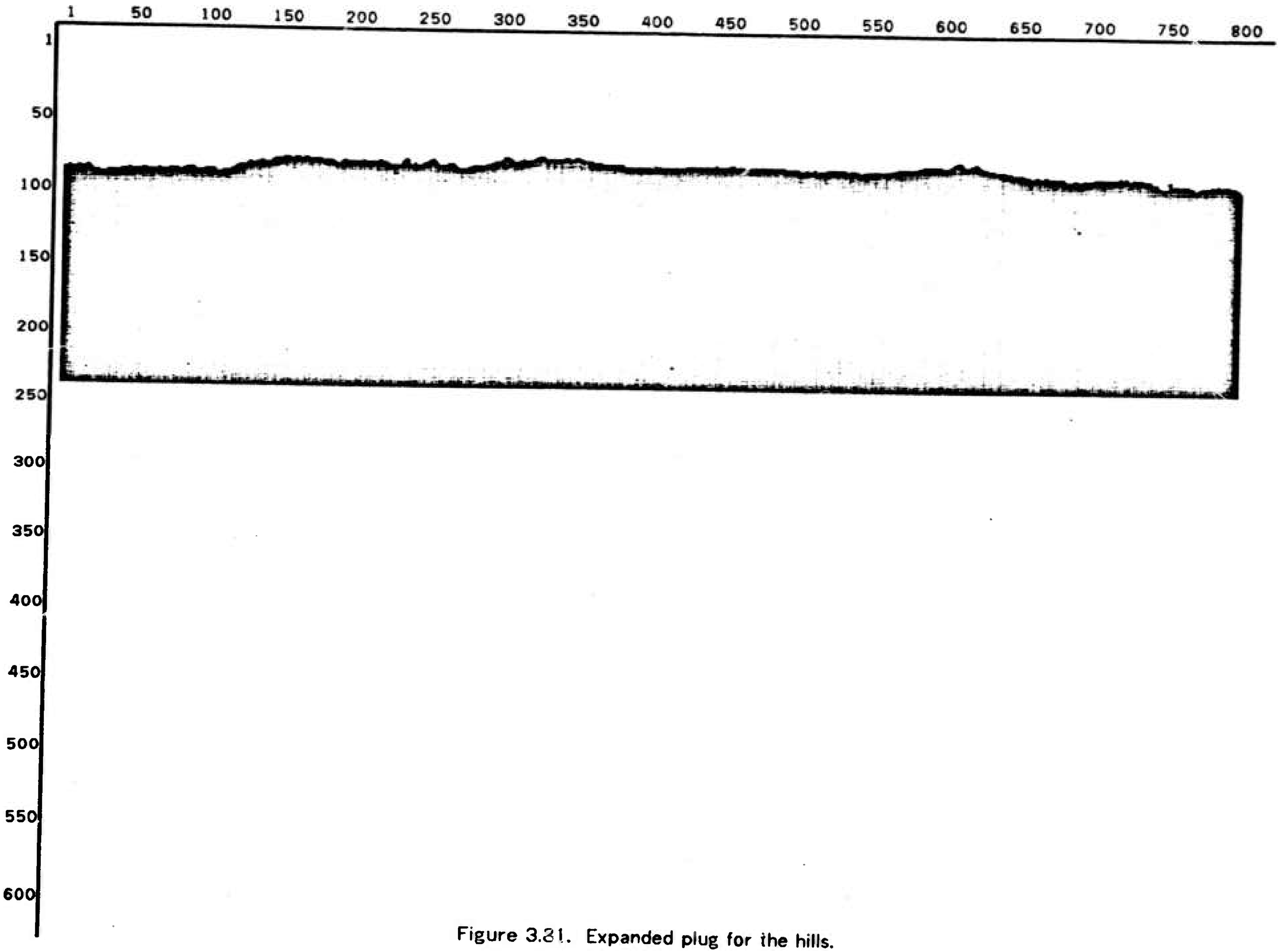


Figure 3.31. Expanded plug for the hills.

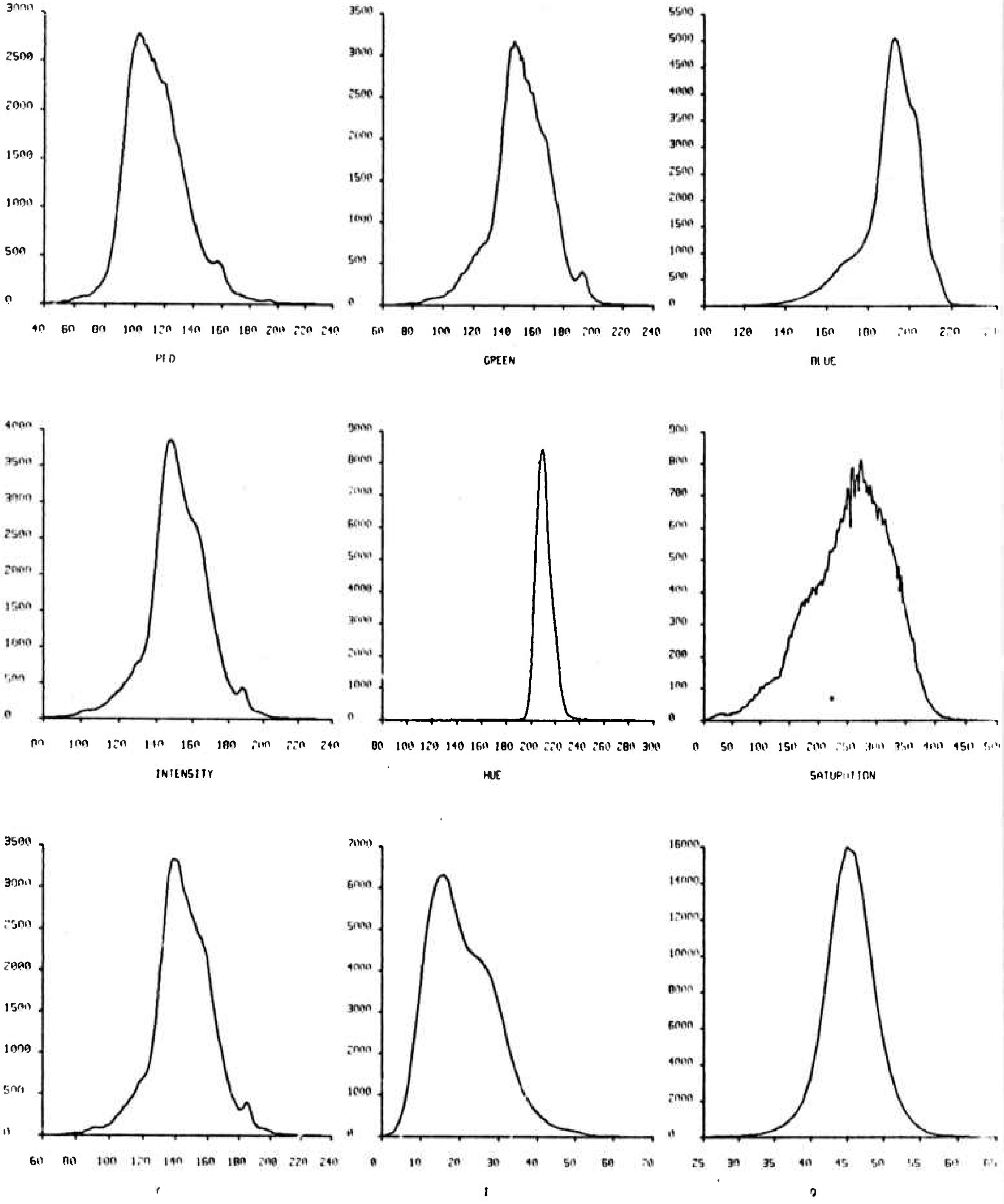


Figure 3.82. Nine parameter histograms for figure 3.81.

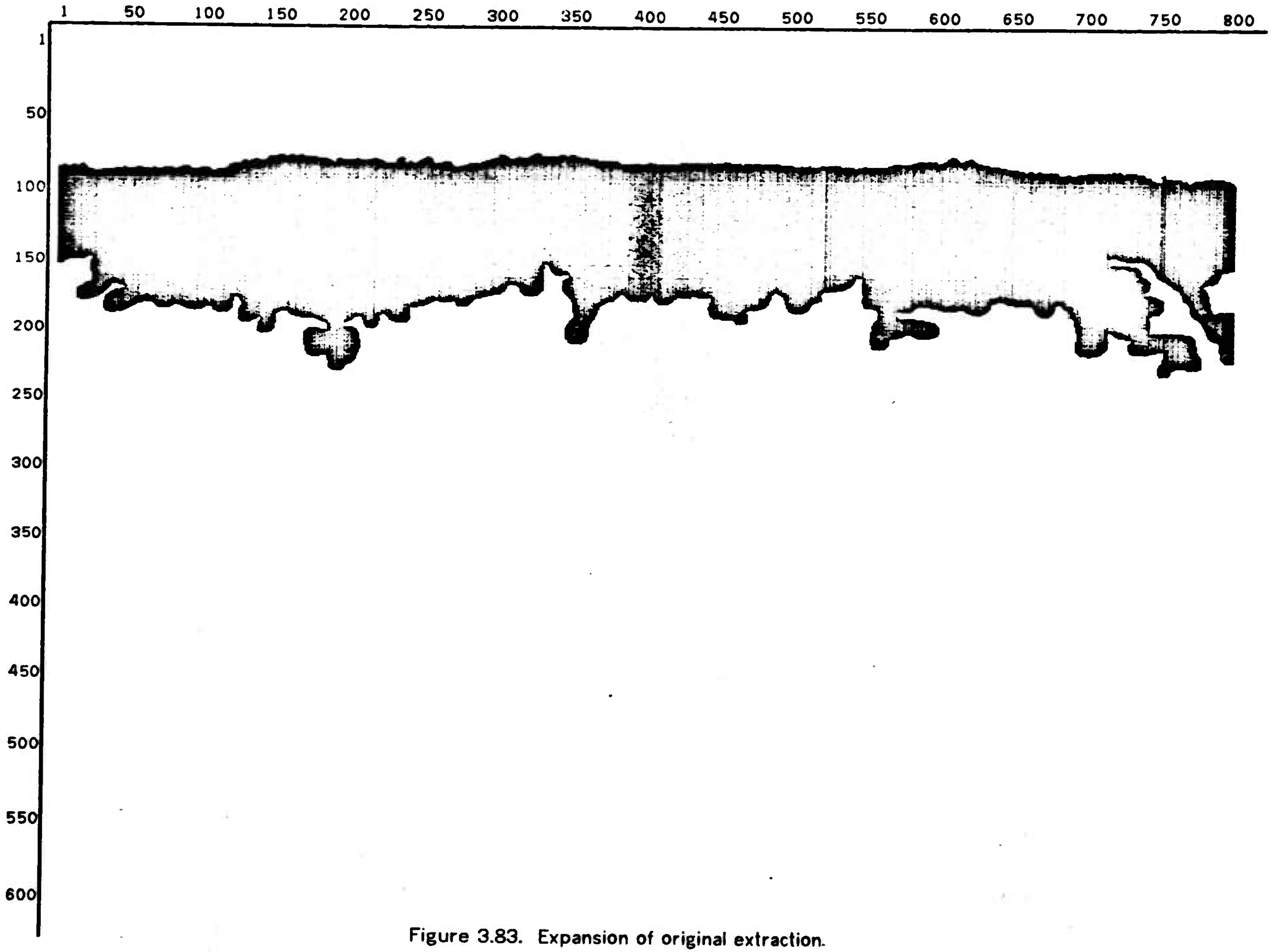


Figure 3.83. Expansion of original extraction.

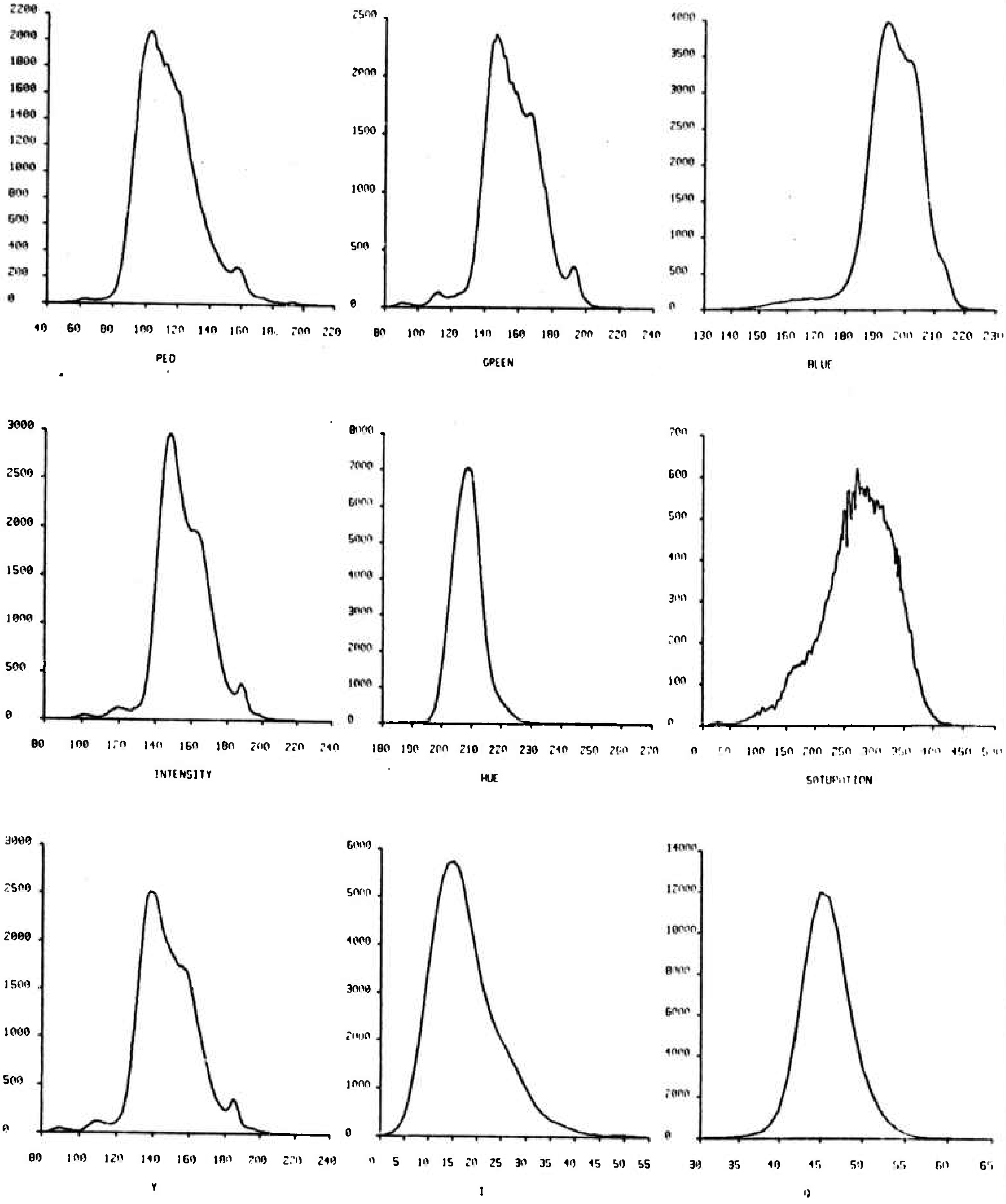


Figure 3.84. Nine parameter histograms of figure 3.83.

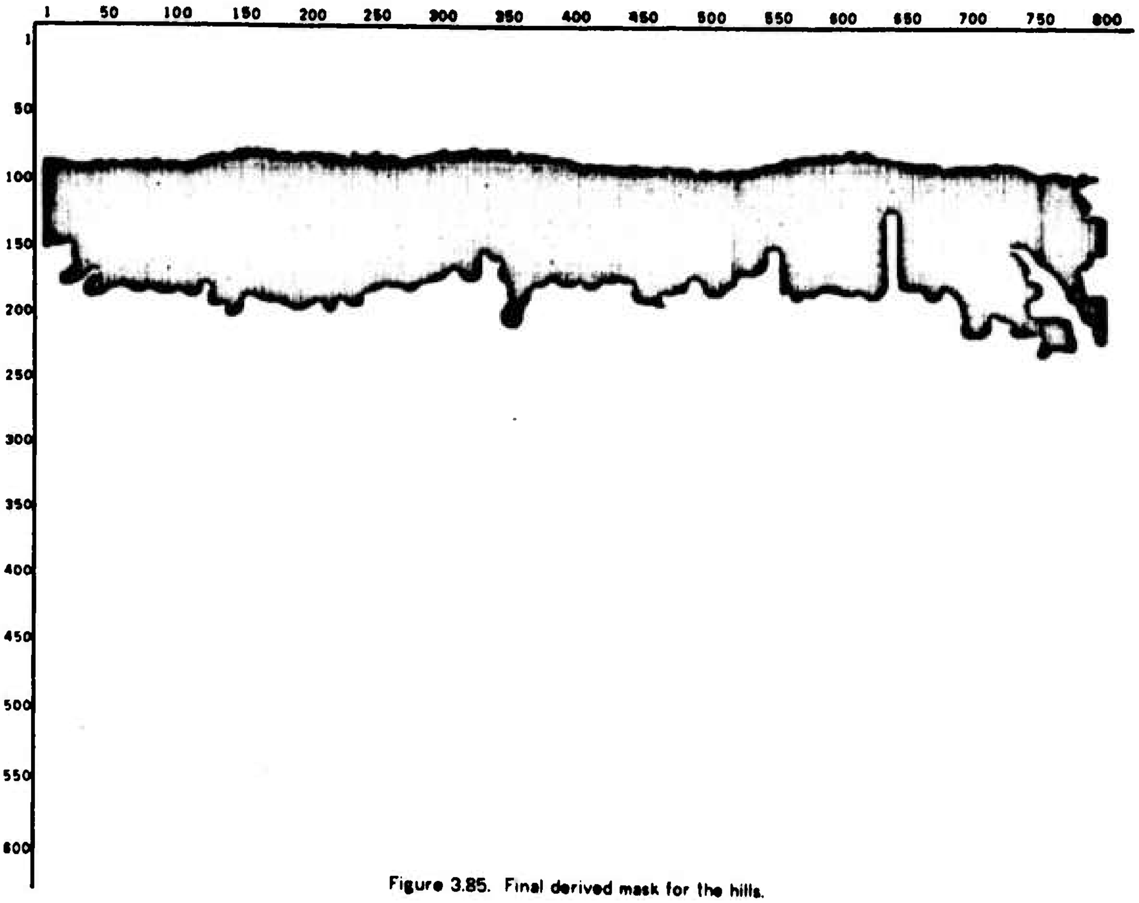


Figure 3.85. Final derived mask for the hills.

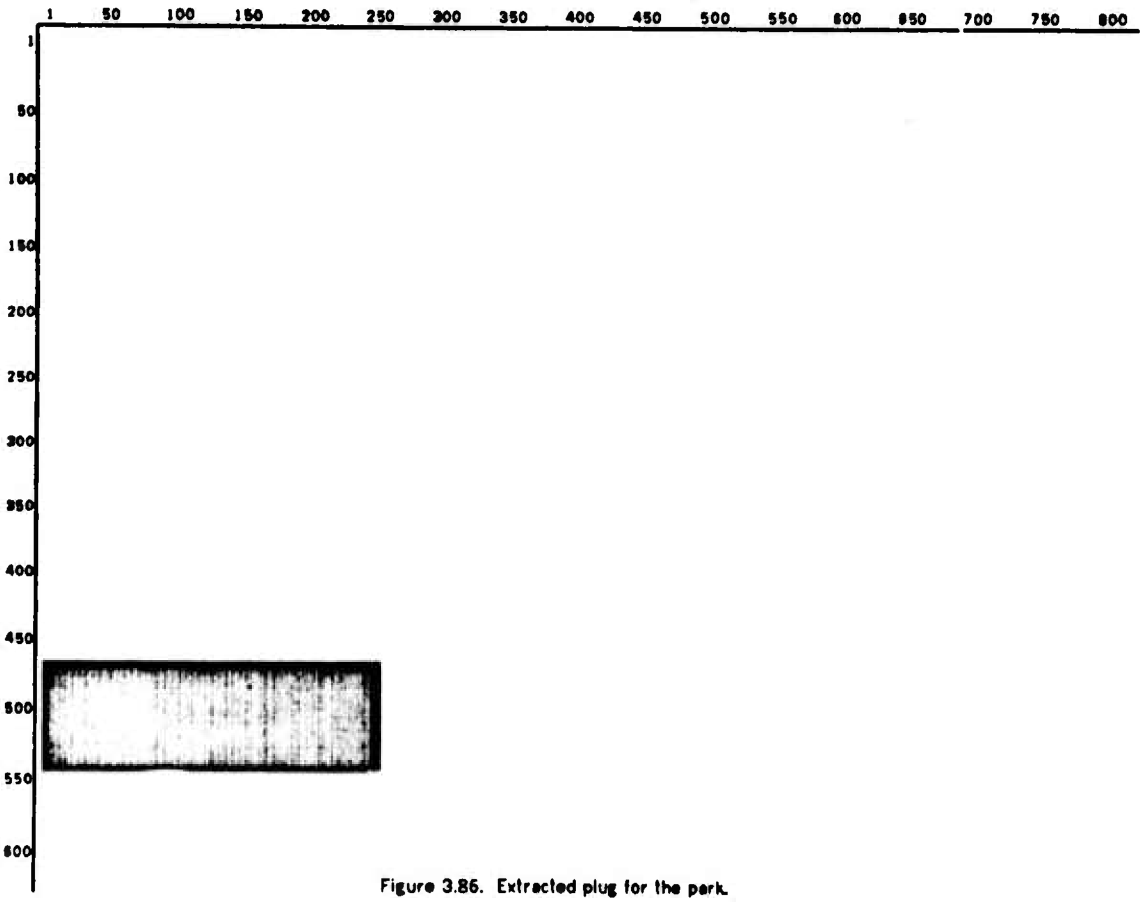


Figure 3.86. Extracted plug for the park.

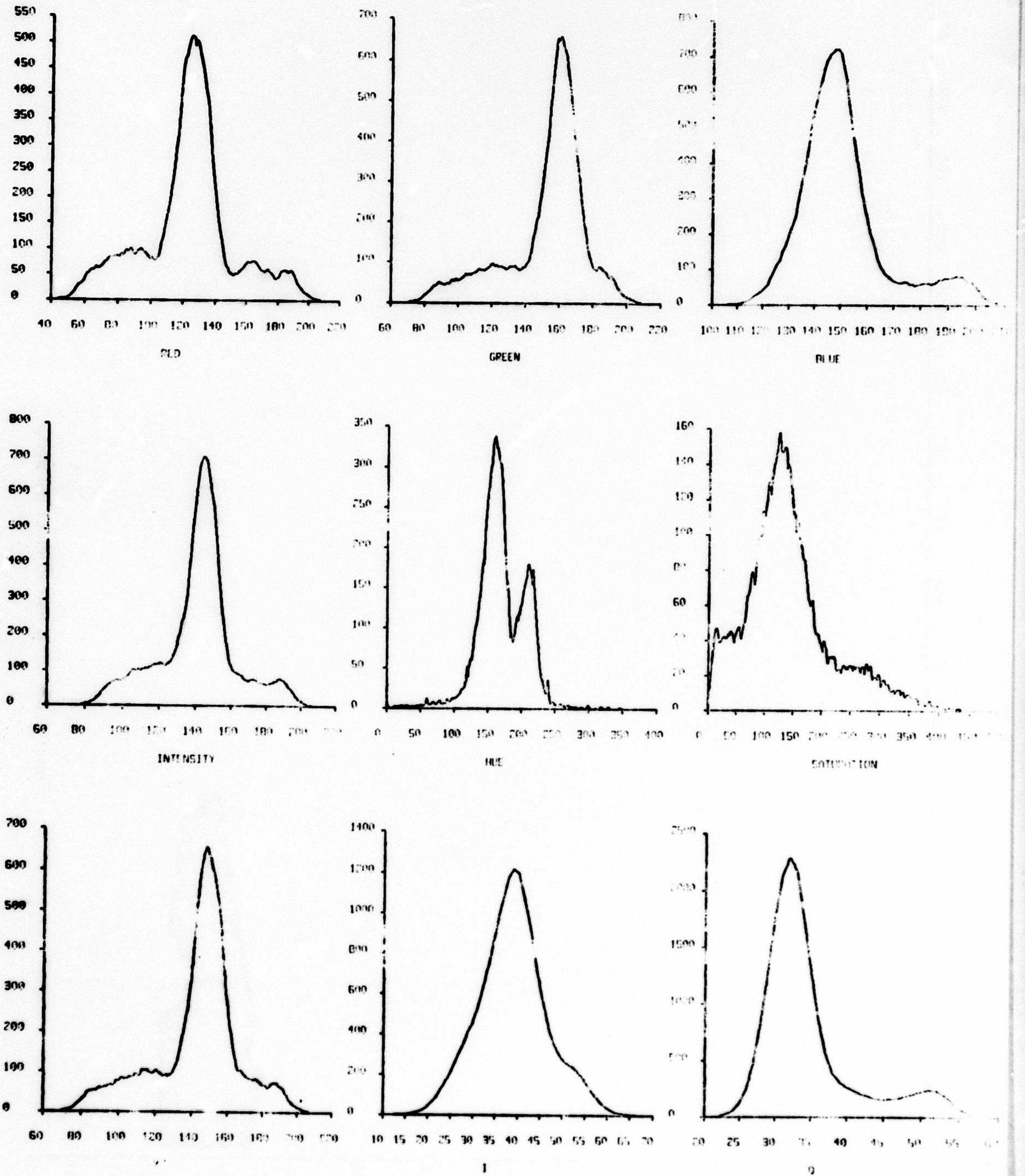


Figure 3.87. Nine parameter histograms for figure 3.86.

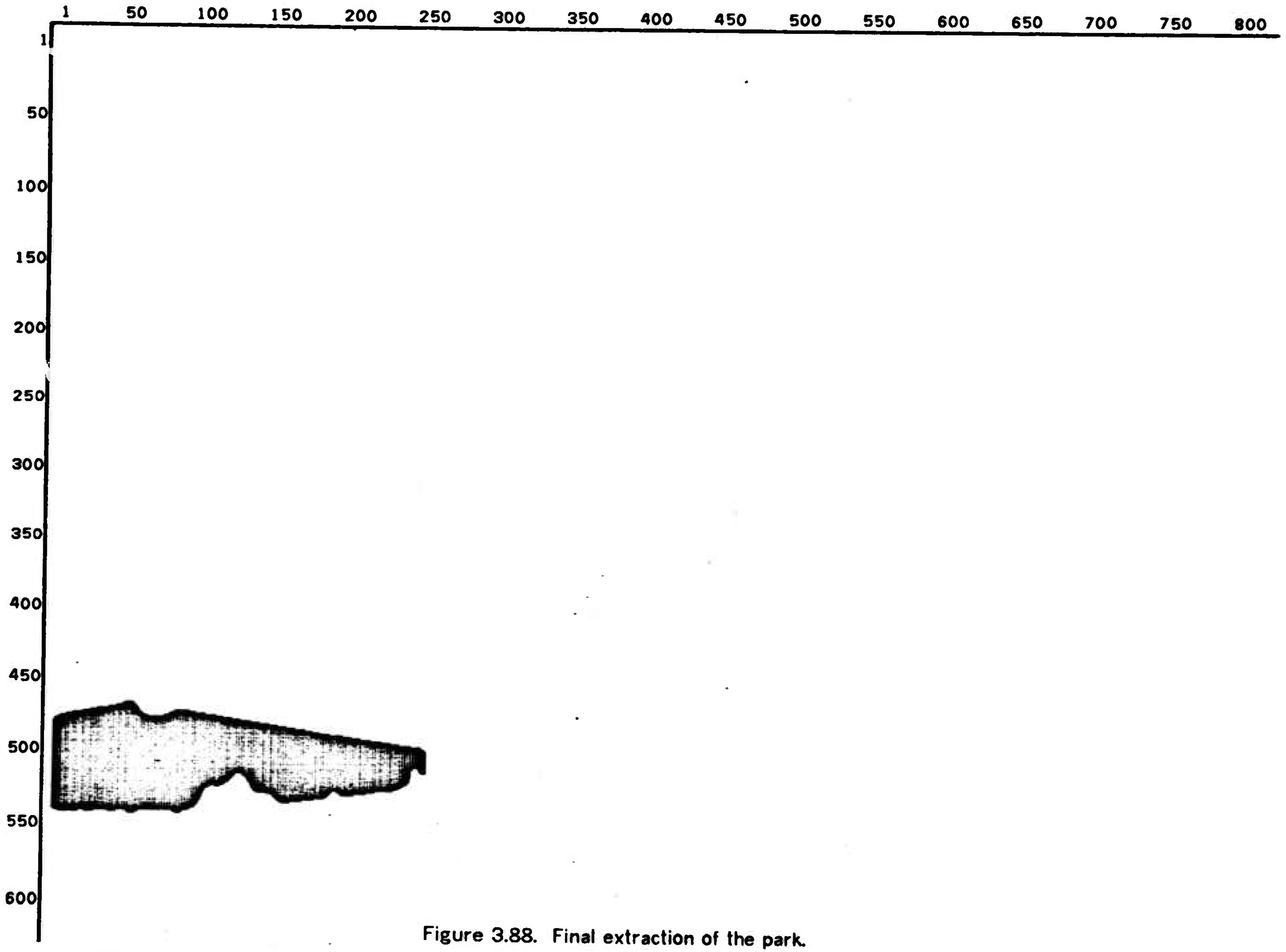


Figure 3.88. Final extraction of the park.

analysis of the extracted regions. We feel that the assumption is valid for most cases because, if there had been substantial areas of the picture which were homogeneous, they would have manifested themselves in histograms of the original scene. It is possible that modification to the procedure will become necessary if scenes are encountered which have regions which are similar in homogeneity but dissimilar in other respects; presuming, of course, that treatment by forced isolation is necessary.

Once we have processed the homogeneous regions of the picture there are two options available. One can return to processing the picture in the normal way, or we can continue to force isolation by treating the heavily textured portions of the picture. The latter course is elected for a couple of reasons. First, in a domain which is as featureless in available parameters as the one here, we are not likely to achieve segmentation of busy areas by elimination of surrounding non-busy regions. Secondly, it is to our advantage to achieve as great a reduction to the scene as possible. This increases chances of splitting remaining parts of the image into a number of closed areas which are more easily processed. Even if resulting regions cannot be further refined a greater degree of partitioning will have been achieved than was available before.

The procedure followed for the treatment of busy regions is basically similar to the one which was just discussed for homogeneous regions. There are three points of difference which can be noted in the flow chart. The first difference occurs in step 45 which requires an averaging of the parameters to get a smoothing effect. This permits further treatment by thresholding, as has been remarked upon earlier. The second point of departure follows step 47 when a transfer to the basic algorithm is made. This ensures that the heavily textured area is treated just as any other subpicture, with the exception that averaged parameters are used to effect thresholding. The reason that this step is taken here, and not for the previous process, is that the scene is heavily textured and the busy areas of the picture are more likely to be composed of regions which are differentiable along some dimension. The third difference encountered in this phase of the procedure is that we don't employ a square plug. There are two reasons for this. First of all, we don't expect to determine boundaries as precisely as before so we don't need the large expansion. Secondly, regions for which parameters are averaged are not as readily separable as in the case of homogeneous areas. So, until the matter can be explored in greater detail, we settle for expanding the extracted busy region one window size to compensate, in a small degree, for the imprecise busy calculation.

Application of the procedure just discussed to the skyline scene produces some interesting results. First the heavily textured regions of appropriate size are extracted from the busy matrix (figure 3.89). The histograms of the averaged data which are derived using this mask are shown in figure 3.90. We observe a number of available peaks which promise a useful segmentation. Thresholding on the basis of limits provided by the saturation histogram and following up with the standard adjustments results in the processed segment shown in figure 3.91. Complete recursive processing of the template of figure 3.89 yields the additional segments shown in figure 3.92.

After complete processing of the forced isolation phase of the new algorithm we

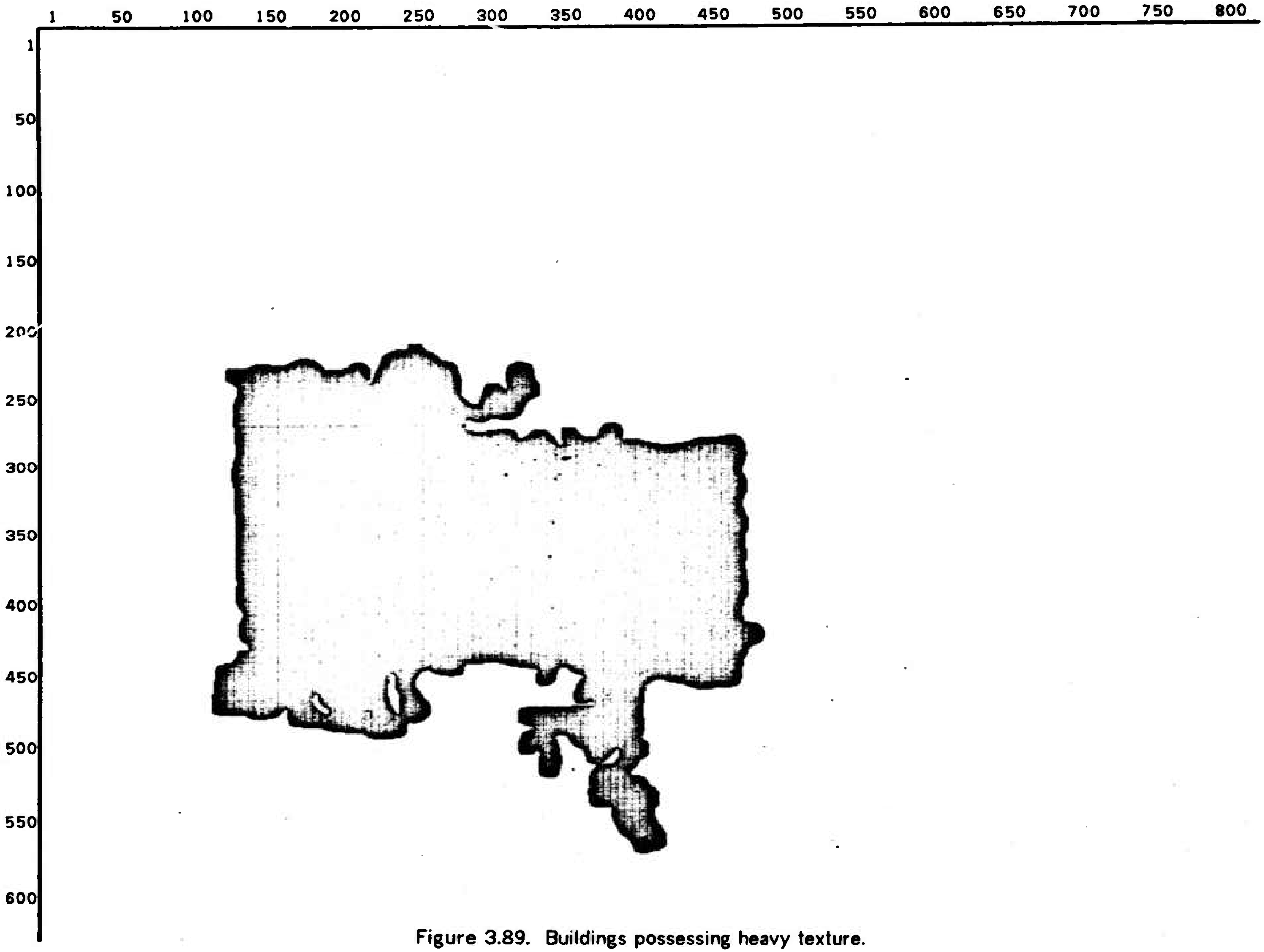


Figure 3.89. Buildings possessing heavy texture.

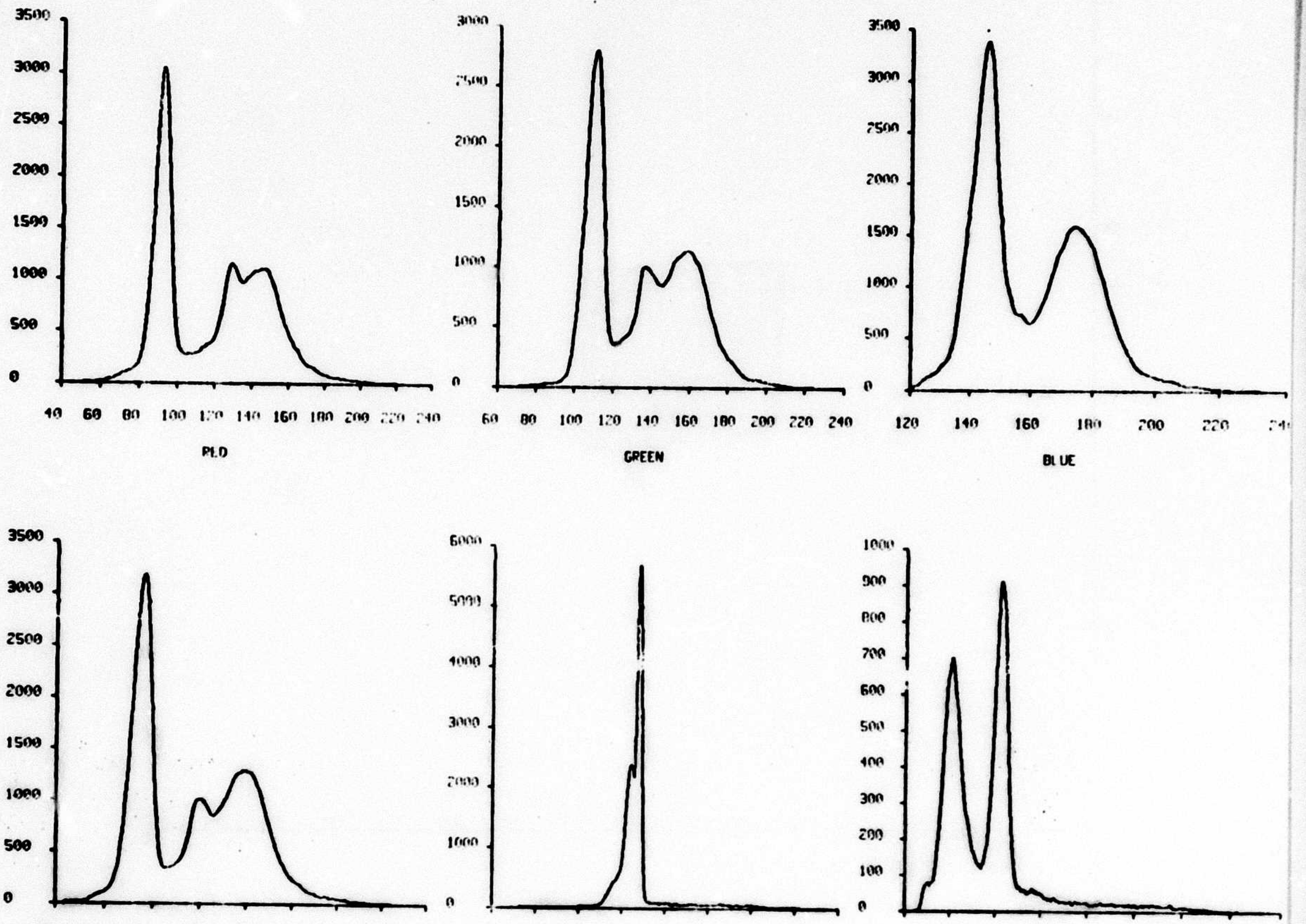


Figure 3.90. Nine parameter histograms of averaged data for figure 3.89.

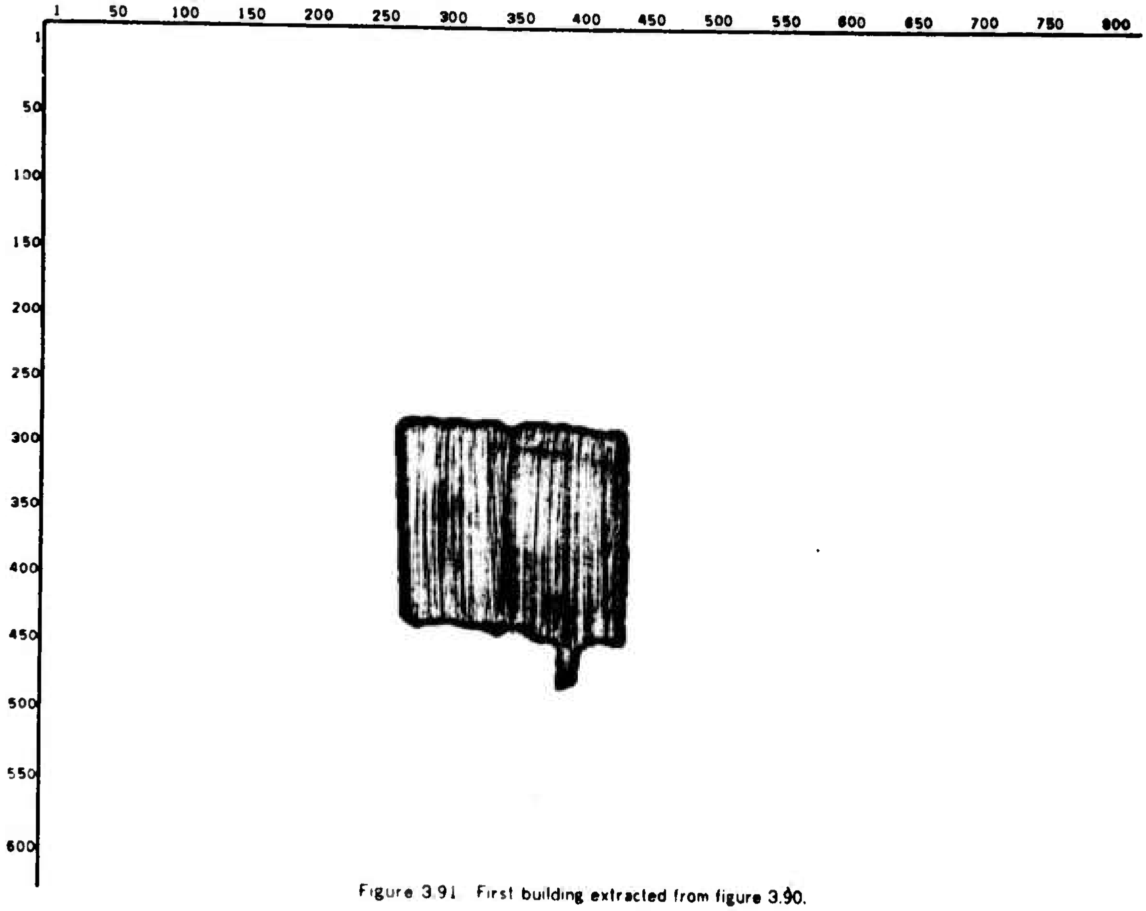


Figure 3.91 First building extracted from figure 3.90.

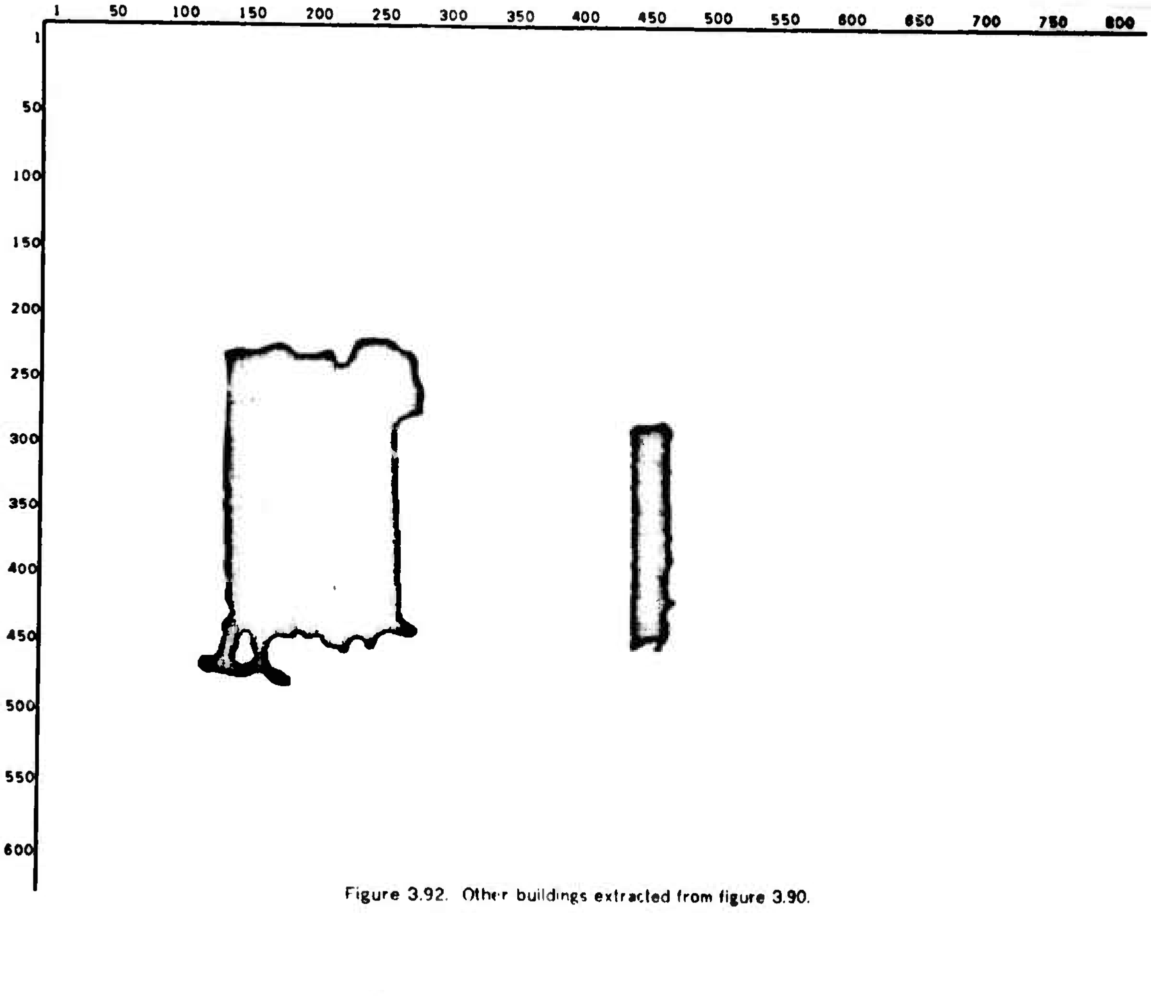


Figure 3.92. Other buildings extracted from figure 3.90.

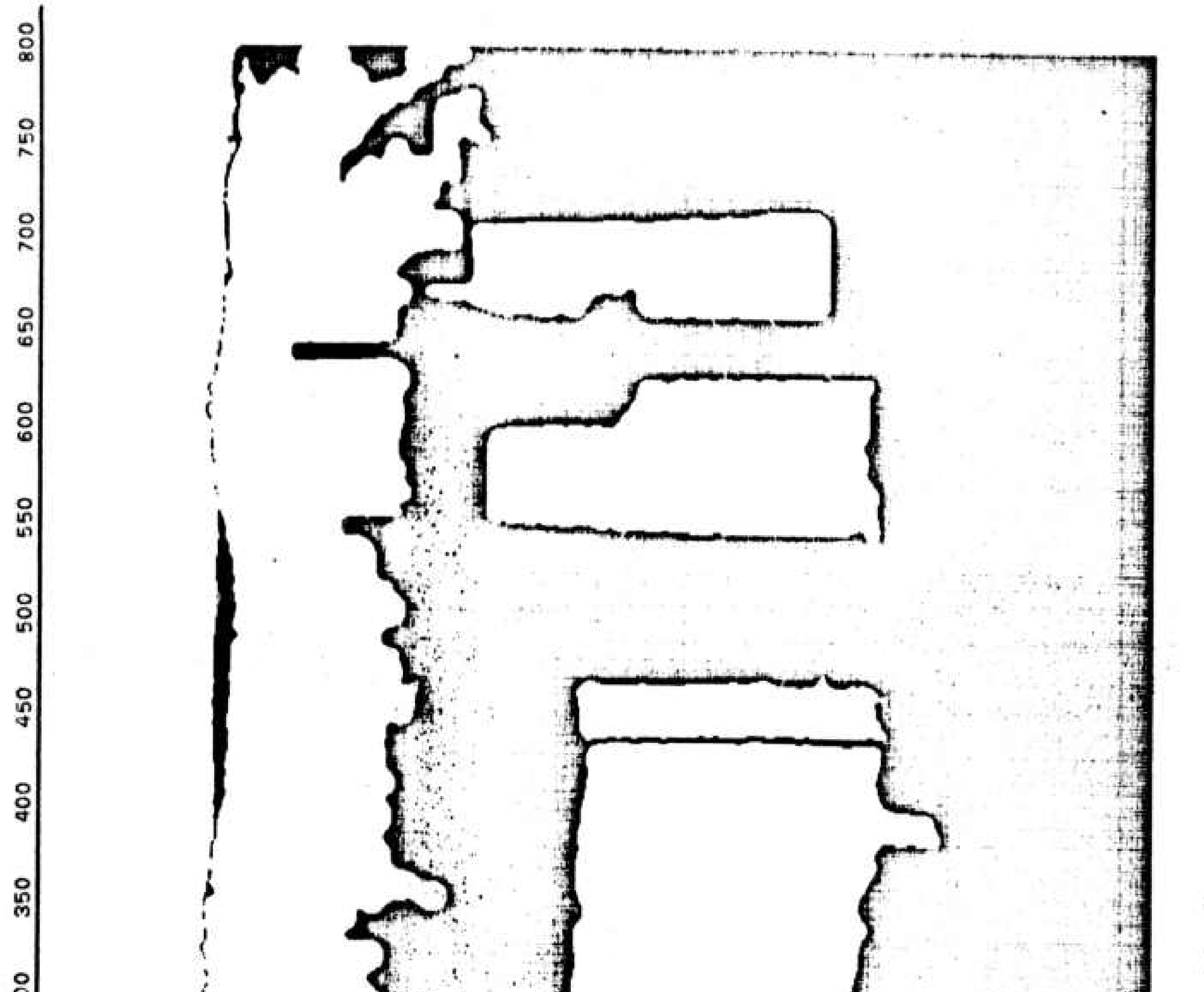


Figure 3.93. Resulting template after masking out light and heavy textured regions.

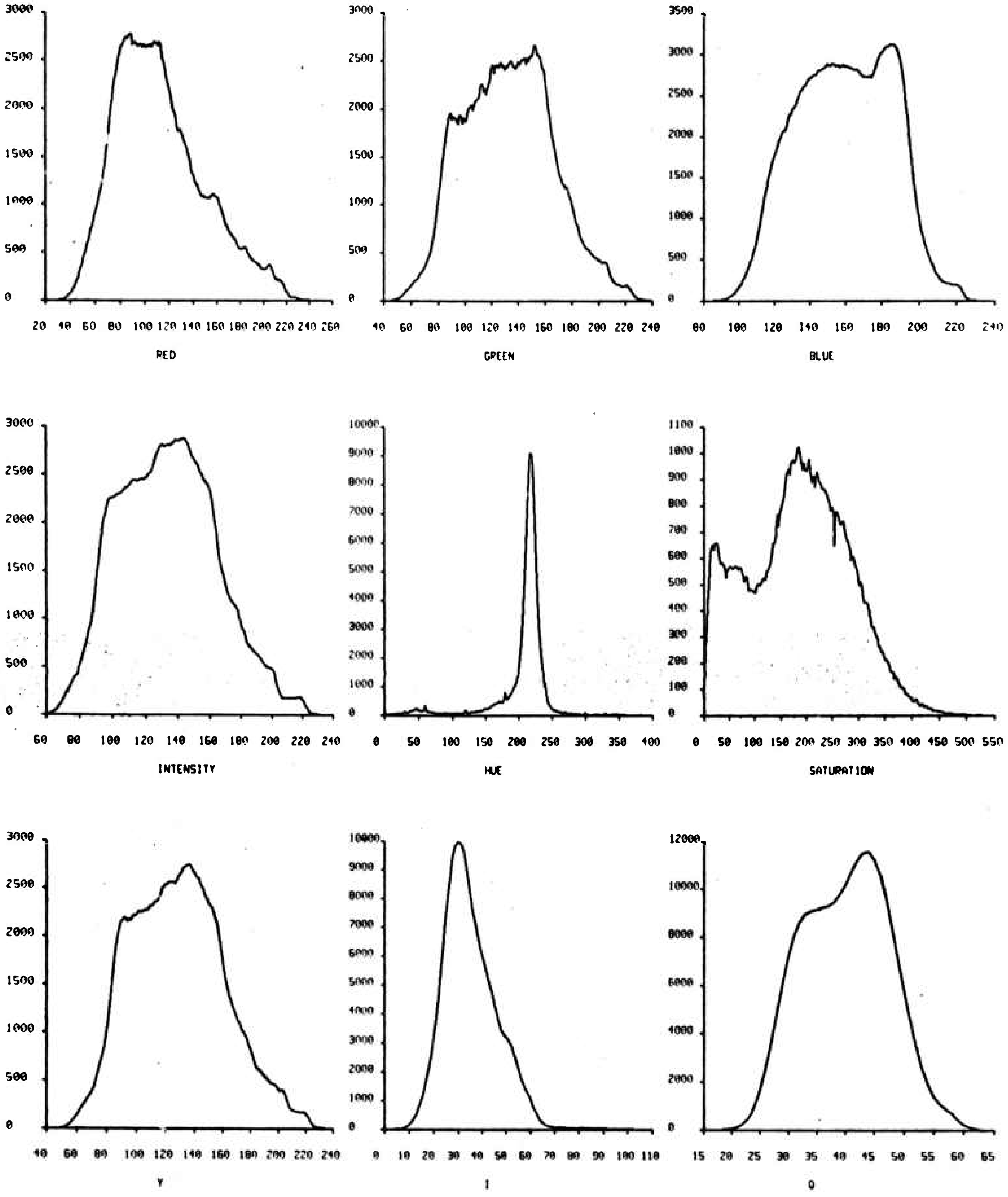


Figure 3.94. Nine parameter histograms of figure 3.92.

are left with the template shown in figure 3.93. Histograms of this result (figure 3.94) indicate a cutoff value for the saturation parameter in the achromatic range of the curve. Following through on this limit we are able to eventually produce the segmentation of the skyline shown in the results section below. This concludes our discussion of the development of the basic algorithm.

The general segmentation procedure has been developed over the domains of three very different types of scenes. It has reached the state where, in our opinion, it can produce very useful results for a fairly wide class of scenes. To check its applicability we applied the procedure to the additional three scenes in our catalog. The results of these additional segmentations were satisfactory and are shown in the following section.

Results

In this section we want to illustrate the decomposition of the three scenes which were analyzed in detail in the implementation section. We also want to present results obtained for three additional images that were segmented with the final procedure. The scenes are presented in the series of pictures which follow. The original scene is shown and followed with a proof sheet that shows the decomposition of the scene. Another photograph which outlines the extracted regions in white will come after this. Not all segments extracted at each level are presented. The purpose was rather to demonstrate the path that recursive descent followed and give some idea of the kind of partitioning we were able to achieve. We apologize for the smallness of the images, but it did not seem to be appropriate to add a lot of additional photographs to a dissertation already overburdened with figures. An appendix which is to be published separately will show the decomposition of all scenes in great detail.

Segmentation

Notice the great amount of detail that is obtained in the decompositions of the room and house scenes. This is due to the richness of color and the high resolution of the digitized pictures. It is our belief that the algorithm will function equally well for any scene possessing this variety of information in any measureable parameter.

In the decompostion of the skyline the breakdown of the homogeneous areas can be observed on the second row and the breakdown of the textured regions on the third row. The fourth image from the right in the fourth row of the same picture shows that we were not very successful in separating all the buildings in the background. The result is still a useful first order approximation.

The decomposition of the girl shows that we were not able to differentiate her blouse from the wall. This is a good example of the problems that arise, even in simple scenes, when we do not have sufficient discrimination among the parameters. This will arise time and time again in any segmentation process that considers a wide range of scenes. If a range map had been available the separation could have been made. This is just a case of having sufficient sensory sources of information. The converse is also true, if fewer parameters are available we can expect to extract much less information from a given picture. In the present circumstances we would have to rely on higher level knowledge using available mechanisms to derive the isolation desired. Notice that in the same decomposition we have shown the extraction of the eyes, mouth, and teeth. This is not actually achieved in the original extraction. The eyes are texture areas that are too small for consideration at a lower level. The mouth was not isolated from the face because the discontinuity in the histograms were not sharp enough to warrant the further refinement. The results show, however, that the finer segmentation can be easily attained if the proper motivation from higher level knowledge is available.

The car scene decomposes on a fairly gross level. In our opinion, this is what is wanted at a first level of segmentaion. As we pointed out earlier additional refinement can result in fragmentation which makes the recognition much more difficult. If finer detail is sought, higher level knowledge can supply the proper direction.

The decomposition of the bear is a very interesting result. Observe that the rocks are separated out on a first level by discriminating on the saturation parameter. This is making use of the special knowledge that we discussed earlier. A further refinement is then obtained on the basis of hue. Considering the lack of structure in the scene and the heavy texture, we believe the segmentation to be quite a good one. There does remain the problem of separating the darker portion of the rocks from the body of the bear. There is also the difficulty of associating the small white portions of the picture with the bear. The latter problem should be much easier than the first to solve.

Time and Space

It should be clear that the large scale pictures, the time sharing system of the PDP-10, the number of sensory parameters, and the variety of picture operations all contribute to a system requiring large amounts of storage space and heavy expenditures of computational time. We have summarized the time and space requirements for the segmentation of the skyline scene:

number of bits accessed = 10^9,
number of bits stored = 10^8,
number of operations = 385,
total CPU time = 9 hours (approximately).

Operators:

	Histogram	Smoothing	Region Extraction
no. of ops.	216	92	20
% of time	23%	66%	7%

	Masking	Thresholding	Misc.	Totals
no. of ops.	20	27	10	385
% of time	1%	2%	1%	100%

The smoothing operations listed include the contraction and expansion operators.

Heavy I/O requirements increases the real time processing to a factor of 2 to 3 times the CPU time. Thus, we are talking about 18 hours or more to process a fairly complex scene. Complete automation would add another substantial increase to the total time requirement. This would not be due to supervisory overhead, but rather to the necessity of executing every step in the algorithm. The experimenter can occasionally skip steps that will not affect the outcome of the process. For example, it makes no sense to perform an expansion and contraction at a given point of the execution if they will produce no effect. The user can also direct the extraction routine to skip point clusters that are clearly too small to qualify as a processed segment. A machine supervisor, on the other hand, must extract all regions to see which ones qualify for acceptance.

If the segmentation scheme presented in this chapter is to find some practical application, speed-ups in time and reductions in space requirements will be necessary. Discussion relating to these issues is presented in chapter 5.

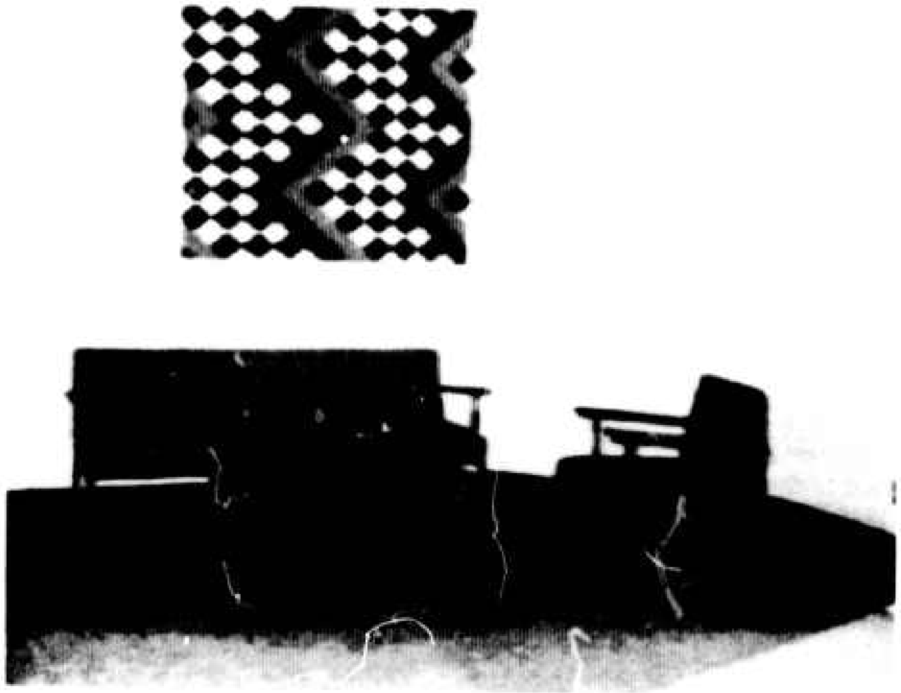


Figure 3.95. Room scene.

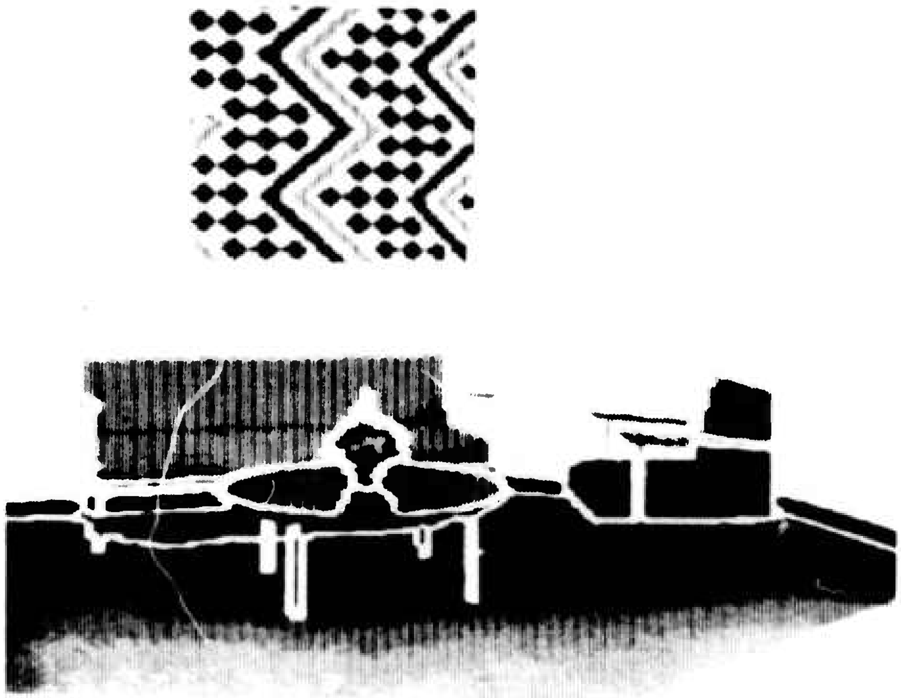
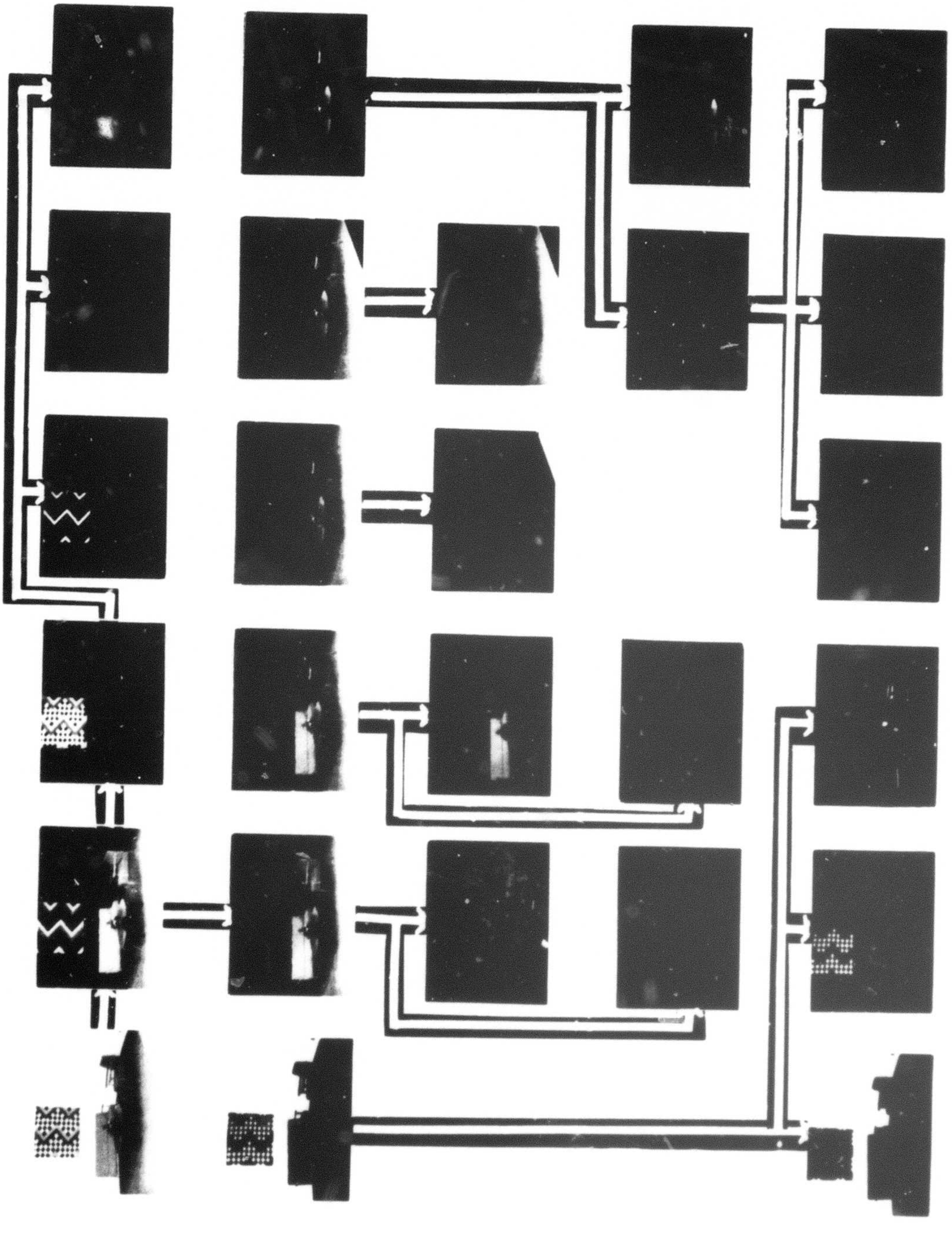


Figure 3.96. Resultant segmentation of the room scene.



Computer Decomposition of a Room into Component Regions

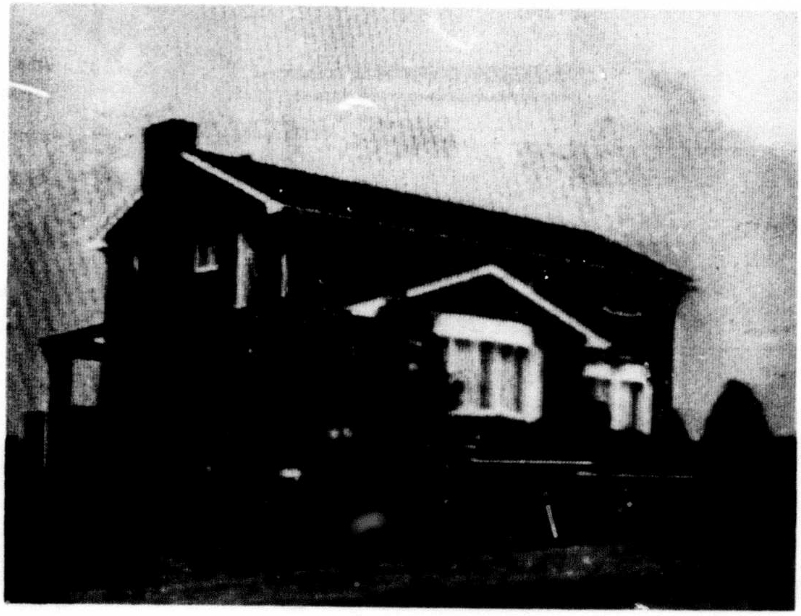


Figure 3.98. House scene.

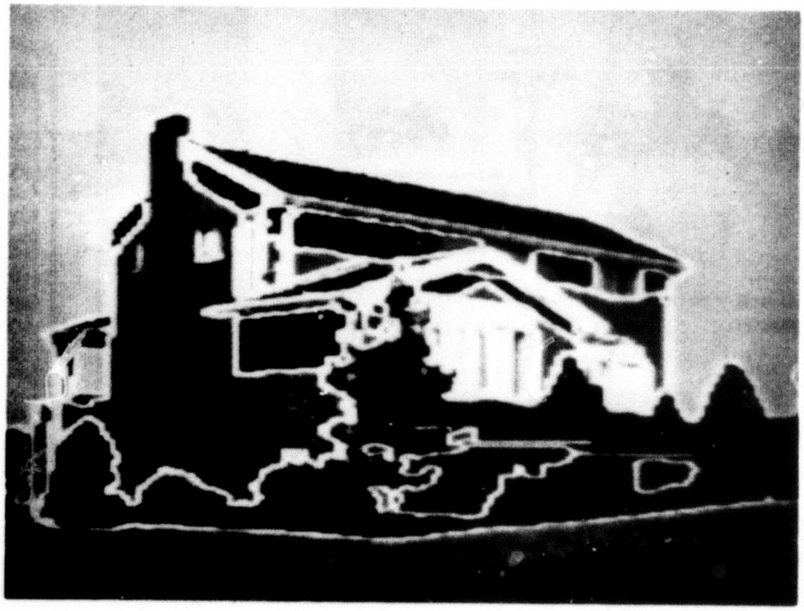
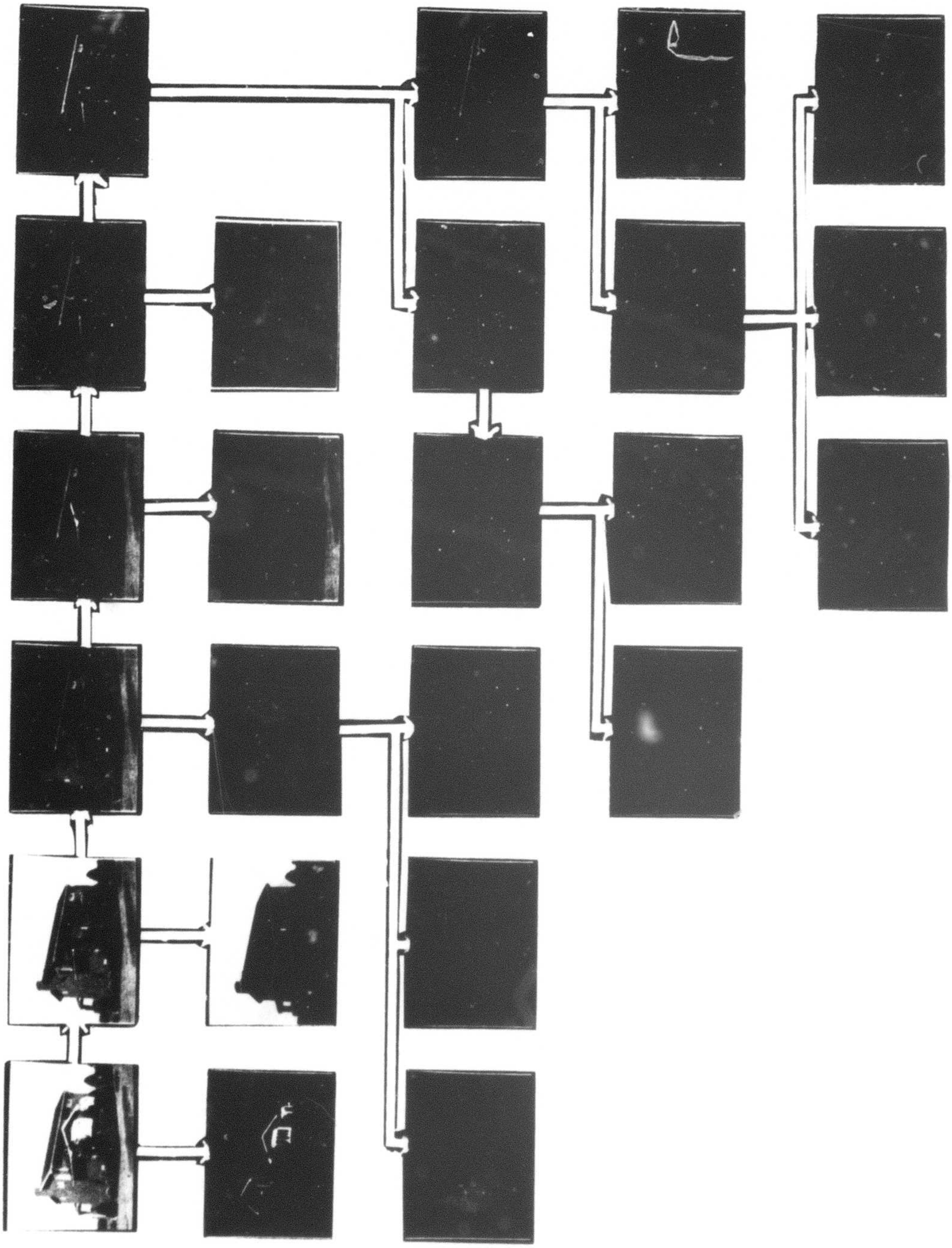


Figure 3.99. Resultant segmentation of house scene.



Computer Decomposition of a House into Component Regions

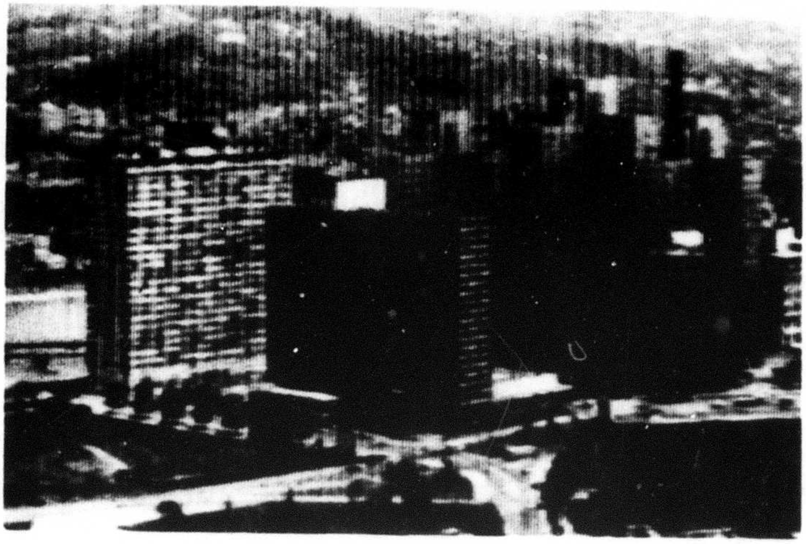


Figure 3.101. Skyline scene.

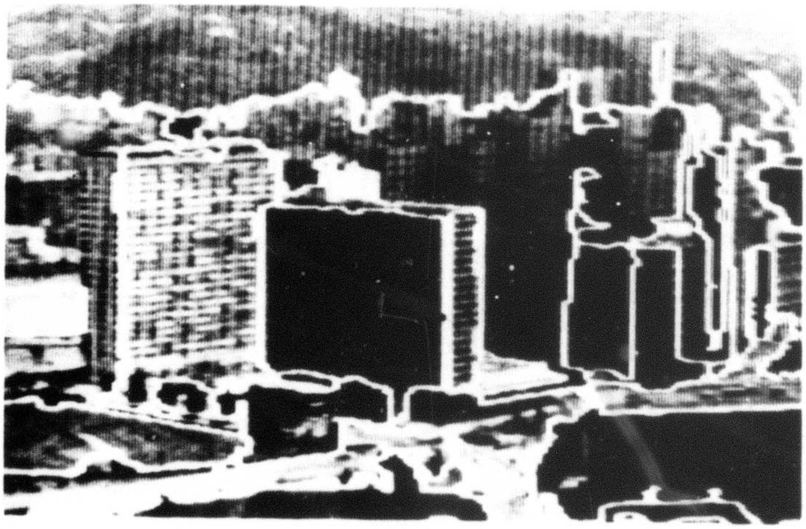
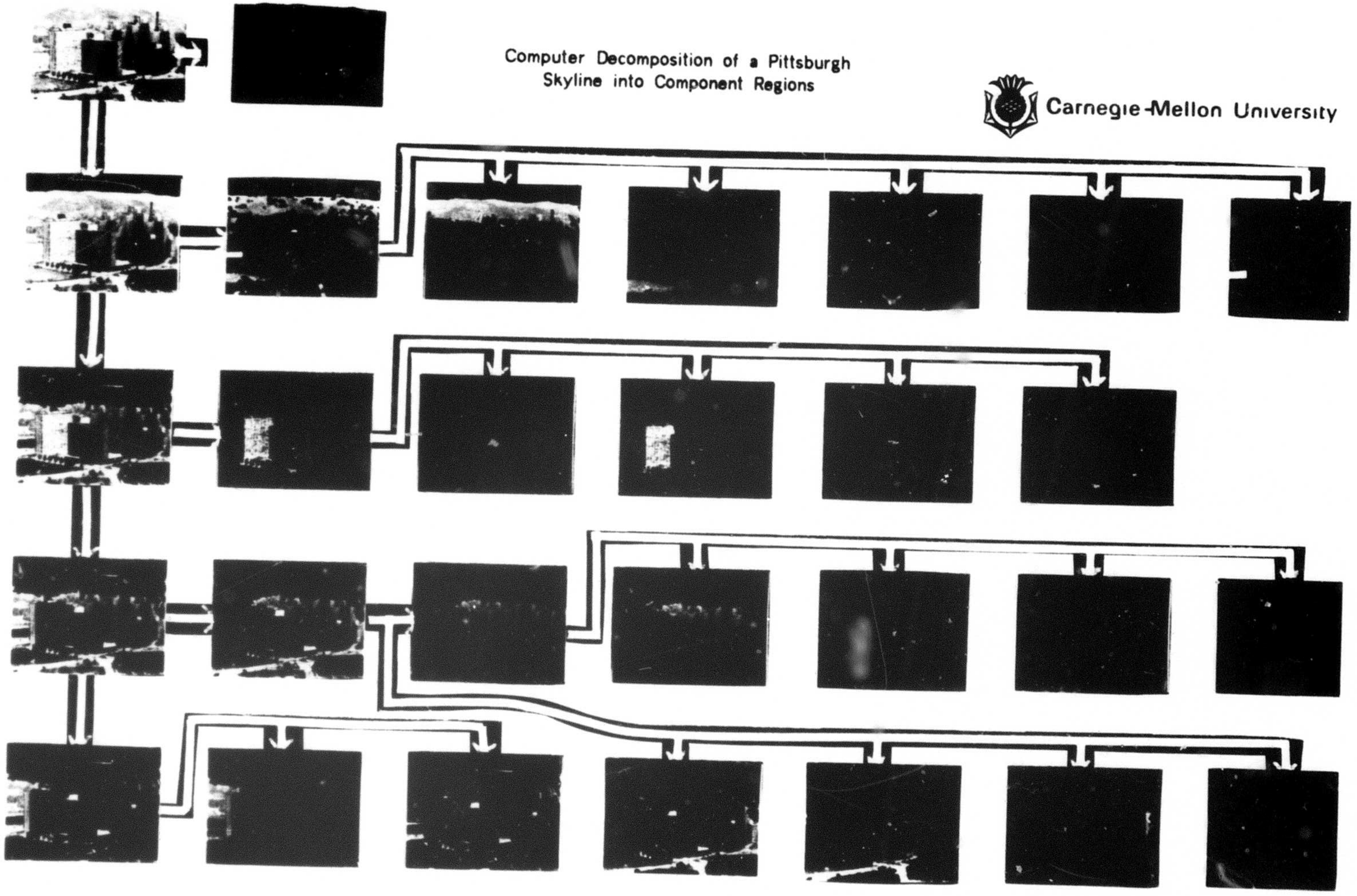


Figure 3.102. Resultant segmentation of skyline scene.

Computer Decomposition of a Pittsburgh
Skyline into Component Regions

Carnegie-Mellon University



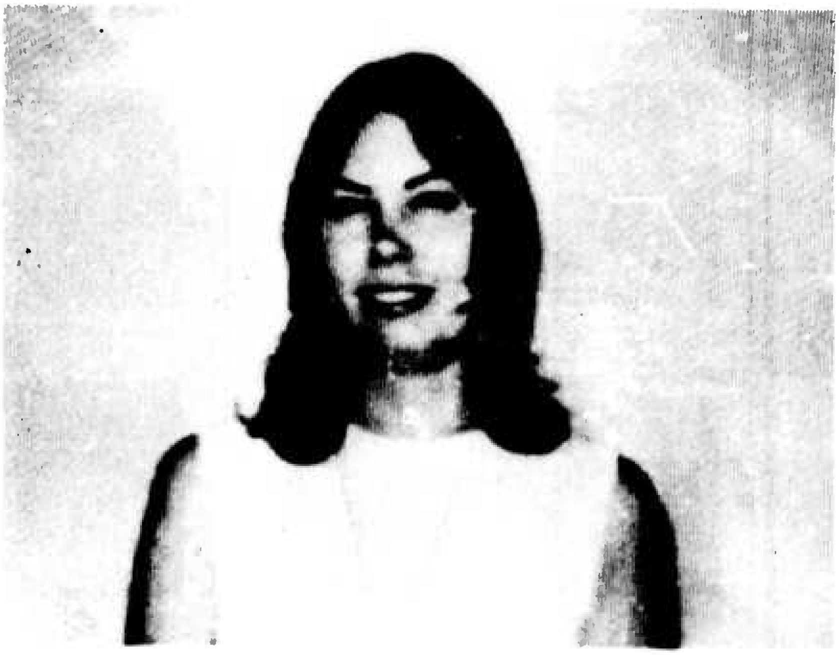


Figure 3.104. Girl scene.

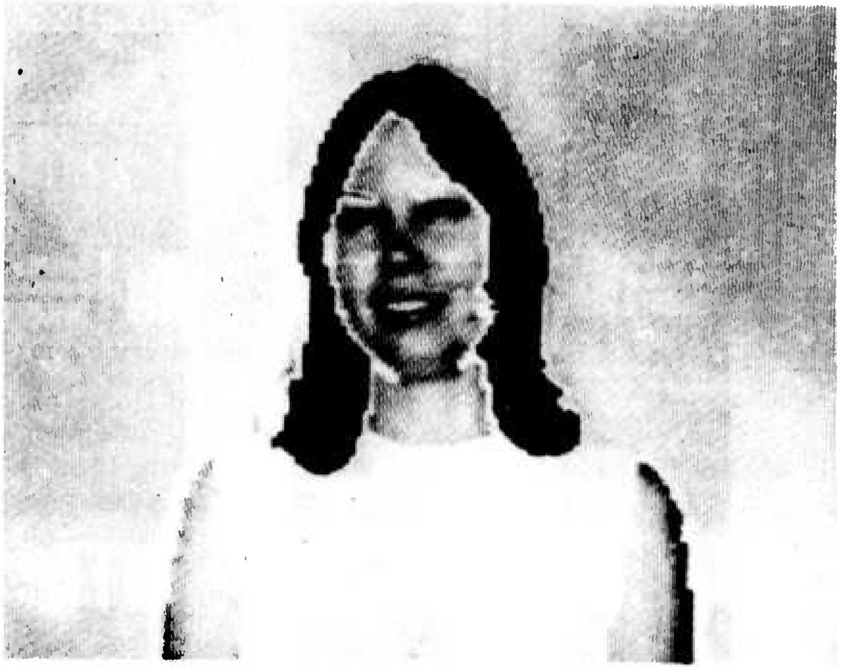
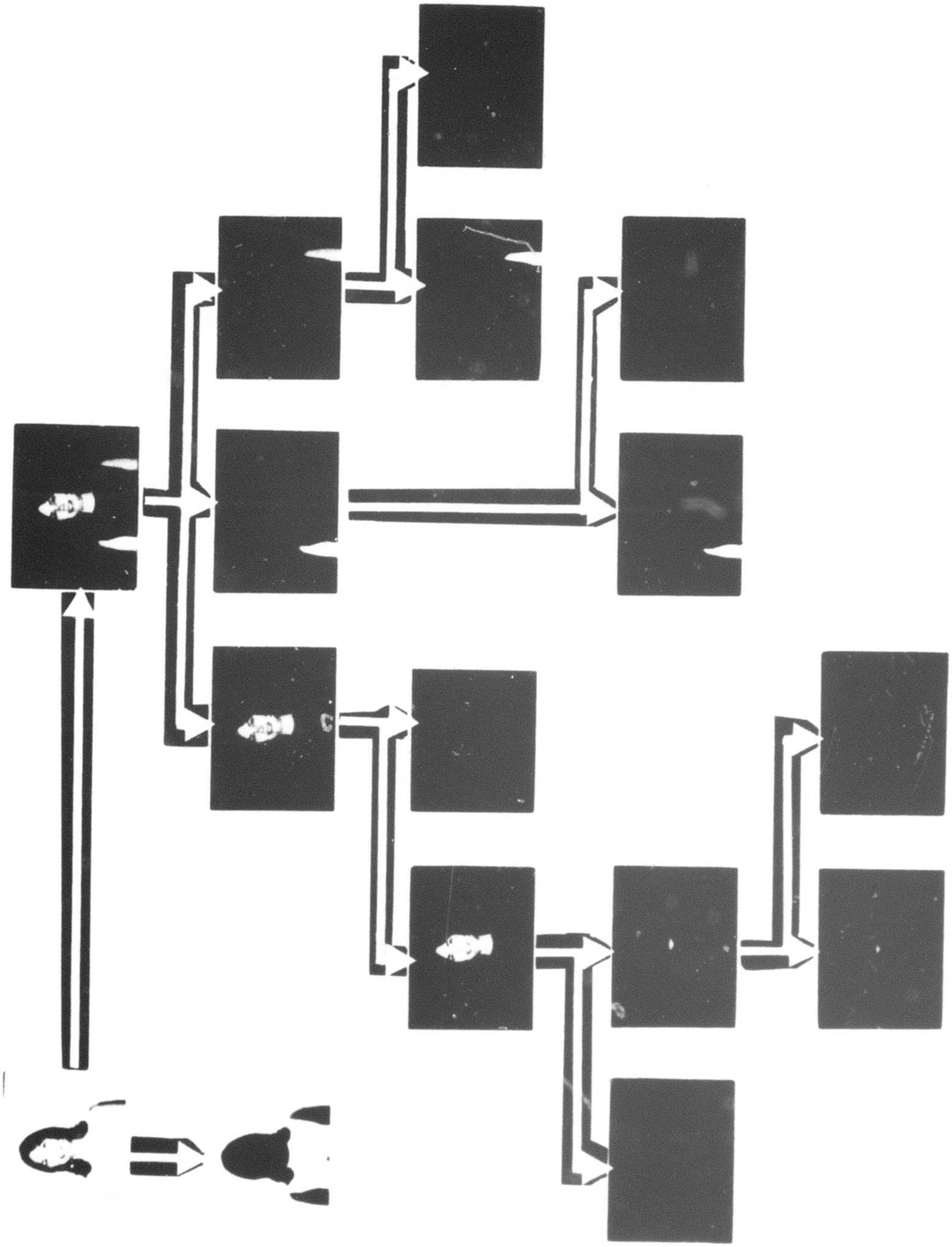


Figure 3.105. Resultant segmentation of girl scene.



Computer Decomposition of a Face into Component Regions

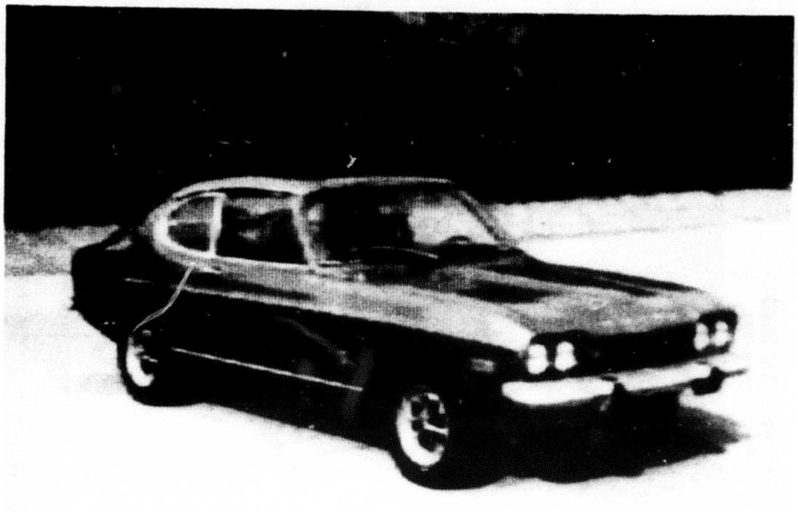


Figure 3.107. Car scene.

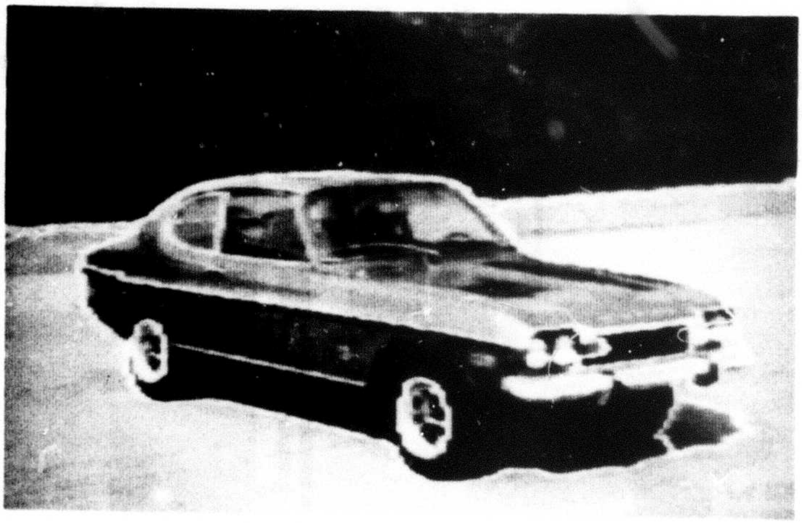
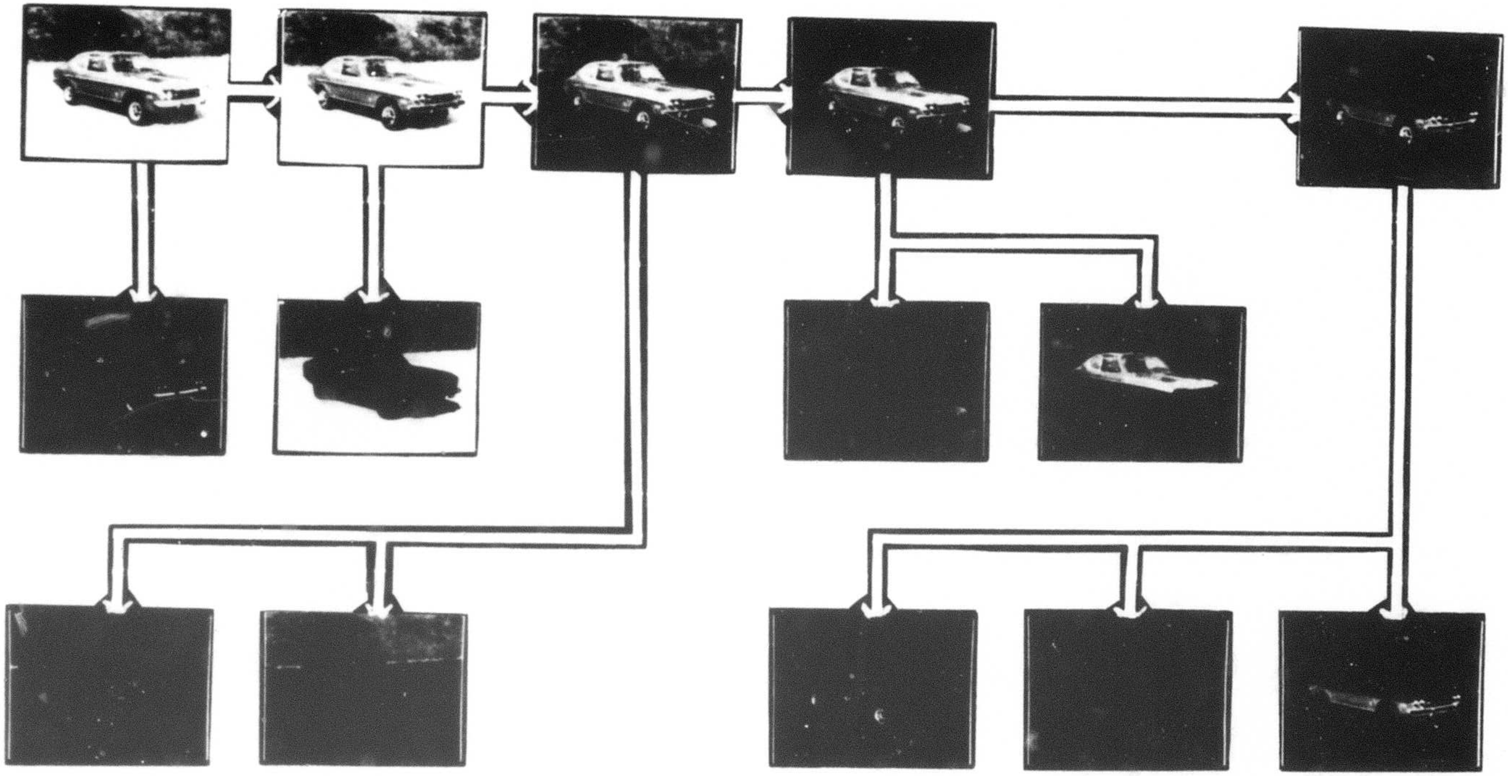


Figure 3.108. Resultant segmentation of car scene.



Computer Decomposition of a Car into Component Regions

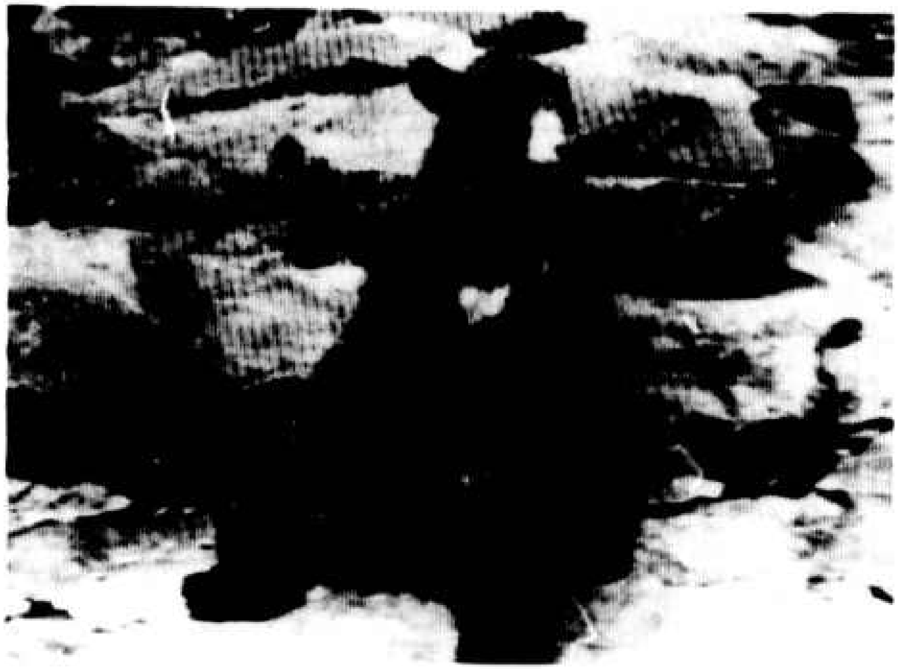


Figure 3.110. Bear scene.

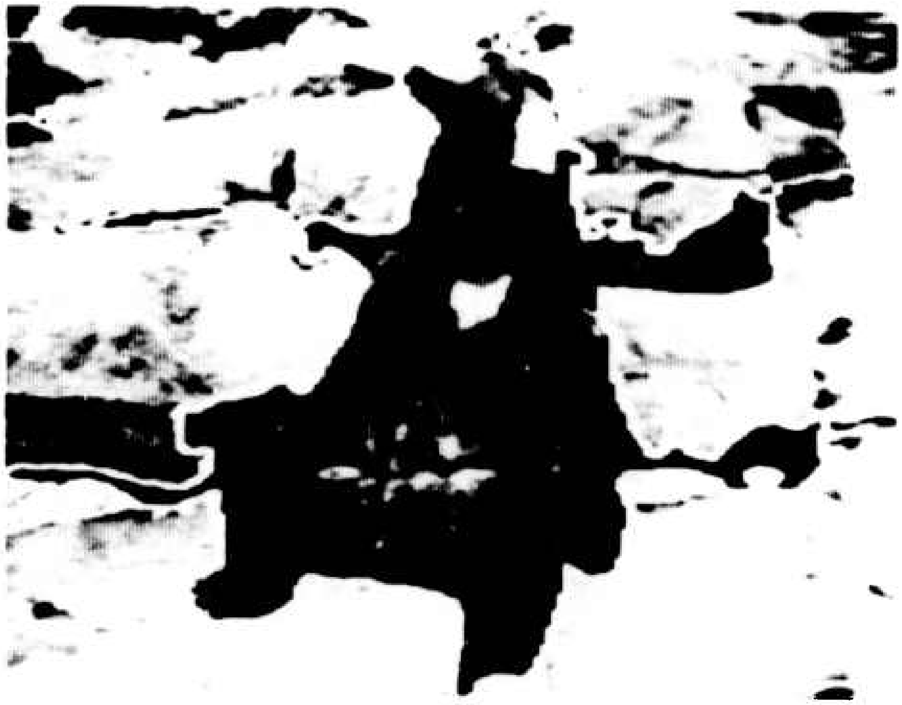
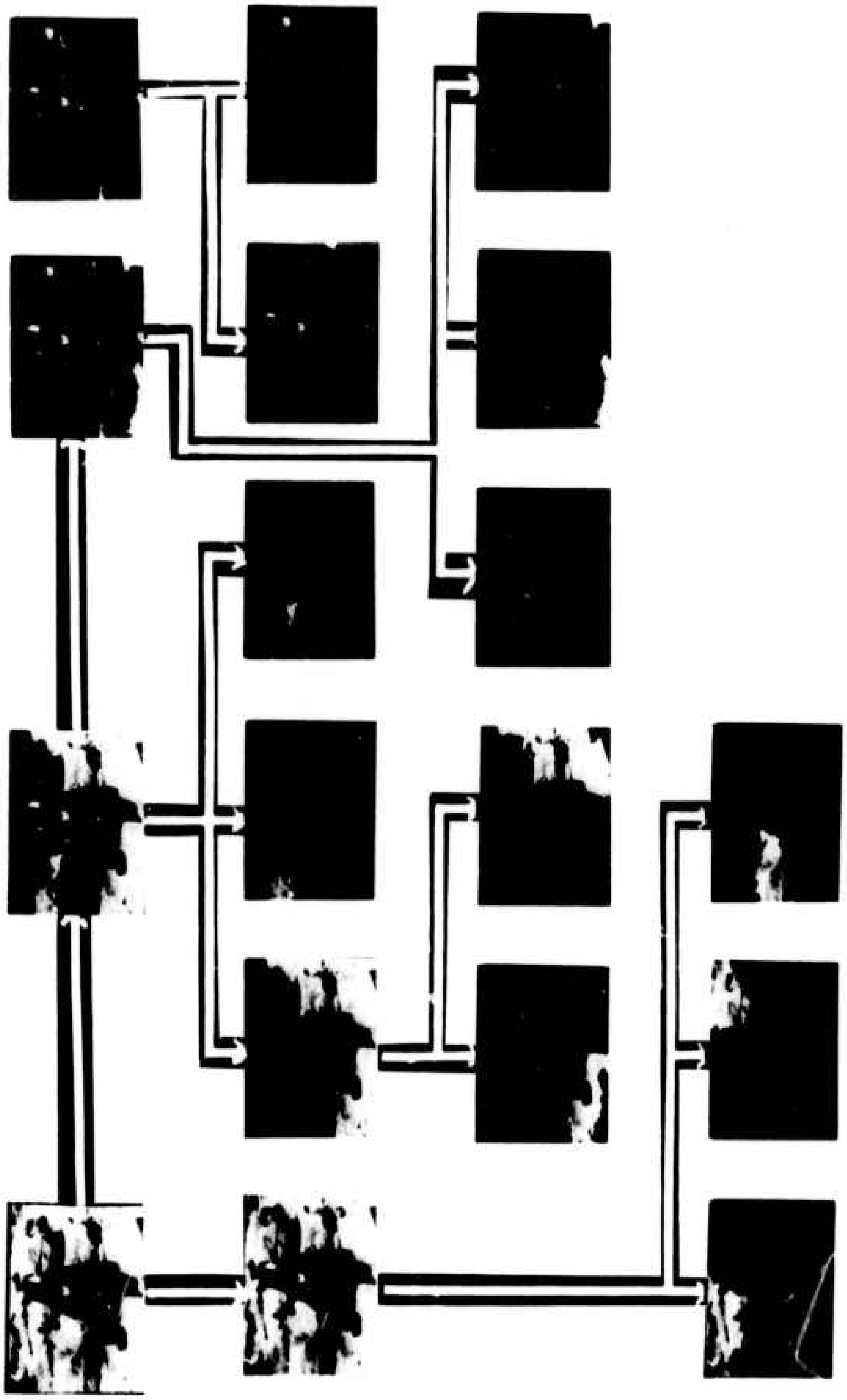


Figure 3.111. Resultant segmentation of bear scene.



Computer Decomposition of a Bear into Component Regions

4 OCCLUSIONS, SHADOWS and HIGHLIGHTS

In all but the most sterile of natural scenes there exist two ubiquitious phenomena which can play a prominant role in subsequent analysis. These are occurrences of shadows or highlights and occlusions. We speak of two conditions only because we regard shadows and highlights as obverse faces of the same problem, i.e., variations in lighting from the standard. This is, of course, an over simplification, but one which can be made at this stage of the research. In this chapter we want to discuss some of the problems in scene understanding that arise because of the presence of one or both of these conditions, and what can be done to alleviate their effects. What we have to say will not be especially startling, but does lay a foundation which will provide a basis for ongoing research. Thus, we are making a first attempt to come to grips with issues which have been recognized in the past, but which have not been carefully defined nor systematically treated for natural scenes.

Some investigators have contended with problems of occlusions and shadows, mostly in block environments. Guzman (1968), Waltz (1972), and Grape (1973) are among those who have constructed systems for the block world which segment scenes correctly in spite of instances of occlusion. They do not, however, explicitly discuss the problems involved nor do they specifically identify the existence of the condition. Waltz was also able to partition line drawings of polyhedral shapes with shadows, while identifying shadow lines and shaded areas. He accomplished this by using a light source model and judicious case analysis of vertices. In the domain of more complex scenes the contributions have been even more restricted. Yakimovsky (1973) isolates a shadowed region in one of his road scenes but the process has no applicability for a general treatment of the problem. Lieberman (1974) makes an occlusion inference when sky segments are detected among trees but this is for a single instance in a single scene.

From our viewpoint there are two basic issues that arise concerning shadows, highlights, and occlusions and the role they play in scene understanding. The most fundamental question is how they affect the identification of simple objects. Highly specialized recognizers for very rigid scene environments will probably not require elaborate steps to achieve identification. More general types of systems, however, must rely on matching extracted regional features and relations with models that embody the knowledge of the real world (Yakimovsky, 1973; Tenenbaum, 1974). If we construct such models for a given scene, what happens if we alter the positioning slightly? What can we say about a large shadowed area that may have appeared? Can we differentiate a desired object on a larger background from possible shadows? Can we still identify a region that has taken on a different shape due to an occlusion? Must we then construct additional models to recognize the new structure? If such is the case, what effect will another alteration have? We cannot possibly model all structural variations for even a single class of scenes. To achieve some generality we must provide mechanisms which can reduce sensory data to common structures which can be matched against some reasonable set of models that determine an object. A general system should attain a similar degree of understanding for scenes in which a table in a room (fiqure 2.4.b) occupies different positions, or in which a house (figure 2.4.c) is photographed at different times of day. We should be able to achieve this goal without formulation of new models for each occurence of variations of this type.

A second issue concerning the influence of occlusions, shadows and highlights upon scene analysis is whether the existence of such a condition need be identified if it does not interfere with the recognition of major areas of interest. For example, if we manage to segment out the entire floor area in figure 2.4.b, and identify it as such should we be concerned with the fact that parts of it are shaded? Or, should we be concerned with the establishment of the fact that various shrubs occlude the side of the house if we have already recognized the basic structure? The answer to these questions largely depends upon design goals and the power of the system. If only specific objects are to be identified, then the matter of shadows may only be of concern if they hinder the identification of those objects. If an understanding on the order of that achieved by humans beings is desired, then shadowed areas must be delineated and identified, and occlusions recognized. This can only be achieved, however, within the limits of the system's ability to discriminate areas of concern. Our own experience has been that we can detect occlusions where clues are clear cut, and shadows which are fairly large and moderately heavy. Some of the lighter shadows on the rug of figure 2.4.b, for instance, elude our best efforts. The problem has been mainly one of segmentation; If a shadowed or highlighted area can be isolated, it can be detected. We have not yet succeeded in constructing higher levels of knowledge which can make use of lighting sources and established locations of objects to more carefully direct searches for areas of slight variation. Nor can we reconstruct hidden surfaces for which no direct evidence of shape is provided to the viewer.

If the issues raised above are to be treated successfully, the required sources of knowledge must, first of all, be able to detect the existence of an occlusion, shadow or highlight. Then, if some adjustment is to be made to compensate for the particular effects caused by the condition, the representation of the affected region must be altered in some way. In order to accomplish these ends, certain pictorial features have to be identified which will trigger a response that corrects the problem. Case analysis provides a methodology for formally classifying invariants that can force an action for a specific type of condition. Just how classification is accomplished for occlusions, shadows and highlights is the subject of the discussion which follows.

In what follows we will make constant use of the term "region". We will employ this term in two senses. The intended sense of the term will usually be made clear by the context in which it is used. In one case we shall be directly referring to actual sections of the scene which are of interest because they possess certain attributes. The attributes may class the area as a distinct object or simply as a part of the scene possessing uniformity over some number of parameters (e.g., color). At other times we will mean by region some structure contained in the global data base which summarizes our knowledge of a closed area of the scene in question. The structure makes specific reference to the actual scene through a boundary given in some form of picture coordinates. The knowledge consists of properties and relationships which are thought to be important for a proper representation of the actual portion of the image.

In keeping with our proposed model, we want to treat the general issues from the standpoint of implementation through sources of knowledge. For this reason we will divide the remainder of the chapter into two main sections: knowledge necessary to satisfy design goals, and control structures needed for to convert knowledge into action.

Required Knowledge

There is a good deal of overlap in the kinds of knowledge required for occulsions and for shadows and highlights, which is why they are being treated together in the same chapter. The two subsections that follow consider some pragmatics which can be employed to reduce the effects of the two conditions.

Knowledge about Occlusion

Occlusion is defined in Webster's New World Dictionary as "the prevention of the passage of (something) by closure or blockage". In the case of vision we construe this to be a blockage of light rays or a shutting off from view. In this sense every object in existence occludes something else. In order to talk about this condition intelligently we must further restrict the definition to the precise frame of reference provided by the limits of extent of the scene under analysis. We also require that there exists sensory evidence of the occlusion. The one exception to this stipulation is for those objects in a scene, completely hidden from view, for which there is strong evidence of existence and which would be observable if an occluding object were moved. For example, since our world model tells us that sofas have legs in each corner, it is reasonable to suppose that the table in figure 2.4.b is completely hiding the right rear one.

We can further restrict our task domain by excluding from consideration certain intances which conform to the specifications stated thus far. Although the baseboards of figure 2.4.b and the shutters of figure 2.4.c fit the definition, they have a number of properties which prompt us to treat them as separate entities not amenable to occlusion analysis. Their semi-permanent nature and particular function suggest a fixed relationship with their underlying structures. In a sense they can be treated as a part of that structure. They differ from the hedges of figure 2.4.c, which are also semi-permanently fixed, in that the latter may occur anywhere on the ground surface.

Occlusion is a three-dimensional condition and regions refer to areas of pictures which are two-dimensional representations of real world objects. In the discussion to follow when we speak of a region as being occluded it should be understood that we are alluding to the actual object represented by that region.

Detection Issues

As was intimated earlier, we want to develop the main argument by means of a case analysis. Before attempting this, however, a clear understanding of the available knowledge facts is in order. We must know what pictorial clues signal the possibility of an occlusion. These clues could be embedded in the world model by denoting which objects are likely to function as occluding structures and which are likely to be occluded by others. Walls and floors in indoor scenes and skies are probable instances of the latter class, while sofas, trees, and shrubs may be of either. Another possibility would be the use of a library of unobstructed shots of all objects to match occluded areas by differencing techniques. The drawbacks inherent to this approach are the necessity for exact registration for proper alignment and the impossiblity of getting clear views of all subjects (we can hardly ask the bear in figure 2.4.e to move). More general and helpful indications of occlusion can be gotten from local clues, i.e., from regional properties and relationships directly extractable from segemented regions without consideration of contextual knowledge from world models. This is not to say that higher level knowledge is not desireable or necessary to a general vision system. We are merely saying that a low level approach to the problem can provide some immediate dividends that can serve as a springboard to further analysis.

Ideally, we would like local clues which would constitute a necessary and sufficient condition for the existence of an occlusion. Unfortunately, these are not readily evident, if indeed they exist at all. There are, however, three clues of this type which constitute a necessary condition for the existence of an occlusion. These are immediate proximity, discontinuity, and dissimilarity. That is, a picture cannot depict one object occluding another unless: their two-dimensional regional representations are in juxtaposition; there is sensory evidence to show that the continuity of shape of one structure may have been interrupted; and the two regions are dissimilar in at least one feature. The one qualification to this statement is that the occluded object be at least partially visible. For instance, the existence of a right rear leg for the sofa in figure 2.4.b might be hypothesized on the basis of world knowledge but it certainly is not supported by visual proof.

"Discontinuity" is a term which is intuitively clear but difficult to define in a precise way. It refers to those properties of a picture which indicate that a uniformity along some dimension has been interrupted. The very fact of the interruption also indicates along what lines we would have expected the boundaries of the occluded region to have continued. These concepts, which must appear somewhat fuzzy at this point, will be illustrated by further explanation and examples given below. For now, consider the sofa in figure 2.4.b. The continuity of color and texture surrounding the vase of flowers supports the conjecture that the flowers hide a portion of the sofa. We can also conclude that the exact portion occluded corresponds to that area determined by the boundary that lies in common and a straight line drawn between the first and last points that the sofa has in common with the vase.

In addition to two-dimensional clues, the three-dimensional property of relative range would be very useful in detecting the presence of occlusions. Let us postulate for the moment that we have relative range information available in the form of some number for those surfaces which are nearly orthogonal to the camera focal axis, and in the form of minimum and maximum values for other surface orientations. Consider some additional inferences that might be made. Transformations from range data and picture coordinates to real world coordinate systems is an easy step (Duda and Hart, 1973) and will yield useful height information. This would perform the same function as range for surfaces of horizontal extent. Range or height disparity between adjacent regions is a strong indicator of occlusion, for it is usually the case that real world object borders overlap in the two-dimensional image. They do not, however, constitute a necessary condition for occlusion, nor, even coupled with the two-dimensional clues, do they constitute a sufficient condition.

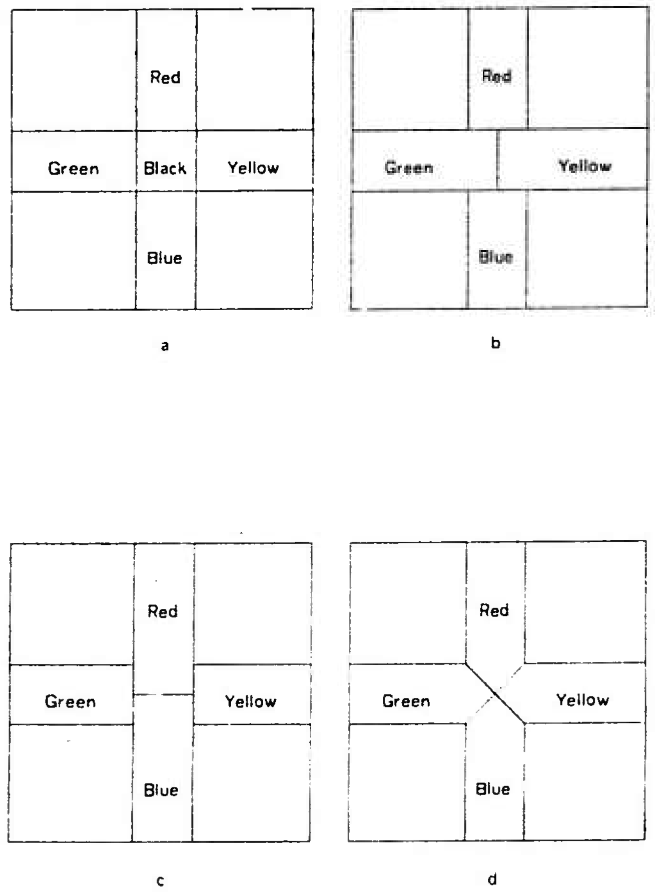


Figure 4.1. Some examples of continuity ambiguities.

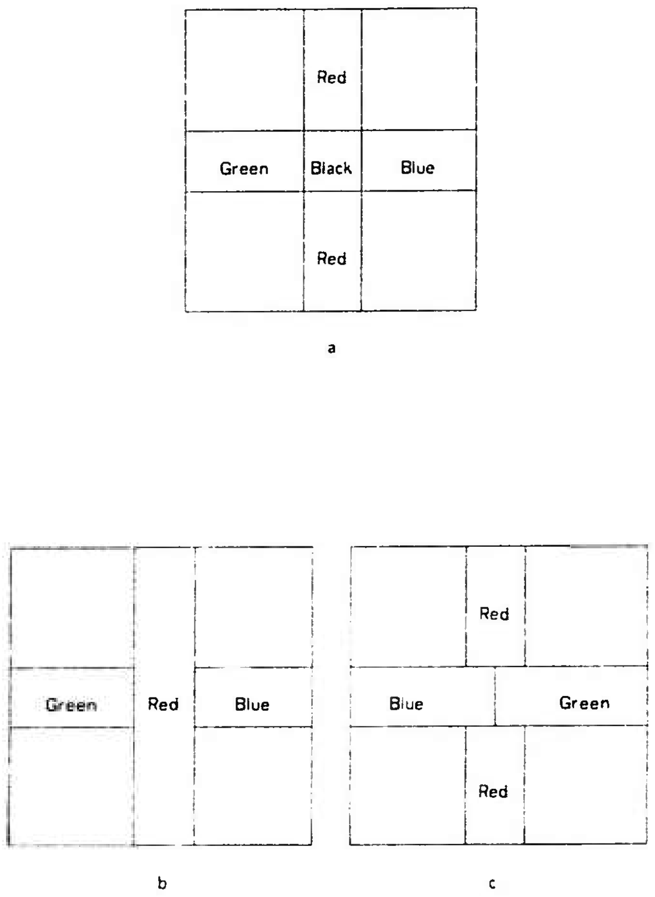


Figure 4.2. More examples of continuity ambiguities.

Restoration Issues

Up to now we have talked about hypothesizing regions which may compensate for occlusion effects, without really examining how such a thing might be accomplished. What we really want is a specification of those features and relationships which would be extracted by the system from a scene in which the actual occlusion is removed. This might be done with unobstructed images of the object in question, but the approach not only suffers from the shortcomings noted above but, it also requires an identification, which was what we were seeking to establish in the first place. It could be accomplished through a synthesis of the necessary values for all parameters over the affected area of the image. The required features could then be extracted in the usual way. Unfortunately, the formidable problems associated with texture synthesis rule out this approach. We have compromised by estimating those properties and relationships which we feel are necessary to the recognition process.

One of the critical tasks in this respect is the correct determination of boundaries. Not only is this important for derivation of new relationships and geometrical properties such as shape, size, height to width ratios, position, etc., but it also provides the strongest visual conformation of correct analysis to the human eye in an interractive system with graphics capability. Correct location of boundaries is highly dependent upon the nature of the objects involved and upon local contextual information, especially indicators of interrupted continuity. For example, consider figure 4.1.a where the simulated black area represents an occlusion. The most reasonable hypothesis might be figure 4.1.b, although figure 4.1.c is certainly possible within the local context. But which hypothesis is best in the contexts of figures 4.2? Should all hypotheses be made? The situation becomes even more confused as we leave the domain of man-made objects. Not even a human can estimate the shape of the rocks behind the bear in figure 2.4.e with any degree of confidence. There are no general solutions to difficulties such as these, but limited pragmatic alternatives which can be given within the framework of an existing system will be discussed below.

Besides regional boundaries, there are a number of additional local features that must be estimated for the affected area. We have already mentioned geometrical and relational properties which usually need to be recomputed because of boundary alterations. Other likely kinds of features (e.g., hue, saturation, intensity, texture) are statistically determined within the specified region and the same measures can be assumed for the region to be hypothesized. After all, the assumption is that the hidden area is similar to the one which is open to view and thought to be occluded. There are always possibilities of peculiar circumstances where the wall behind a framed painting might be of a different color or have a large hole in the plaster, but similarity of features is still the most reasonable hypothesis and errors made in situations of this kind will have to be discovered by later verification in the context of world knowledge. The more difficult task is to establish three-dimensional relationships and modify old two-dimensional relationships.

The Case Analysis

Now that we have established some of the necessary features that must exist for an occlusion to be present in a scene and what can be done to "restore" an occluded area, we can classify the condition as to a number of specific types. The classes are determined along lines of decreasing continuity features. Each case presents its own particular obstacles to detection and restoration of occluded areas. We do not make any claims to an exhaustive consideration of all possible configurations, but we do feel that they cover our chosen scenes and have a wide range of applicability to natural scenes in general. As we discuss each case we will point out ambiguities and difficulties that arise for detection and hypothesization mechanisms and give our own particular choice of action to be taken.

The greater part of the material covered below focuses on the employment of local visual clues to detect occurrences of occlusions. It is these clues that play the predominant role at all levels of analysis. They are the only strong indicators available on a low level basis when identifications are yet to be made. Since they constitute a necessary condition for occlusion, it is also required that they be utilized to verify hypotheses proposed by other knowledge sources. It should be kept in mind, however, that knowledge from world models will be available and could be used to postulate the presence of an occlusion or verify the hypotheses provided from local clues.

Case 1: One region is contained entirely within the boundaries of a second region.

The implicit understanding in this case is that some continuous background surface is interrupted by a smaller region. Examples of this are pictures on a wall or clouds in a sky. An instance which occurs in our own set of pictures is the abstract design, shown in figure 2.4.b, which hangs upon the wall. In some sense this is a degenerate occurrence of the case in question as there is no expanse of wall between the design and upper edge of the image frame. It can still be construed, however, as fitting the definition and it is convenient to assume that the expanse of wall is cut off by the picture border.

Examining this example in terms of the local clues that have been proposed earlier, we see that continuity is expressed by the continuous expanse of wall (occluded region) which surrounds the design (occluding region). It is explicity established by determining that the design has only the wall (or image border) in immediate proximity. Range information, which is negative in the sense that it yields no disparity, offers no further confirmation. At this point in the analysis we have established the possibility of an occlusion (we have detected the necessary condition), but we have no way of knowing whether the design is painted or hung upon the wall (indeed we do not even know that there is a design or wall). Such a decision, of course, constitutes the verification process which might be initiated by some module which embodies knowledge about the real world.

In some cases range data could be a decisive factor in resolving possible ambiguity. For instance, consider a scene which shows a blank wall with a window.

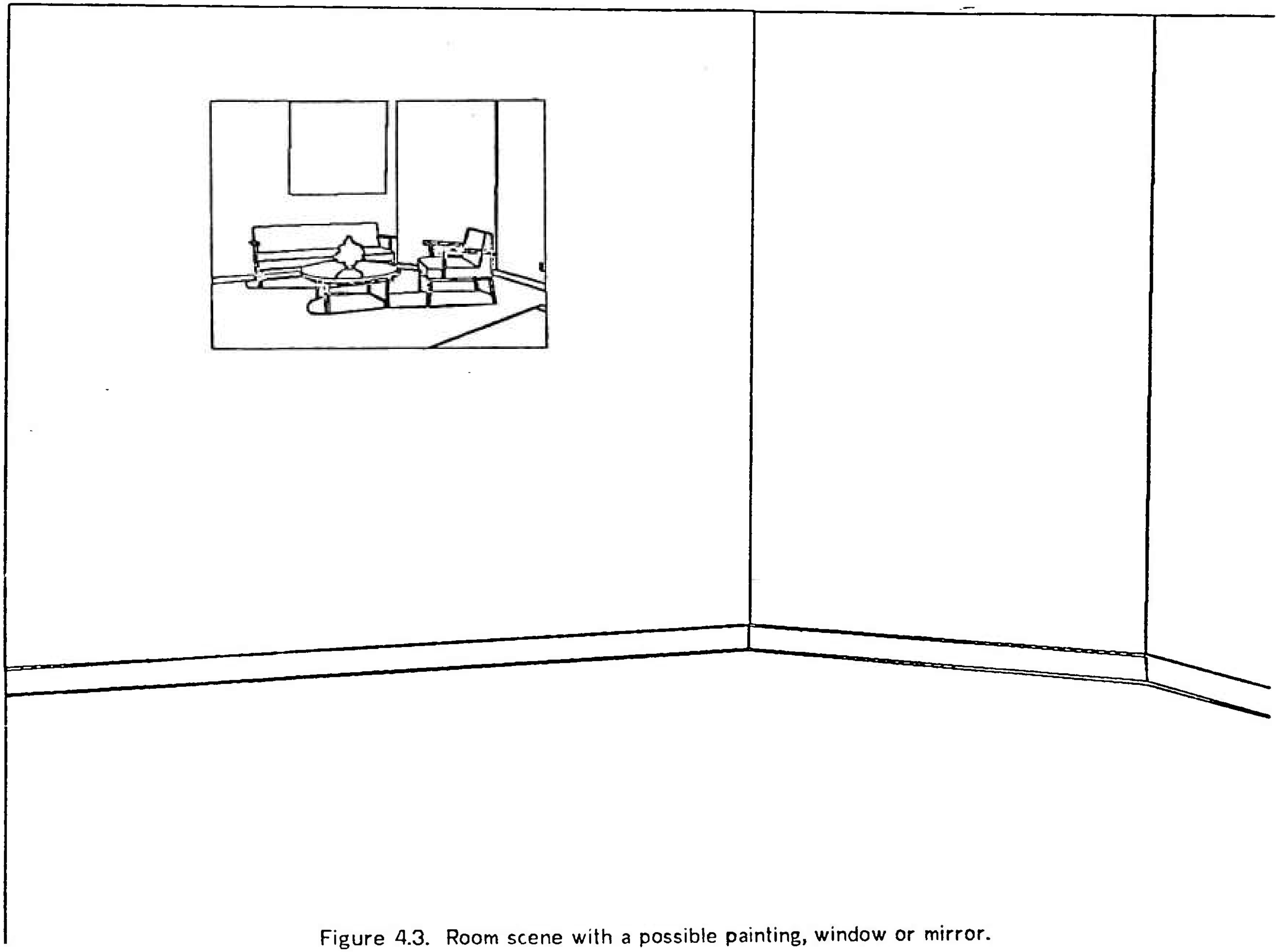


Figure 4.3. Room scene with a possible painting, window or mirror.

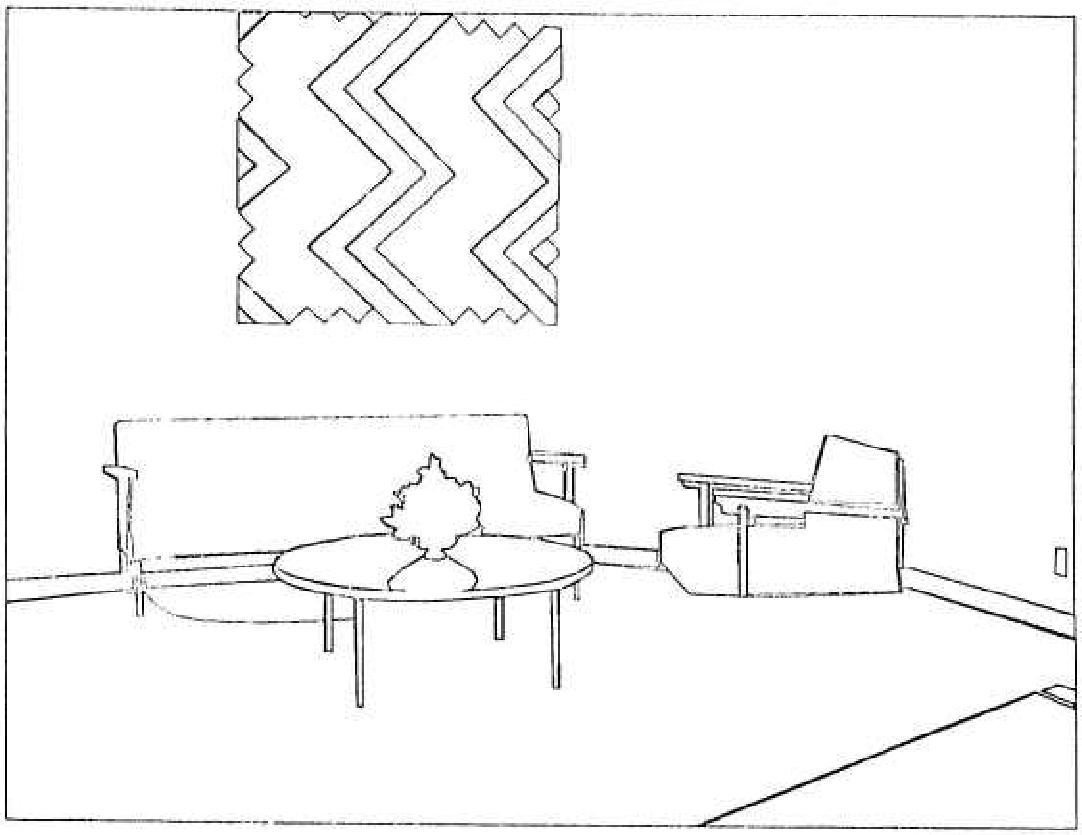


Figure 4.4. Segmentated representation of scene in figure 2.4.b.

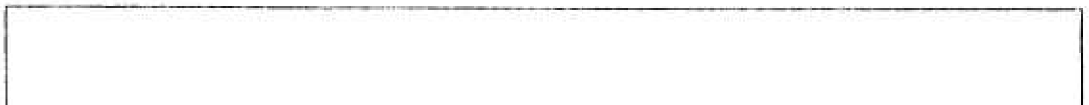


Figure 4.5. Removal of design from figure 4.4.

Through the window another blank wall is visible. Although the two-dimensional clues would be the same, range disparity would certainly decide the issue. If the window had curtains behind it or if glare destroyed transparency, range information would be lacking and we would be faced with the same ambiguity demonstrated above. To explore the problem further, examine figure 4.3, which could represent a line drawing of a room scene where there might be a painting, window, or a mirror in the wall area. Range disparity would resolve the painting-window question, but would not be decisive in disambiguating the window-mirror problem. Remember, it is not a matter of identification that is raised here but rather a question of what occludes what.

We have raised these issues concerning ambiguity to emphasize the difficulty of the basic problem. With the infinite variations of stuctural complexity that exist in the real world, it is painfully evident that the kinds of local clues we can detect do not establish the existence of an occlusion with certainty. Nor do we expect them to. With problems of this magnitude we must restrict our attention to relatively simple environments, as represented by the room and house scenes, in the initial experiments. With this limitation the kind of range ambiguity just described is not an issue. Range and/or height disparties allied with the two-dimensional local clues prove to be decisive. The difficulties involved do emphasize, however, the need for the hypothesize-and-test paradigm. The system must be permitted to hypothesize errorful regions while counting on the model to provide mechanisms to verify the validity of the decision. It may happen that a particular type of ambiguity may occur only in certain scenes so that hypotheses can be verified or rejected, depending on context.

Once the decision is made that there is sufficient evidence to support the conjecture of an occlusion, the hypothesis takes the form of an insertion of a new region, which represents the unoccluded object or surface, into the global data base. For most instances of the current case, boundary reformulation is rather simple. Borders which delimit the occluding region are simply excluded (figure 4.5 is an example of actual recomputation of boundaries for figure 4.4). With few exceptions most other attributes can safely be assumed to be the same as for the occluded region. Note, however, that if circumstances were reversed and the contained region were the occluded body the decision would not be so simple. Considering only the two regions under scrutiny, there are no indicators to provide a basis for extension of the occluded region's boundaries.

Case 2: One region borders a second region on three or more sides but does not completely surround it.

The immediate question that is raised is what is meant by the term "side" in reference to a region? For regularly shaped objects, such as the abstract design of the room scene, it is very clear what is meant by the top, the right, the left, or the bottom of the region that delimits that object. In the case of more amorphous shapes, such as the vase of flowers, it is not so clear where one side leaves off and another starts, or even what is meant by a side. To provide some frame of reference we define for each region an external minimum bounding rectangle (MBR) oriented such that its sides lie in the vertical and horizontal directions. A region is said to border a second region on a given side (left, right, top, and/or bottom) if: 1) they share a

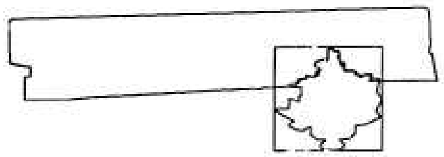
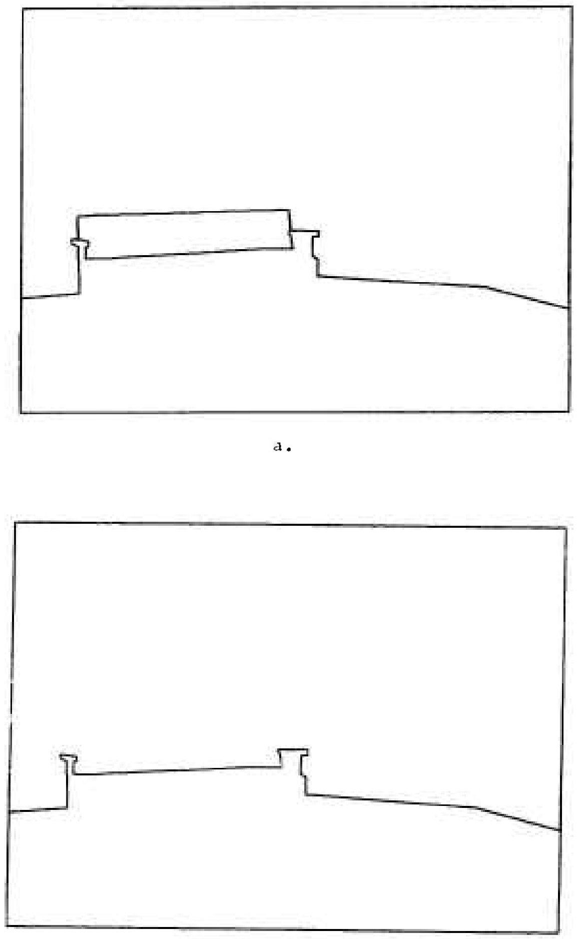


Figure 4.6. Vase and sofa of the room scene with MBR of vase shown.



a.

b.

Figure 4.7. An examole of a tyoe 2 occlusion and its restoration.

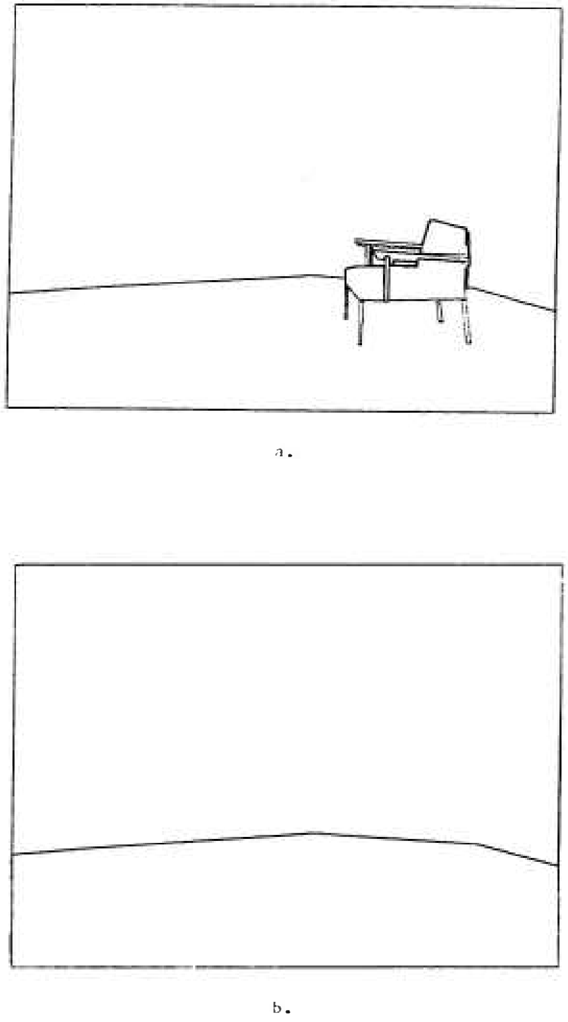


Figure 4.8 Another example of a type 2 occlusion and its restoration.

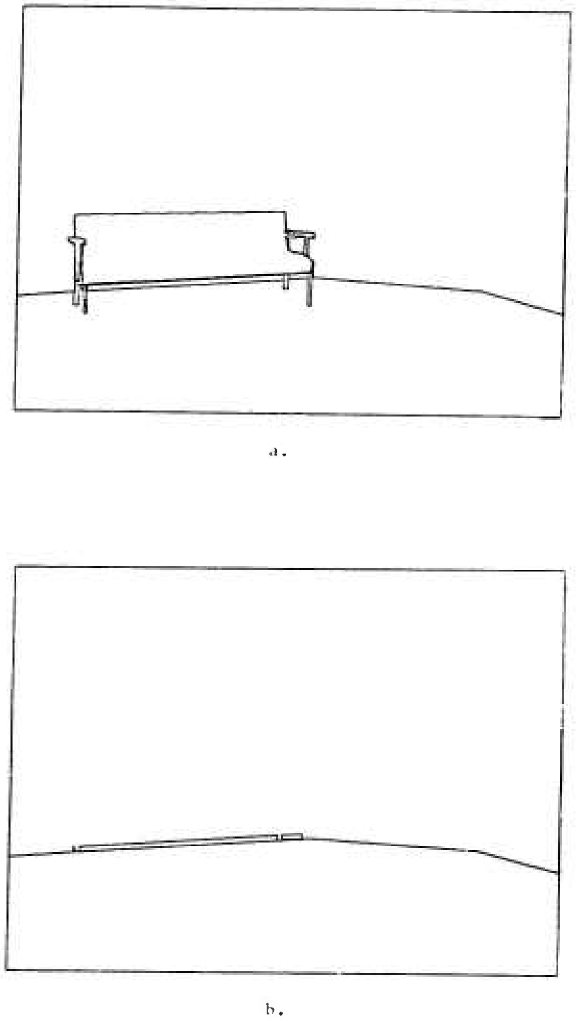


Figure 4.9. Yet another example of a type 2 occlusion and its restoration.

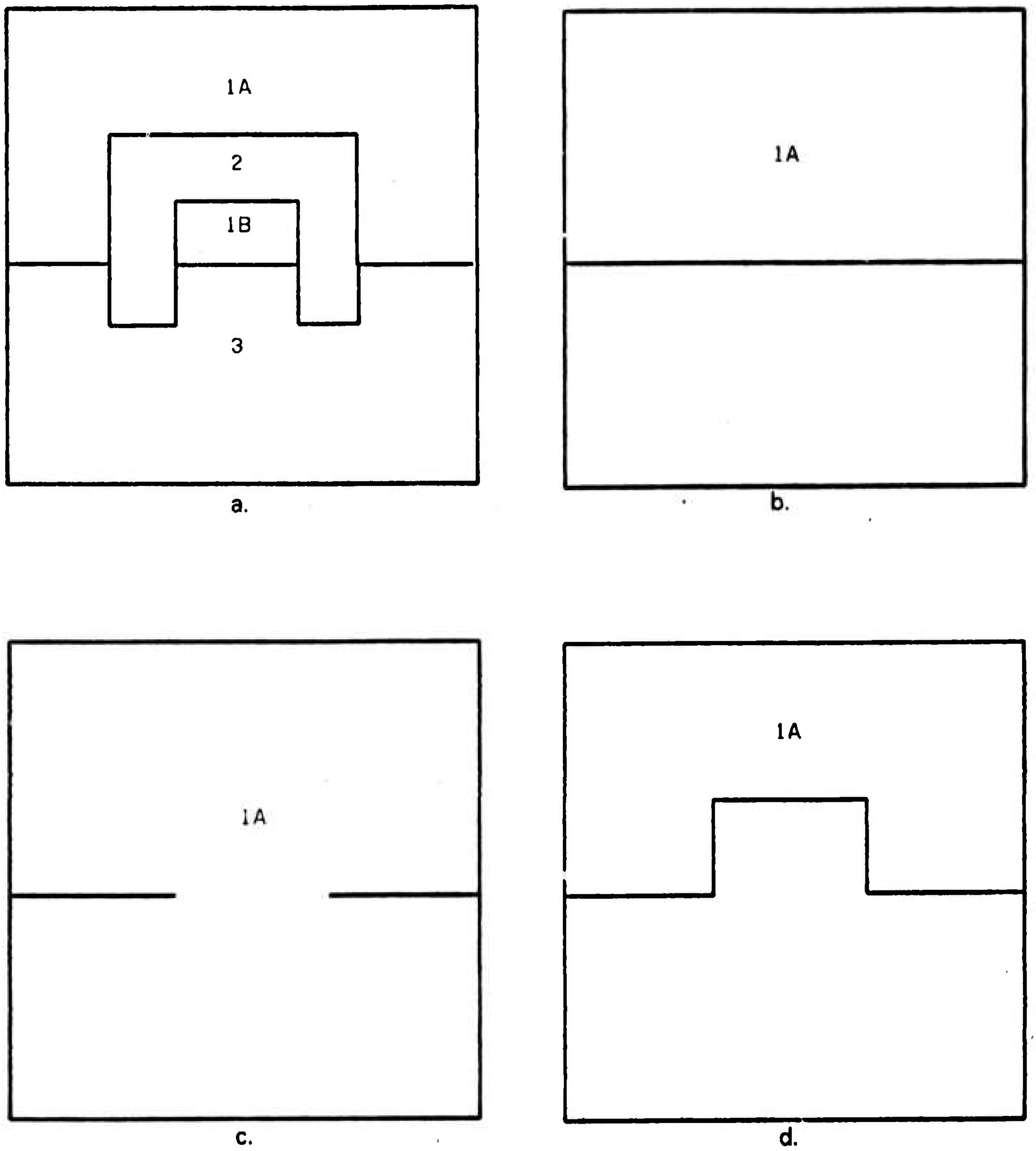


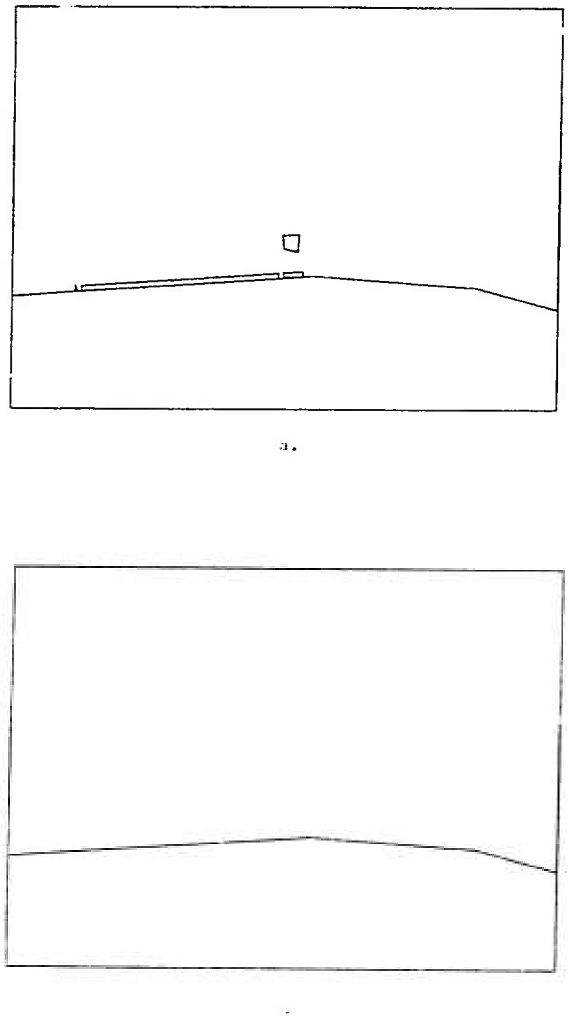
Figure 4.10. Example of a type 2 occlusion restoration.

common boundary, and 2) the first region has any point in common with the given side of the MBR of the second region. Given this definition, it is clear that the vase in figure 4.6 is bordered by the body of the couch on the left, right and top. Note that when we establish the directional relationship between one region and its neighbor, we have usually discovered at the same time the inverse relationship, i.e., the direction of the original region from the neighbor. Thus we know that the vase is below the sofa body in figure 4.6 without explicitly testing the boundaries of the sofa with respect to the MBR of the vase. We can use this fact to advantage when we cannot employ the MBR test in one direction. This situation occurs when one region is contained entirely within the MBR of another in such a way that its boundaries do not intersect the sides of that MBR. An example of this can be seen in figure 4.7.a. In such circumstances the proper relationship of the sofa back with respect to the wall can be discovered by first establishing the directional relationships of the wall with respect to the sofa.

Most of the kinds of ambiguities that were described for the first case are also possible for the current case. The points made then are still appropriate, so no more need be said on the matter at this time. What we will concern ourselves with here are issues regarding the recomputation of boundaries for the occluded region. As noted earlier, boundary recalculation is intimately bound up with the degree of continuity present in the scene for the given type of occlusion. For occlusions of type 1, in which the contained region is the obstructing body, the occluding region's boundaries are eliminated; thus, no matter how irregular its shape, an accurate border determination is derived for the occluded region. In the current case, continuity is less pronounced, so we must be prepared, in some instances, to accept a certain degree of error for boundary determination.

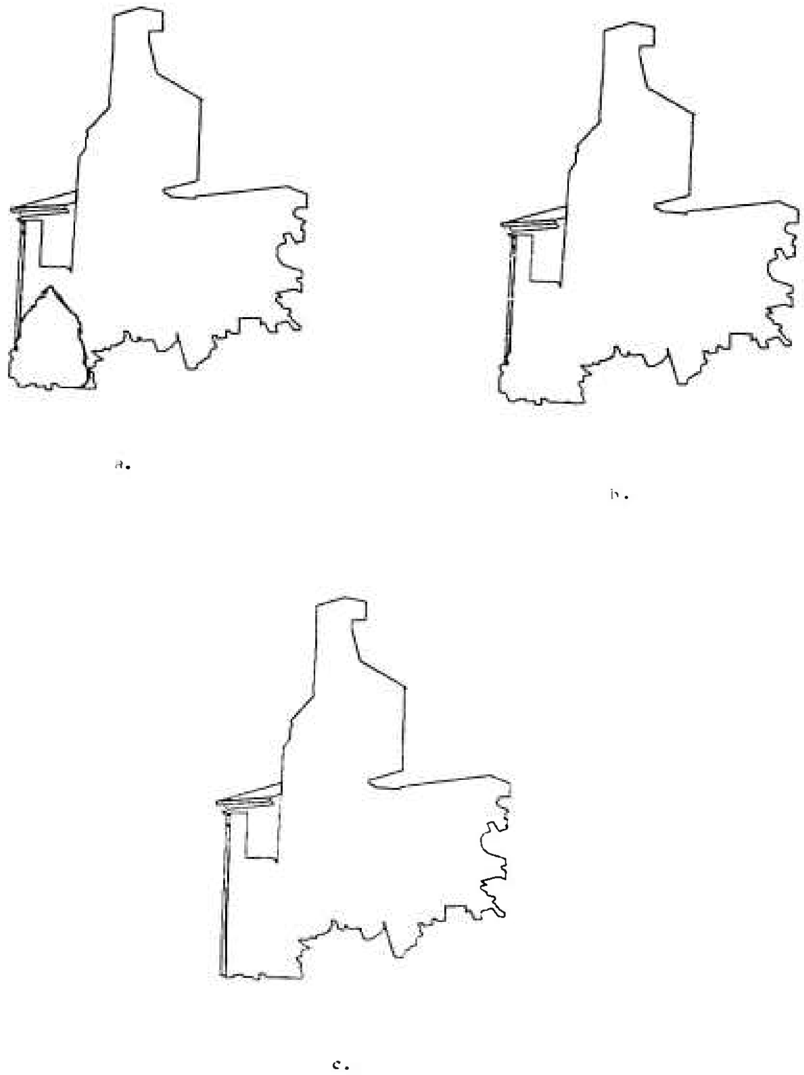
Of particular interest are the three subcases of a type 2 occlusion as they occur in figures 4.7.a, 4.8.a, and 4.9.a with respect to the wall. The first subcase is characterized by the fact that the MBR for one region lies within the MBR of a second. The proper boundary extension is computed by eliminating the portion of the boundary of the occluded region which is in common with the occluding region, and inserting in its place that part of the obstructing region's border which does not lie in common. Using this procedure we obtain the new boundaries illustrated in figure 4.7.b. In figure 4.8.a we observe a variation: the boundary of the chair extends beyond the MBR of the wall. This is an indication that continuity has been interrupted at the first and last common point of the two regions. To restore order, that part of the border of the partially obstructed region that lies in common with the occluding region are deleted. Lines are then extended from the first point (following the border in either direction) that lies in common, to intersect a line extended from the last such point (figure 4.8.b). These lines will have the same slope as some small line segment preceding and including each of the extreme points (for figure 4.8.b the lines coincide). We do not permit that the line segments intersect beyond a fixed distance outside of the MBR of the occluded region. If such is the case, the approach described in the next paragraph must be employed.

The third subcase, which might at first appear to be the same as the preceding one, has the peculiarity that it does not completely occlude the wall on the bottom (figure 4.9.a). Of course, on a local level it is not known that the small region under



a.

Figure 4.11. Steps in the removal of the sofa as an occlusion.



a.

b.

c.

Figure 4.12. Boundary restoration for a wall of the house of figure 2.4.c.

the sofa is a part of the wall. The only hint of a difference is that there is an additional bordering region which lies below the occluding region and overlaps the MBR's of the occluded and occluding regions. The construct in figure 4.10.a makes the issue somewhat clearer: areas 1a and 1b represent a surface occluded by the object designated by region 2. Region 1b overlaps the MBR's of regions 1a and 2. Deducing that fact allows us to make the correct reconstruction. This can be accomplished by making a line extension as before (figure 4.10.b) and then deleting that portion of the extension which lies in common with the overlapping region (figure 4.10.c). The final adjustment is made by adding the uncommon portion of the overlapping region (figure 4.10.d).

It should be clear that the procedure in the preceding paragraph that results in figures 4.9.b and 4.10.d does not produce the desired final result. The next step should be to merge regions on the basis of similarity. In fact the reader, will note that we have implicitly taken such a step before. The regional representations for both the sofa and chair of figures 4.8.a and 4.9.a must contain references to regions which represent the sections of wall which are seen through the arms of the furniture. Previously, we were concerned only with the issue of recomputing regional boundaries in reference to the outer borders of the occluding regions. In reality the external boundary recomputation actually calculates new borders as shown in figure 4.11.a. A joining procedure gives the final desired output shown in figure 4.11.b.

The types of boundary adjustment described above will run into trouble for more natural scenes. Consider the particular type 2 occlusion shown in figure 4.12.a, which shows a bush occluding one of the walls of the house. A more general approach is required in order to compensate for the loss of regularity which is, in a sense, also a loss in continuity. In such circumstances we can delete the common boundary between the two regions and then complete the broken boundary of the obstructed region by inserting the uncommon border segment from the occluding region (just as we did in the first subcase). This results in a region which closely approximates reality (figure 4.12.b), but which overlaps inaccurately on its lower left side. To correct for this an investigation can be made to see if there are additional neighboring regions of the bush which might correspond to an object which is also occluded and which delimits the wall boundary. In this case we find a drainpipe which is obstructed by the bush on the left side, and which is of the same approximate range as the rear of the wall. From this we infer that the derived region for the drainpipe bounds the wall to the left, so that the proper restoration is given in figure 4.12.c.

The final step of the hypothesization of a new representation involves re-estimation of the standard regional properties and relationships. As in the former case, attributes such as average hue, saturation, intensity, and color can be assumed to be the same for the new region as they were for the occluded area. Geometrical properties will require recalculation because of the boundary extensions. Finally, two-dimensional positional relationships between neighboring regions will have to be recomputed due to the elimination (with respect to the new region) of the occluding segment. The occluding region will have a new three-dimensional relationship (as it does in all cases) which places it in front of the newly created region.

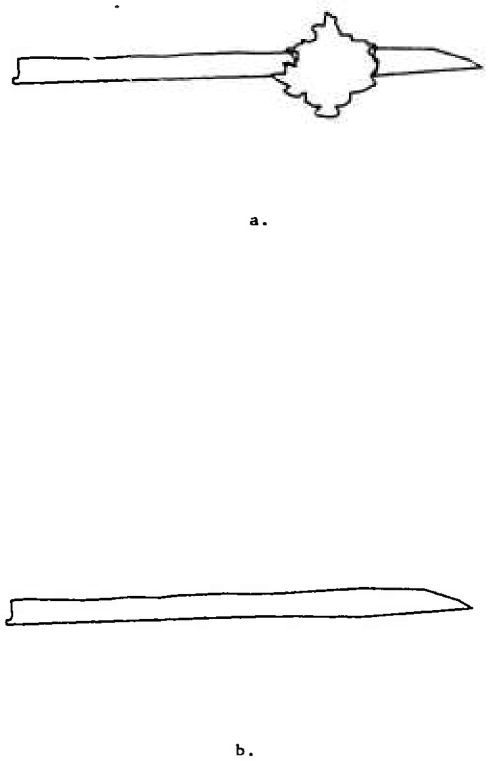


Figure 4.13. Seat of sofa from the room scene.

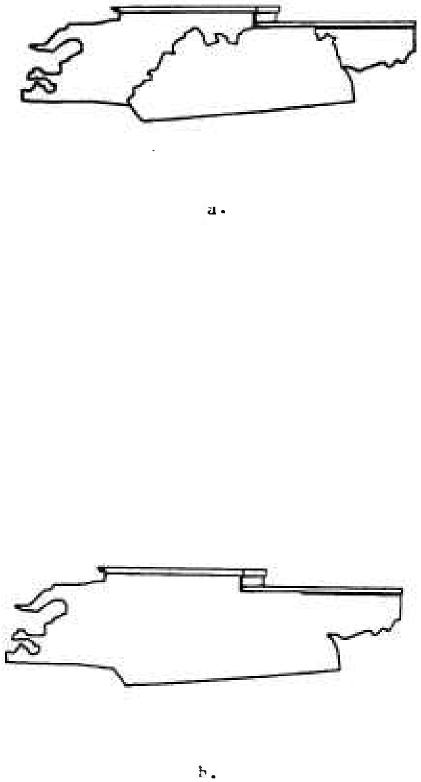


Figure 4.14. Bush and wall areas from house scene.

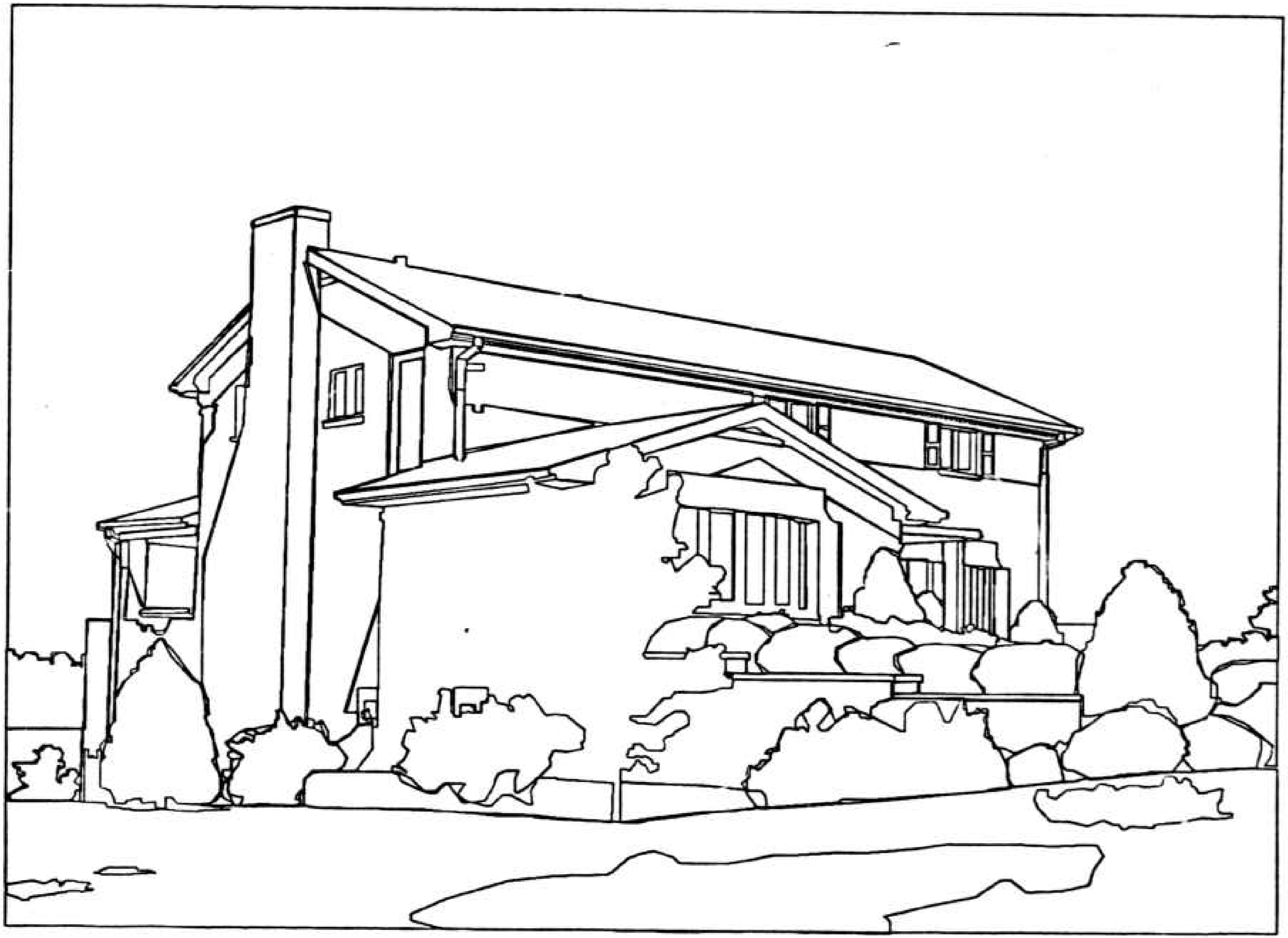


Figure 4.15. Adjusting side and front of house to compensate for occluding tree.

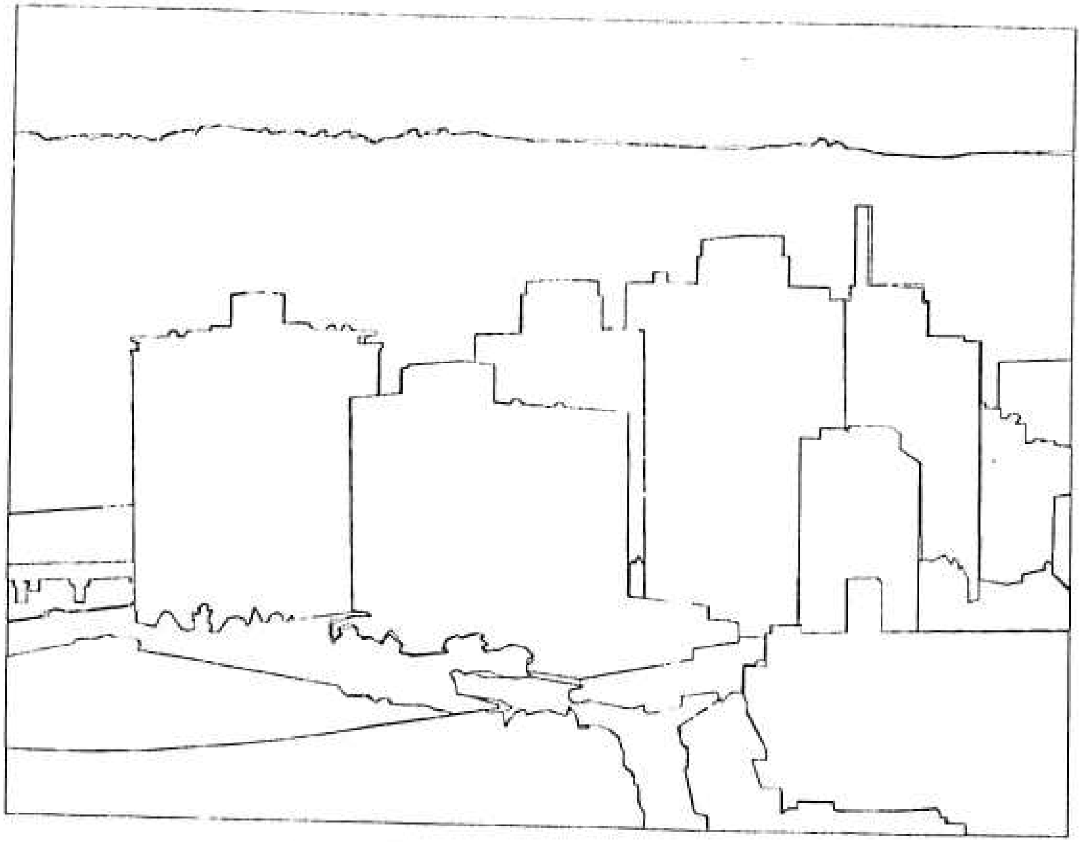


Figure 4.16. Hand segmentation of skyline from figure 2.4.f

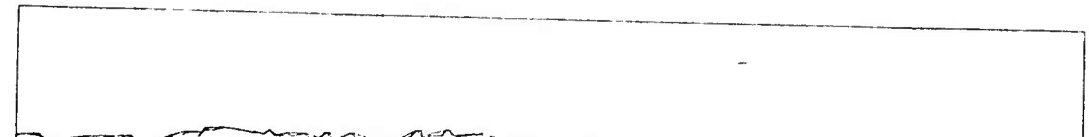


Figure 4.17. Hand segmentation of skyline with some buildings restored.

Case 3: Two or more regions of similar properties bordering a region of dissimilar properties.

There are any number of examples of this type of occlusion ranging in complexity from the relatively simple instance where an object is cut in two or three segments by an occluding member (e.g., the chair in figure 2.4.b), to the very difficult case of an irregular shape showing glimpses of one or more surfaces which it might occlude (e.g., the tree occluding the left front of the house). The simpler occurrences of this case of occlusion are relatively easy to check by looking for another region of similar properties which borders the suspected occluding region. If such a region is found more or less on the opposite side, the hypothesis could be strengthened, as this is a very strong indication of continuity.

As always, the problem of boundary recomputation is much simpler for regular man-made objects than for nature's handicraft. The two parts of the couch seat shown in figure 4.13.a can be joined by locating the first and last points of the common boundaries of the seat parts and the occluding region and extending lines between respective points on opposite sides (figure 4.13.b). In some circumstances we might want to extend intersecting line segments, based on slope estimates, from the common points. For more irregular types of bodies (figure 4.14.a) we would operate much as we did for case 2: we would extend the boundaries between corresponding common points by inserting the uncommon segment of border of the occluding region which lies between (figure 4.14.b). Adjustments might have to be made if analysis of adjacent regions indicate a correction to the boundary.

When more than two similar regions are involved, additional care must be taken as to the order in which the process goes on. For regular objects of well-defined structure, such as the baseboards of the room scene, this means using some care in iterating on the procedure described above for increasingly longer sections. The problem is much more difficult for the type of situation posed by the occlusion of the left front of the house by the tree In such circumstances we might have to proceed by grafting a region corresponding to the approximate shape of the tree onto the regions of the wall to yield a result similar to that given by figure 4.15. One would then have the remaining task of trimming this gross estimate by analyzing adjacent regions to see if they might indicate occlusions and thereby establish more realistic boundaries. The use of such a method implies that we would be able to determine when features of regularity could be exploited.

Completion of the hypothesis entails computation of the rest of the regional properties and relationships which describe any region. The process is esentially the same as for the previous cases and requires no additional explanation here.

Case 4: One region borders another on two sides.

Examples of this type of occlusion are fairly common and can be seen in figure 2.4.f, or perhaps more clearly in figure 4.16 which is an idealized hand segmentation of the skyline. On close examination it can be seen that this case is but the limit of a type 2 occlusion as less and less contact exists between one region and another on one side. As such, it is processed in essentially the same way. Application of the techniques listed above will result in new boundary limitations as shown in figure 4.17 for some of the structures of 4.16.

One of the kinds of difficulties that can arise with respect to this class is the categorization of two regions which have borders in common along an essentially straight diagonal line (e.g., the upper left side of the house with respect to the eaves in figure 2.4.c) as a type 4 occlusion. This comes about because of the choice of the MBR orientation. Due to problems in boundary recomputation that can arise because of a lack of continuity indicators, we would rather consider this as an instance of class 5 (see below). To remedy the situation we can perform an additional test to see if the common boundary of the adjacent regions is essentially a straight line with a slope that indicates the desired orientation.

Case 5: One region borders another on one side.

This type of occlusion is the further limitation of a case 2 instance as there becomes less and less contact between the enclosed region and the enclosing region on two sides. A borderline case is shown in figure 2.4.b by the body of the couch as it barely extends on either "side" of the table. Occlusions of type 5 have particular problems which prevent them from being handled in the same way as the other classes.

In the first place, existence of this condition is less likely to be indicative of an actual occlusion. The principal difficulty is that there is no continuity (local context) to tell if one object continues behind the other or whether they butt at that juncture. For example, if we consider the room scene, there is really no local evidence to indicate whether the baseboard is hiding a portion of the rug, a portion of the wall, or neither. The same is true for the hedges in figure 2.4.c with respect to the grass and the side of the house. What we can do is determine whether an occlusion is possible without assigning any great degree of confidence to the decision. Range and height disparities would be helpful in this respect if the neighboring regions have similar surface orientations. If orientations are dissimilar, then the range limits of the occluded region must be well within or greater than the limits of the other region.

Even if one could decisively detect an occlusion of this class with the available local clues, a second difficulty arises with respect to the determination of new boundaries. Clues may be available to indicate bounding dimensions along which to extend the region, but how far should the extension go? It is safe to extend the side of the house as far down as the hedge limits in figure 2.4.c, but if one used this same criterion for the partially hidden leg of the chair in the room scene the result would be grossly inaccurate. In some circumstances it might be possible to use principles of symmetry to restore the hidden surface.

The difficulties that have been brought out with respect to the handling of occlusions of the current class prompts us to consider conditional hypothesization of new regions. Instead of complete restoration, a notation in the data base of a possible occlusion for the regions involved could be indicated. If later verifications were made, the world model and other higher level knowledge sources might suggest the proper course of action for restoration.

Case 6: An object is completely hidden from view.

This is the class of occlusions briefly mentioned earlier. No direct sensory evidence exists to indicate the presence of a specific object, so hypotheses concerning its existence cannot be generated by the knowledge source on a low level. The functions of the knowledge source must be restricted to verification of hypotheses proposed by other knowledge modules (e.g., an object identification module). The identification of an object which has a piece missing can be made much stronger if we can at least verify that it might be obscured by some other object in a scene. If an approximate location for the hidden piece can be established, only two things need be determined. We first want to know that another region occupies the same space. Once that is established we must find out if this latter region occludes the object of which the hidden piece is a part.

Knowledge About Shadows and Highlights

What is known about shadows and highlights? Resorting to Webster's again, "shadow" is defined as "a definite area of shade cast upon a surface by a body intercepting the light rays", and "highlights" as "a part on which light is brightest". In turn, "shade" is defined as "comparative darkness caused by a more or less opaque object cutting off rays of light", or "an area less brightly lighted than its surroundings". For our purposes the key notions contained in these definitions are: 1) darkness (brightness) and 2) in relation to surrounding areas. Point 2 indicates that we must establish some norm for comparison. We could consider all regions of a scene as they relate to the most brightly illuminated area of the scene (e.g., the brightly lit portion of the rug in the lower right corner of figure 2.4.b). This could be convenient if we wish to determine simple relative overall lighting effects upon a scene (e.g., whether the couch is more in shadow than the chair). This is not, however, the way humans consider a given scene. They tend to refer to the shading of different areas in terms relative to some degree of lighting which seems average for the scene in question. They would say, for instance, that the bottom right corner of the image of figure 2.4.b was highlighted or brighter than its surroundings, rather than specify that everything else was darker than that bit of rug. This latter approach also seems to be a most reasonable one for machine analysis, and it is the one that we will pursue.

Proceeding on this basis, there are at least two levels of attack for solving the problem. A low level approach has the function of restoring those regions which have had a portion of their surface partitioned out as a distinct entity because of shadow or highlight effects. We want to establish that the illuminated side wall of the house and the shadowed portion above it are in reality the same object. From this standpoint, we would like to see if we can effect a case analysis for shadows in much the same way as we did for occlusions. That is, we want to classify given segments of a scene as possible shadows and eliminate, to some extent, those consequences which might hinder the identification process. In this respect, the only likely dimension along which

COLOR	HUE RANGE	INTENSITY RANGE
RED	0-30	85-120
DARK RED	296-360	30-85
LIGHT GREEN	65-120	95-140
DARK GREEN	155-205	60-95
BLUE	195-210	175-205

Table 4.1. Table of intensity ranges for corresponding colors of the house scene.

it seems pertinent to explore is that of degree of lighting. As noted above this is a comparative measure. As such, this degree of shading can only be decided in relation to some supplied standard or to other regions in the scene under consideration. One of the possible ways of establishing the standard is by determining ranges of intensity for the major divisions of hue in a scene. Table 4.1 shows the ranges of distinguishable hue and the corresponding intensity spreads for the same scene. The mean of the intensity distribution for a given hue would then specify the standard for that color. Extracted features of a region can then be compared to this standard to determine if it is brighter or darker than the average. A further dichotomization can be made on the basis of the degree to which a suspected region varies from the norm. In this way we can distinguish four classes of shading:

1) regions brighter than the average, but which are similar in other respects to some part of their surroundings;

2) regions very much brighter than the average and which have some similarity in hue to part of their surroundings, but which differ in most other respects;

3) regions darker than the average, but which are similar in other respects to some part of their surroundings;

4) regions very much darker than the average and which have some similarity in hue to part of their surroundings, but which differ in most other respects.

This classification is strongly ordered along pragmatic lines of the system's ability to distinguish differences in shading. It may also serve to categorize shadows and highlights in terms of the effects they have upon scene analysis.

Let us pause for a moment to consider the nature of some of these effects. The greatest potential for variation seems to arise in indoor scenes because of multiple lighting sources. In the room, for example, illumination comes from windows, from overhead lights, and from the camera strobe. The diverse sources of light have resulted in a number of shadows of varying degree, some of which are not very obvious. The most subtle effect is the very gradual change in shading of the wall as it nears various objects in the room (e.g., just to the left of the sofa). As we have seen in chapter 3 this condition can cause trouble for the segmentation process. Outdoor scenes can also present their difficulties. On bright sunny days, for instance, shadow effects can be very strong, so strong as to block out edges and texture. The effect is strengthened when the shadowed object has a basically achromatic color. This can be seen in the case of the shadow under the rock to the right of the bear in figure 2.4.e. Such occurrences are not very helpful when we are trying to determine actual boundaries for proper identification.

The second level of approach to the shadow problem is through the application of goal-directed techniques to analyze portions of the scene in terms of models. An effort could also be made to isolate areas of the picture which are shaded but which have not been identified as such by the low-level segmentation. These tasks might be undertaken simply to gain a greater understanding of a scene, or more practically, to analyze areas which are under scrutiny by other knowledge sources (e.g., the object identification module trying to resolve differing attributes). In the quest for such information it would be necessary to make use of such information as location and brilliance of lighting sources, and location of intruding objects which might cast a shadow. Low amplitude differentiation of texture, intensity, hue, and saturation could then be used to examine regions for possible minor variations which might be indicative of shadow effects. Any approximate areas hypothesized in this way would then be evaluated further in terms of the lighting model and placement of already recognized objects. If a verification can be made, the extracted region would be accepted as a shadow. Conversely, one could use the information about detected shadows to hypothesize or verify light sources and three-dimensional placement.

Our primary purpose, of course, is to "explain" shadowed or highlighted areas that differ sufficiently from the norm as to be segmented apart from other regions which depict the same surface. The discussion which follows is concerned mainly with detection and use of local clues to postulate the existence of a shadowed region and hypothesize a new region which negates its effect. This is accomplished through the bottom-up approach described earlier which attempts to classify shadows on the basis of extracted features and relationships. Obviously, the process is highly dependent upon low-level segmentation. In situations where there are areas of low constrast the desired partition might not be forthcoming. For instance, the shadow of the bear's paw refused to be separated from the bear's body on initial segmentation (figure 2.4.e). In such cases, if identification relies heavily on acquiring the shaded portion, higher level routines will have to point out areas which might have obscured a part of the object in question. This would in turn be verified by the same mechanisms described below.

Although we have put forth a classification of four categories based on degree of shading, we will describe only two cases below; the other two are symmetric, substituting "brighter" for "darker", and "highlighted" for "shaded" or "shadowed". When there is a relevant difference, it will be pointed out.

Case 1: Regions which are darker (brighter) than average but which are similar in most other respects to some portions of their surroundings.

Color plays the dominant role in making a first estimate of shadow classification. Let us examine some of the aspects of this property that are affected by shadows. Color can be described by values of hue, saturation, and intensity. We use psychological terms here because they are more likely to be meaningful to the reader. The actual physical analogs are radiant energy, wave length, and degree of white light as determined from transformations from the red, green, and blue sensory inputs. Achromatic colors (shades of gray which range from white to black) can be characterized by low saturation. As saturation increases, hue becomes the primary determiner of what most of us think of as color. Although the achromatic colors can theoretically be entirely free of the influence of hue, we have found in practice that this is not the case. In fact, even though colors may appear white or gray or black they are likely to be, for example, pinkish white or greenish gray or bluish black.

It would have been fortunate if the effect of shadows was similar to the physical

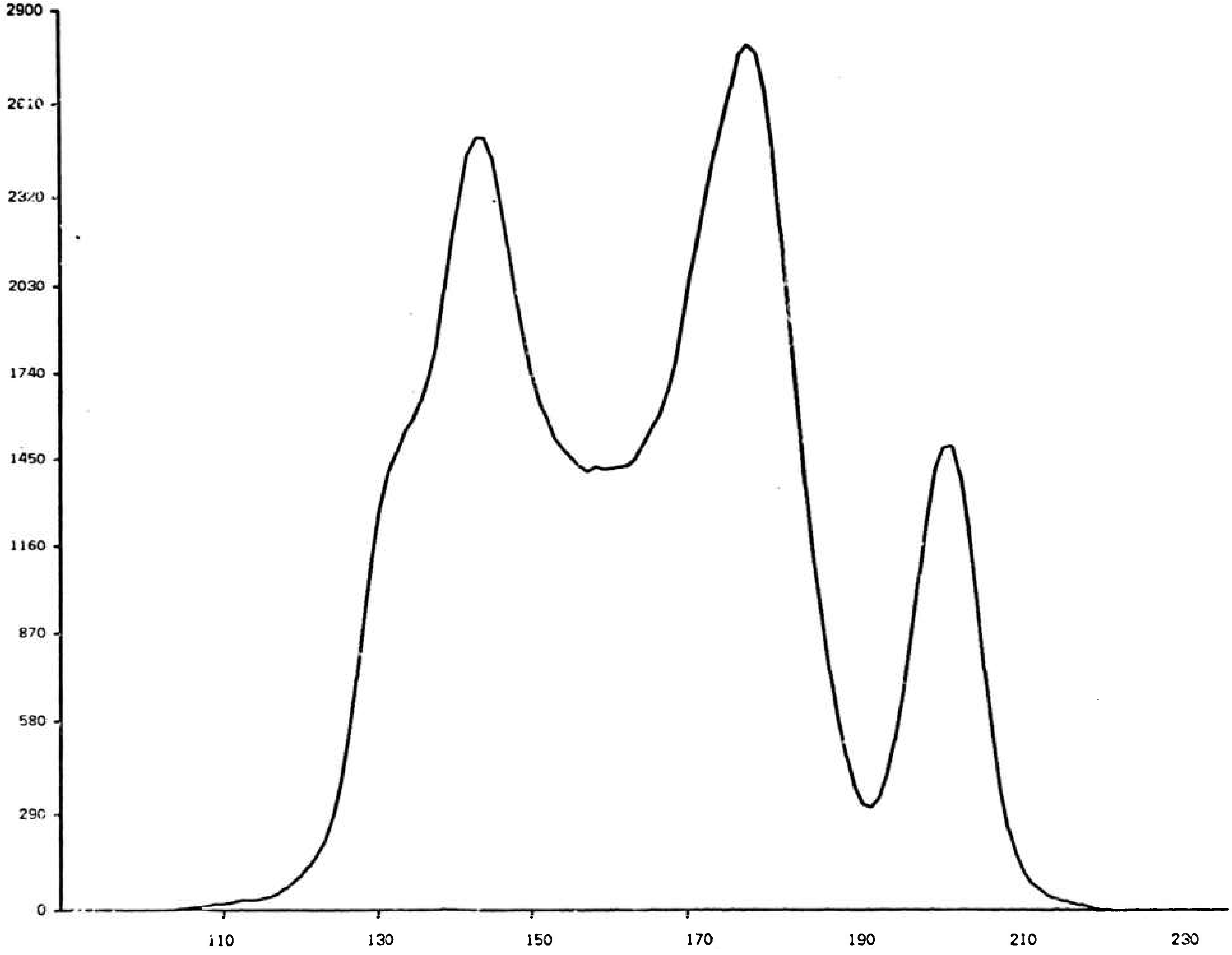


Figure 4.18. Histogram of intensity values for the rug in the scene of figure 2.4.b.

result of removal of white light. Hue and intensity would then remain unchanged while saturation would decrease. This is not the case. Nor is it a matter of a simple change in intensity. Rather, all three components are affected in various ways. Factors such as lighting, colors of surrounding objects, and reflectivity influence what components are changed and to what degree. For example, in the outdoor scenes saturation of a shadowed surface is higher in value than for the unaffected area. If the area is of a neutral color the increase seems to be much more marked. On the other hand, for the indoor scene the opposite effect occurs for the shaded portion of the rug under the sofa. This is probably due in part to the fact that the shadowed portion does not reflect direct light, but only that light which is in turn reflected from other surfaces In the scene. It must be that the shaded region takes on some of the reflected properties in very complex ways. It is also the case that hue is affected in a somewhat unpredictable way. The direction of change is not determinable but the degree is usually within some estimable range. As before, we believe the change is influenced by reflected light. The shadowed areas on the walls of the house, for instance, vary in hue toward the green side. The last of our components, intensity, is always predictable in direction (decreases for shadows, increases for highlights) and varies in amount in accordance with the degree of the lighting change.

On the basis of these observations we can make some tentative evaluation of the lighting effect upon a given region. We can perform this evaluation in terms of a comparison of properties with immediate surrounding regions. Unless an entire surface is shadowed there must exist another neighboring region in proper relation to the proposed shaded area and to the light source which represents an unshaded portion of the actual surface of which the candidate region is a part. We must first determine if such a region exists. This can be done by a comparison of component properties of color. An examination is made of the hue attribute of neighboring regions to see if it is within the same range band in accordance with a property table such as was produced in table 4.1. For example, if we were considering the brightest portion of the rug in figure 2.4.b we would find that it has an average hue of 53 while the largest portion of the rug to its left has an average of 57. Clearly the two areas are similar In this parameter. It now remains to establish which is typical of the norm. If we look at the intensity values for the appropriate range of hue (figure 4.18) we see a distinct peak between 190 and 220 and a large indeterminate area between 80 and 190. A reasonable assumption is that the pixels determined by the peak in the high range constitute a brighter than average portion of the picture. As was determined by segmentation in the previous chapter this is indeed the highlighted part of the rug. With this example in mind we arbitrarily establish the middle portion of the intensity scale (60 to 180) as that range in which areas of average intensity are likely to be found. If we find two juxtaposed regions of similar hue which both have average intensity values within this range, we take the standard as the one with value closest to the midpoint (120). The remaining region is labeled a shadow or highlight, depending on whether its average intensity is lower or higher than the value of the standard.

At this stage of the process we should have discerned whether the region under investigation is darker than the norm, hence a candidate for a shadowed area, and whether it is of case 1 or 2. If we have established a possibility of shadow, we can make a further determination of its suitability by an additional examination of its neighbors. We want to check to see that there is a shadow causing region between the shaded area and the light source (this heuristic is not applicable for highlights). These adjacent regions are evaluated as to direction and degree of contact in exactly the same way as for an occlusion. Regions which are completely surrounded by another region (no picture borders allowed) are unlikely candidates. Shadows of this type can only be produced for outdoor scenes by clouds or flying craft and are usually fairly large. For indoor scenes we will find this type of shadow effected by some sort of suspended object in a strong light (some object suspended by thin wire perhaps). Since these circumstances do not occur in our selection of subjects, the existence of such a condition is enough to disqualify the candidate. What we are saying in effect, is that any shadow area must be in immediate proximity to the object producing the shadow. If the region under investigation is adjacent to a border of the picture, however, the object might be cut out. This means that we must establish the existence of any adjacent region which exhibits the proper directional relationship with respect to a specified light source. For instance, since the sun is just about overhead, the shadow on the upper left side of the house in figure 2.4.c could be produced by the region representing the eaves.

Even though a region might survive the tests proposed up to this point, there is no certainty we have captured a shadow area. In fact, the large darker green area of the grass in the foreground of figure 2.4.c does pass all the tests but is in reality just a darker patch of grass. The erroneous hypothesis would be given a somewhat lower level of confidence because it could only be justified on the basis of a shadow producing object which might have been cut out of the picture. In spite of this, we would be willing to accept such a hypothesis because it could be the proper decision in some cases. Later verification by an object matching routine should fail to verify the hypothesis and correct the error by establishing the orginal segment as the required patch of missing grass.

Notice that the requirement that a shadowed region co-exist with its unshaded counterpart eliminates from consideration, at this level, those areas totally in shadow (e.g., the underside of the front eaves of the house). It also eliminates regions which have similar properties to shaded areas. An example of this phenomenon would be the roof of the house which is very like the upper shadows on the wall. Totally shaded areas could be treated by higher level knowledge sources if it were necessary to explain properties which differed significantly from models.

. In those cases where we have discovered an adjacent region to be a portion of the same surface of which the shaded area is a part, we must complete the hypothesis by merging the two regions to form a new one. This is accomplished by a recomputation of boundaries along with the necessary re-estimation of other properties. The common boundary of the unshadowed area is deleted and the uncommon portion of the shadowed segment is inserted. In this case we know exactly what the proper boundaries are. We do not have to make crude guesses as we did for many of the instances of occlusion. The remaining properties are adjusted just as they were for occlusion. Geometrical attributes are recomputed while most other features are assumed to be the same as they were for the unshaded region. New two-dimensional relationships must be established.

Case 2: Regions which are very much darker (brighter) than the average and which have some similarity in hue to part of their surroundings, but which differ in most other respects.

The predominant characteristic for shadows of this type is that they are so heavy that properties are dissimilar in most respects from the those of the unaffected surface. Saturation may vary considerably, as might intensity. The difference is that intensity will change in a predictable direction. Textural detail is greatly reduced or completely lost. A good example of this is provided by the heavy shadow under the rock to the right of the bear. Saturation for the unshaded rock is .148 and intensity is 114.6, while the values for the shadowed area are .510 and 37.3, respectively. Even the attribute of hue varies to a more marked degree; it can still be very useful, however, in forming a decision as to the presence of shadowed regions. In the instances of case 2 shadows that we have observed, hue has not altered by more than 60 units (17%) from a normally lighted surface of the same type. When the scene is sufficiently rich in variety of color and possesses shadowed areas of reasonable size, we can observe a significant peak in a histogram of the hue parameter. We detect such a peak in figure 4.18 lying between 290 and 360. In the segmentation of the house we found pixels in this range to correspond to the roofs and shaded areas of the brick. Further observation of the histogram shows a following peak in the range 0 to 60. Points under this curve are also red and correspond to the normally lighted portions of the brick. Such observations lead us to pursue an investigation for a case 2 occurrence of shadows when adjacent regions exist which are not classed as case 1 and are within 60 units of hue. The hypothesis is given further credibility when intensities differ by more than 50%, as this indicates a significant change in lighting which is likely to have been the cause of the change in hue. An even higher level of confidence is awarded if we observe the double peak histogram phenomenon. As a final step we check, as before, to see if a proximate region exists which could cast the proposed shadow, and which is consistent with world knowledge concerning light sources.

In the current case, additional problems can arise for recomputation of boundaries because of the increased possibility that different surfaces under the same heavy shadow might be segmented out as a single region. It is also likely, under these circumstances, that any differential which might indicate a low frequency edge will be non-existent. We saw an instance of this before in the example of the dark shadow under the rock which is to the right of the bear's shoulder. We are quite prepared, in this case, to accept the slight error in overlap and make our best effort at joining the shaded area to the rock above. We note in mitigation that although humans are able to perceive that the actual juncture between the rocks is lost in shadow, they too are unable to place the real boundary with precise accuracy. We must also remark that for a different type of scene, such heavy shadows could result in considerably greater errors.

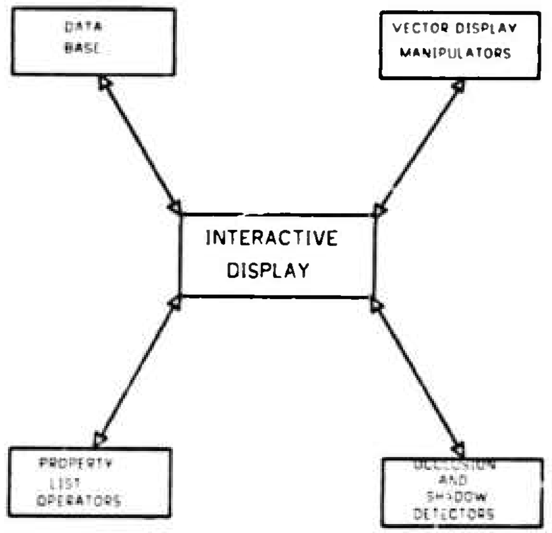


Figure 4.19. Interactive subsystem for processing occlusions, shadows and highlights.

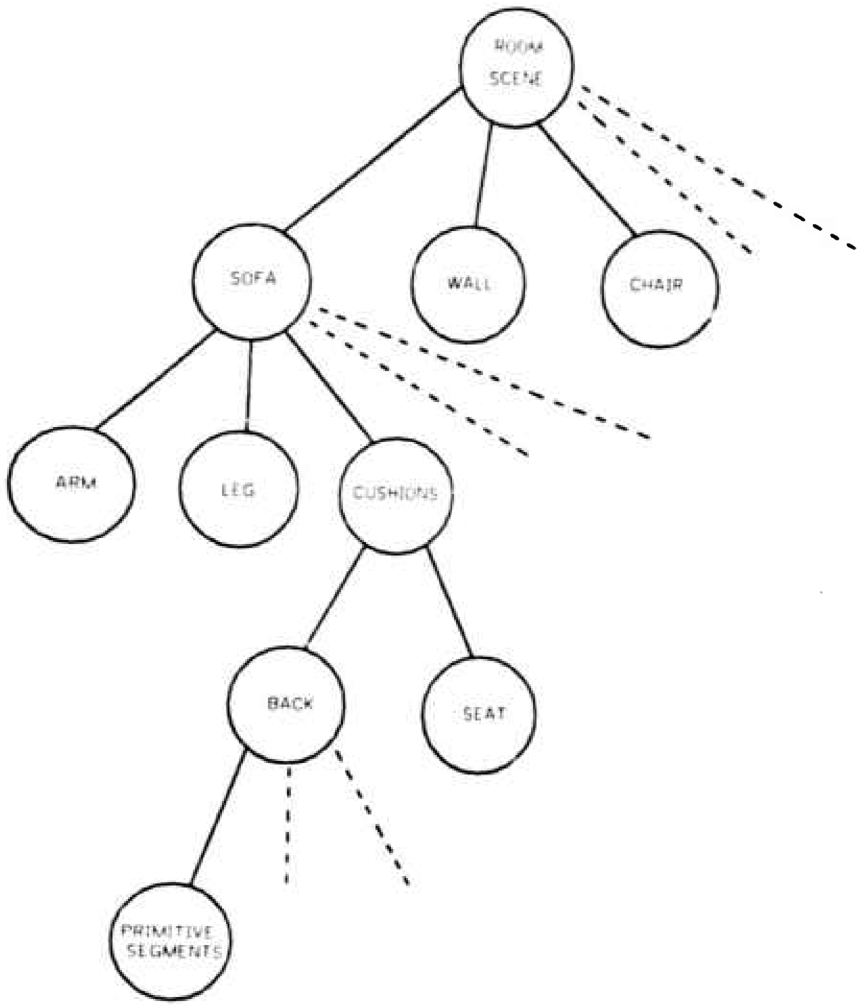


Figure 4.20. Structural organization of the data base upon completion of analysis.

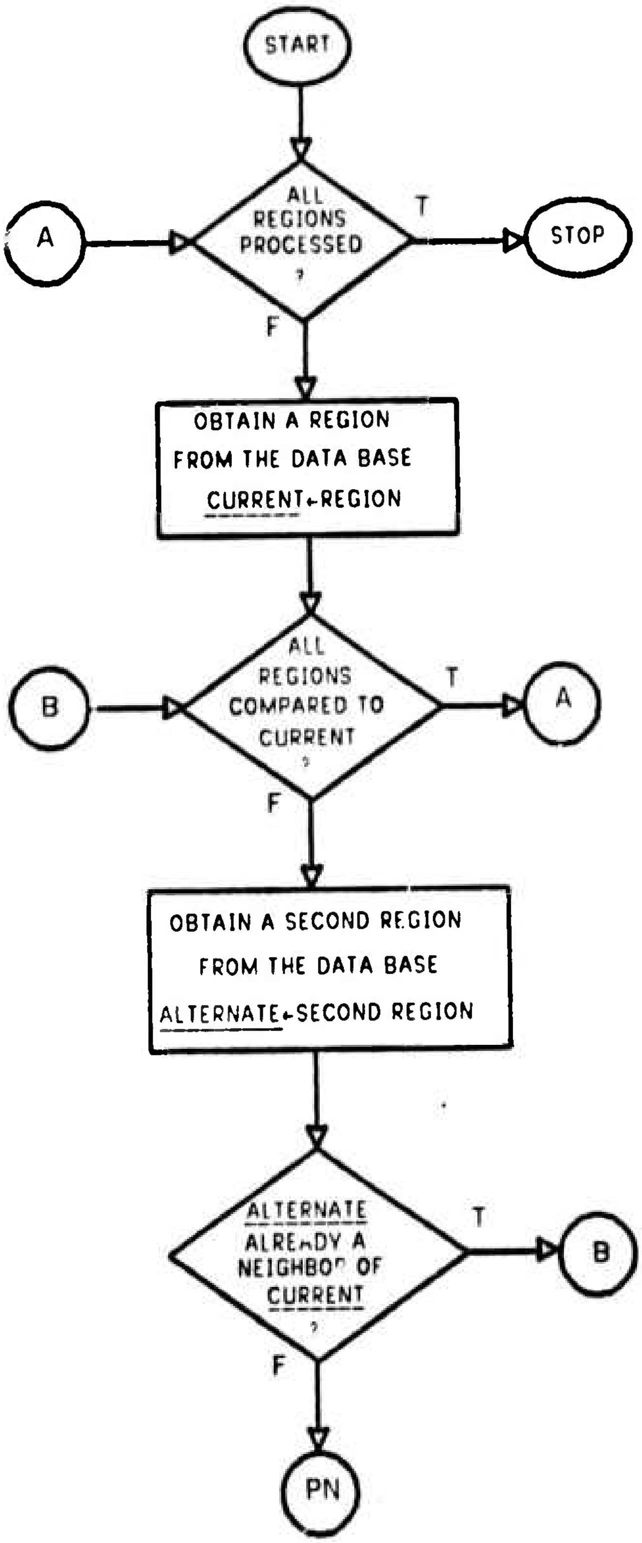

START
ALL REGIONS PROCESSED ?
A
T
STOP
F
OBTAIN A REGION FROM THE DATA BASE CURRENT←REGION
ALL REGIONS COMPARED TO CURRENT ?
B
T
A
F
OBTAIN A SECOND REGION FROM THE DATA BASE ALTERNATE←SECOND REGION
ALTERNATE ALREADY A NEIGHBOR OF CURRENT ?
T
B
F
PN

Figure 4.21. Algorithm for derivation of directional neighbors.

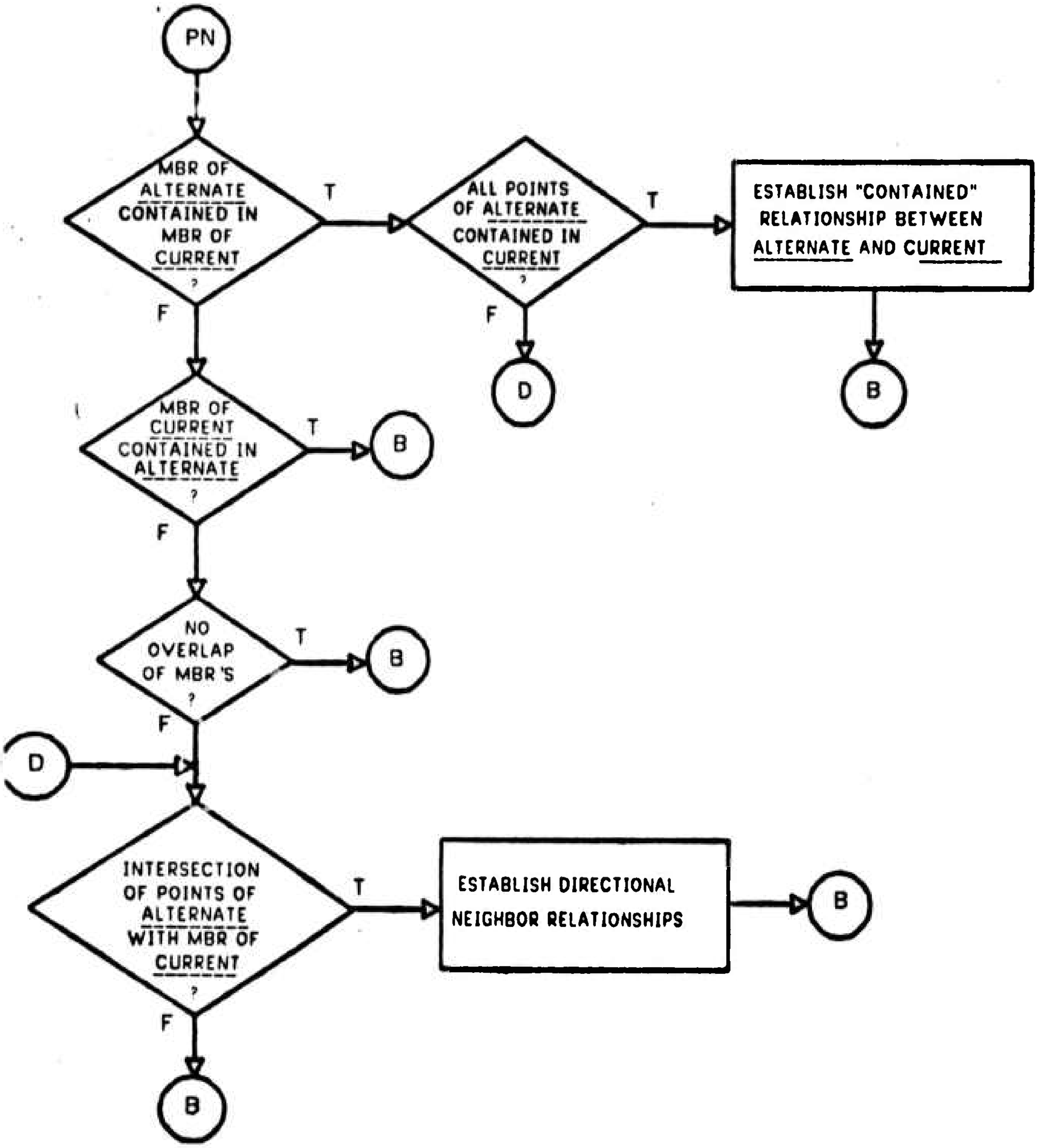

PN
MBR OF ALTERNATE CONTAINED IN MBR OF CURRENT ?
T
F
ALL POINTS OF ALTERNATE CONTAINED IN CURRENT ?
T
F
ESTABLISH "CONTAINED" RELATIONSHIP BETWEEN ALTERNATE AND CURRENT
D
B
MBR OF CURRENT CONTAINED IN ALTERNATE ?
T
B
F
NO OVERLAP OF MBR'S ?
T
B
F
D
INTERSECTION OF POINTS OF ALTERNATE WITH MBR OF CURRENT ?
T
ESTABLISH DIRECTIONAL NEIGHBOR RELATIONSHIPS
B
F
B

Figure 4.21 (continued). Algorithm for derivation of directional neighbors.

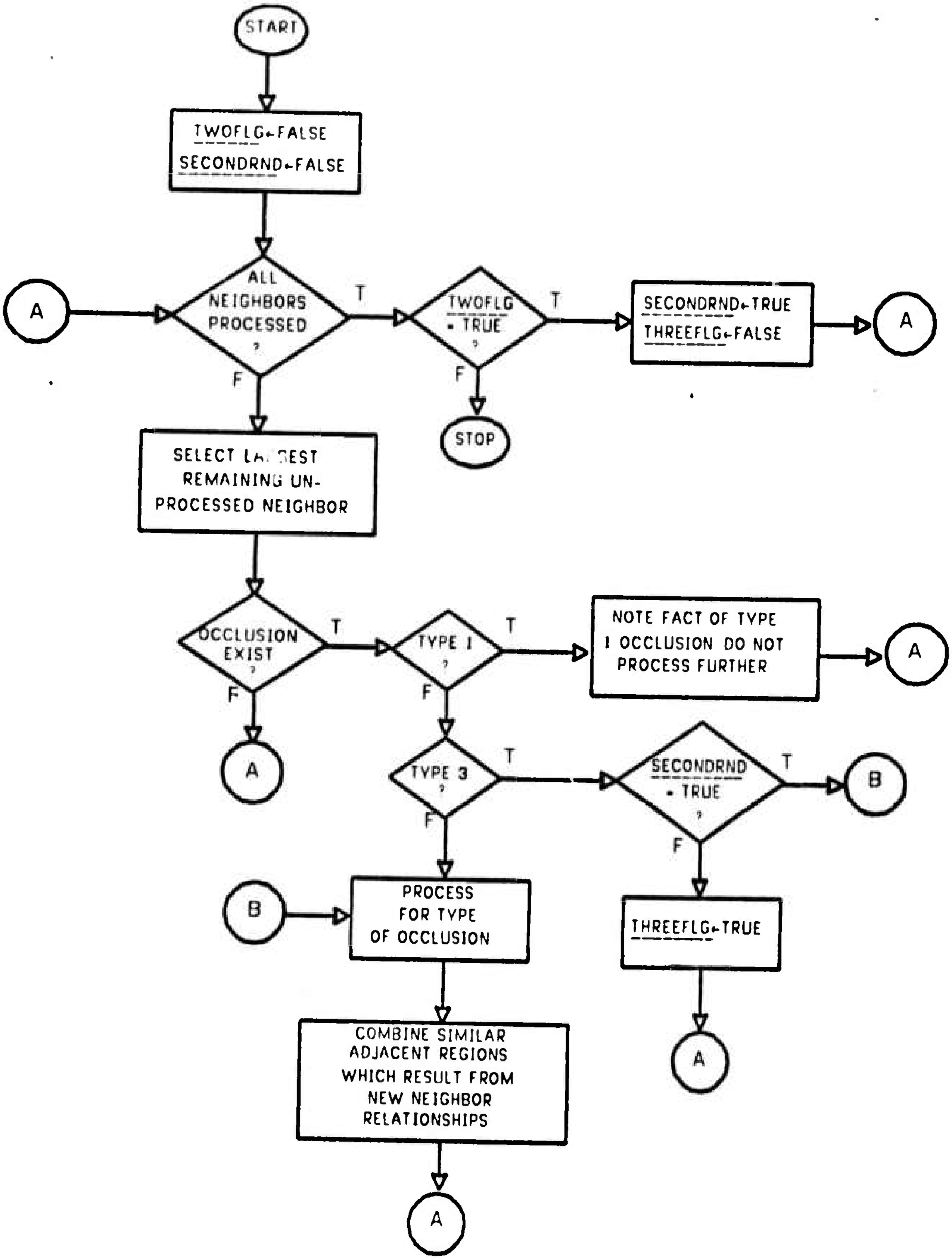

START
TWOFLG←FALSE
SECONDRND←FALSE
A
ALL NEIGHBORS PROCESSED ?
T
TWOFLG = TRUE ?
T
SECONDRND←TRUE
THREEFLG←FALSE
A
F
F
STOP
SELECT LARGEST REMAINING UN-PROCESSED NEIGHBOR
OCCLUSION EXIST ?
T
TYPE 1 ?
T
NOTE FACT OF TYPE 1 OCCLUSION DO NOT PROCESS FURTHER
A
F
F
A
TYPE 3 ?
T
SECONDRND = TRUE ?
T
B
F
F
B
PROCESS FOR TYPE OF OCCLUSION
THREEFLG←TRUE
A
COMBINE SIMILAR ADJACENT REGIONS WHICH RESULT FROM NEW NEIGHBOR RELATIONSHIPS
A

Figure 4.22. Algorithm for detection of occlusions.

permitted adjustment of errors made by automatic computation. We then began to construct subroutines which performed more and more sophisticated border reformulation. The first attempt involved the construction of a procedure which would eliminate the boundary that a specified region had in common with a second region and replace it with the uncommon portion of the boundary of the second region. The line extension subroutines were also constructed.

With the completion of the vector manipulation package we were able to correct occluded boundaries for such simple cases as shown in figures 4.7.a and 4.8.b, if we specified the two regions involved. The next obvious step was to implement automatic detection of the different types of occlusion. This required the calculation of adjoining regions (neighbors) and "contained" regions for all segments of the scene. The procedure is shown in figure 4.21 and is based on the minimum bounding rectangle (MBR) estimation discussed in the preceding section on occlusions. Once the neighbor calculations are made we can specify a region and initiate a computation of all occlusions for that region (figure 4.22). Notice that for type 3 occlusions we must search for an additional neighbor which adjoins the neighbor under consideration. We require that it lie in a direction opposite to that of the region being analyzed for occlusions and that it have similar properties. This is a restriction of the general case of type 3 occlusions but is the only type we are prepared to handle at this time. Notice also that we postpone the processing of a type 3 occlusion until all others have been considered. This is to prevent the section of wall seen through the arm of the couch from being handled as an occlusion of this type. For a number of reasons it is best to remove the upholstered section of the sofa first and then join the section thus uncovered.

Automatic recomputation for two-dimensional and three-dimensional relationships was provided and the control structure described above tried out. The procedure worked well for single occlusions, but encountered a number of difficulties when multiple occlusions were undertaken. For example, if we were to consider the baseboard to the left of the chair in figure 2.4.b we notice that there are three intervening regions between it and the next section of baseboard. How many sections must we allow when we check for a type 3 occlusion? Consider also the sequence of steps shown in figure 4.23 which demonstrates the algorithm for the elimination of the upholstered portion of the sofa that is occluding the wall. If we examine the final result closely we can see that the baseboard which was under the sofa and the rug which was under the table have not been restored. This requires that a two-dimensional relationship be established between the old baseboard and the new wall construct. The old baseboard must also maintain its two-dimensional relationship with the table top. As new constructs emerge, a complex network of relationships between regions in various stages of reconstruction builds up and the problem of determining the proper relationships for occlusion processing becomes increasingly difficult. To avoid the issues raised here we decided to implement the control structure shown, in abbreviated form, in figure 4.24. This is a recursive algorithm which will ensure that, before we remove any occluding region, we check to see if it is in turn occluded. We continue checking occluding regions for possible occlusion until the foremost object in the scene which is in line with the original surface occlusion is obtained. We then proceed to remove these occlusions in the reverse order of their discovery. As the recursion unwinds we ensure that all surfaces occluded by the object currently being removed are restored.

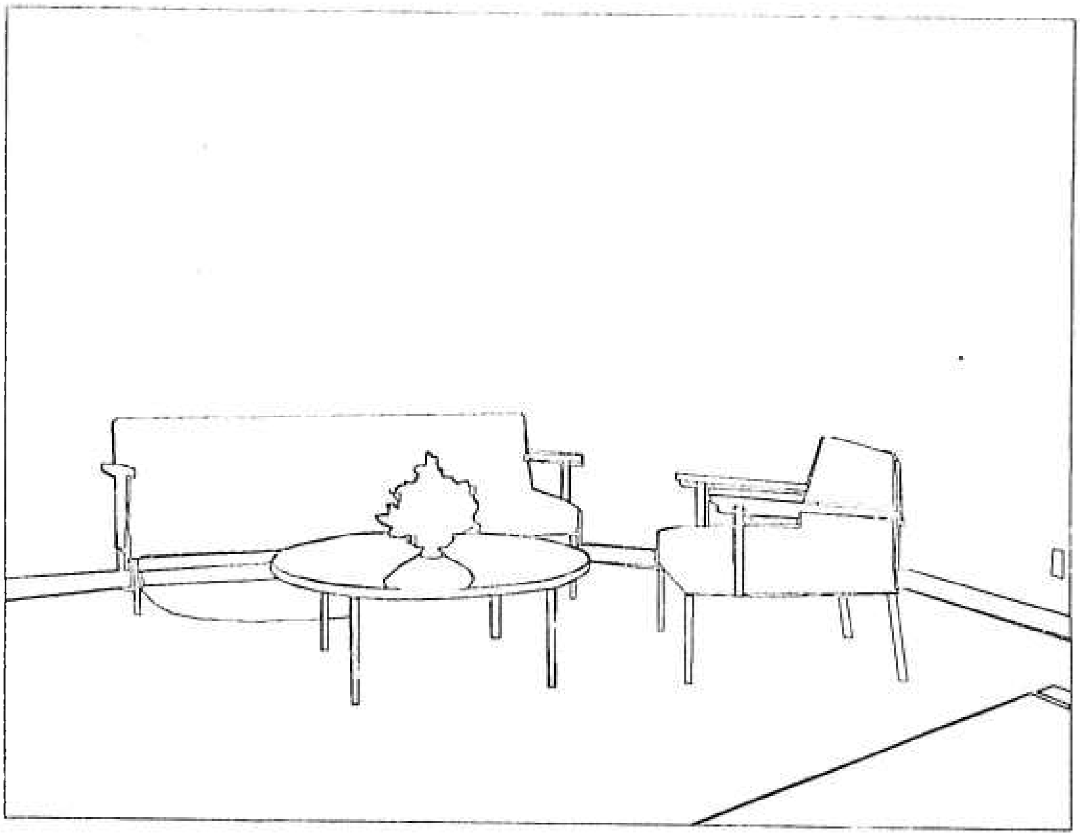


Figure 4.23. A sequence of steps restoring a portion of the wall.

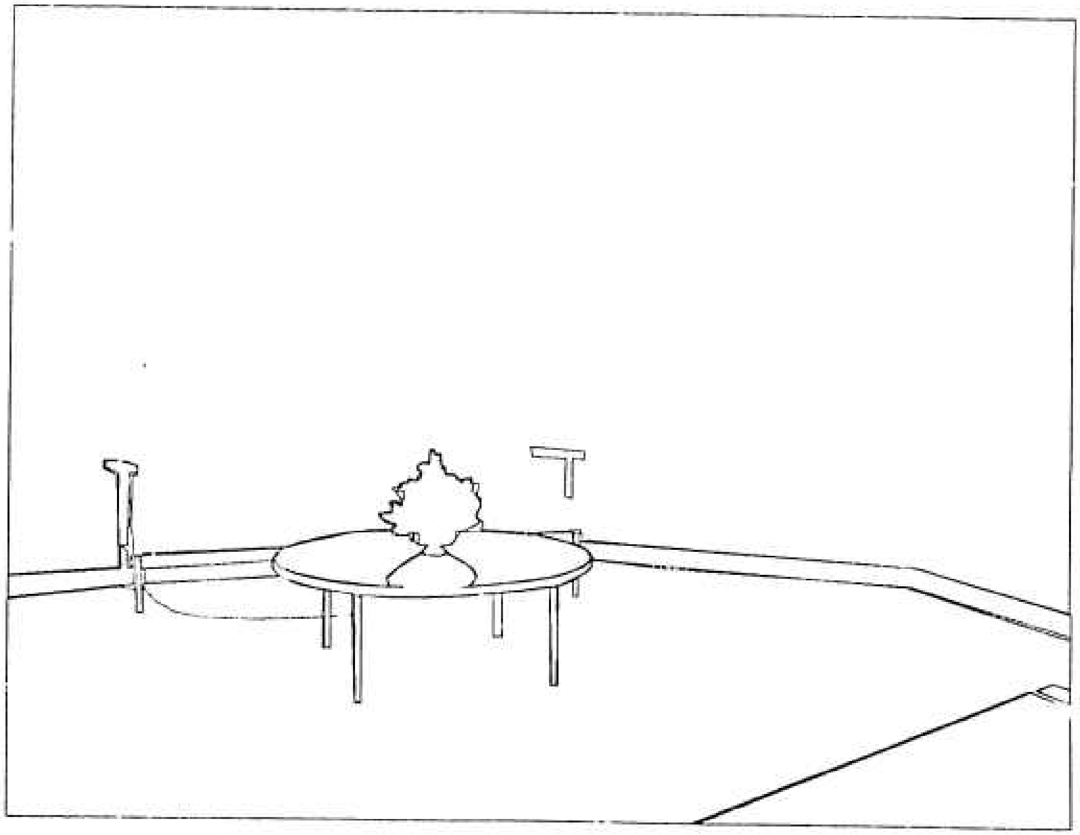


Figure 4.23 (continued). The upholstered section of the sofa is removed.

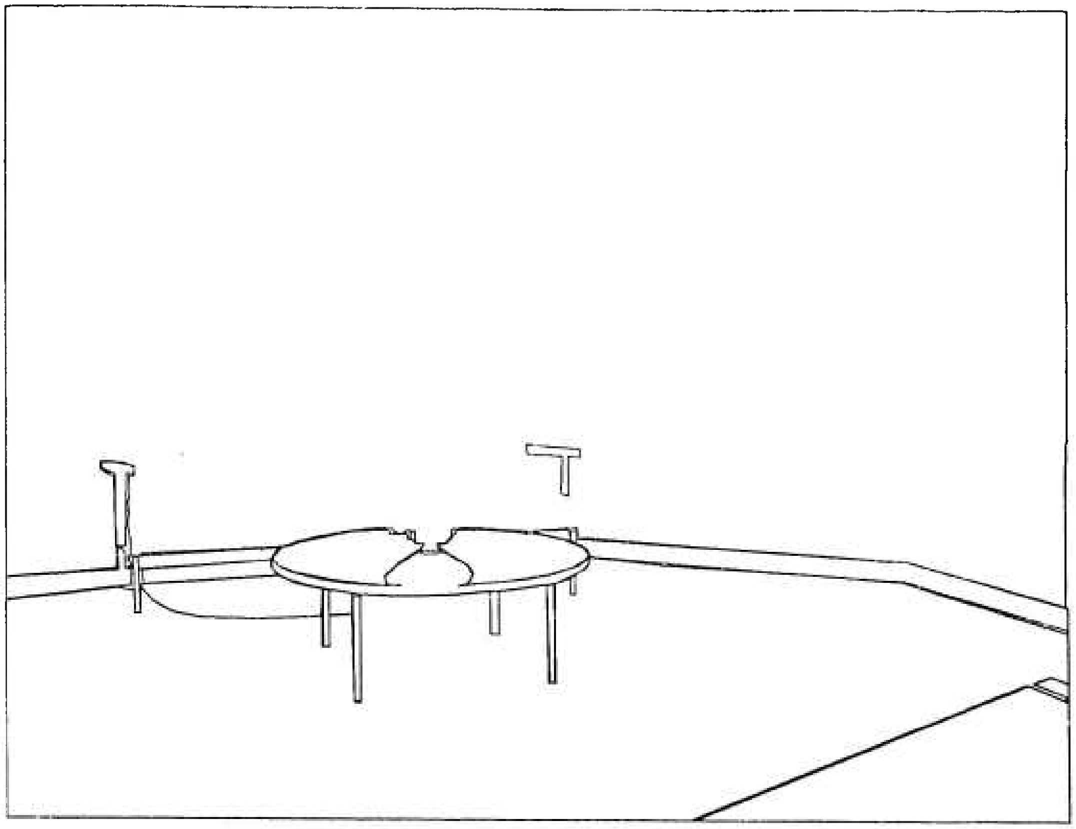


Figure 4.23 (continued). The vase is removed.

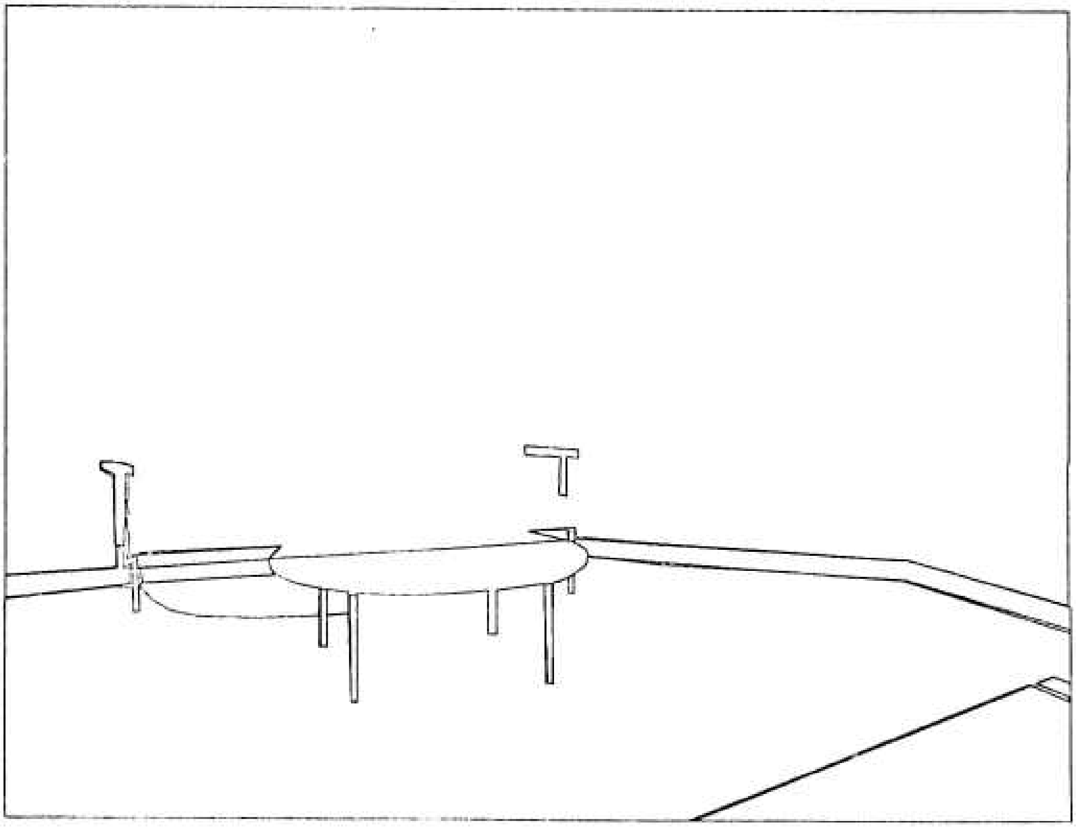


Figure 4.23 (continued). The sections of the table are removed.

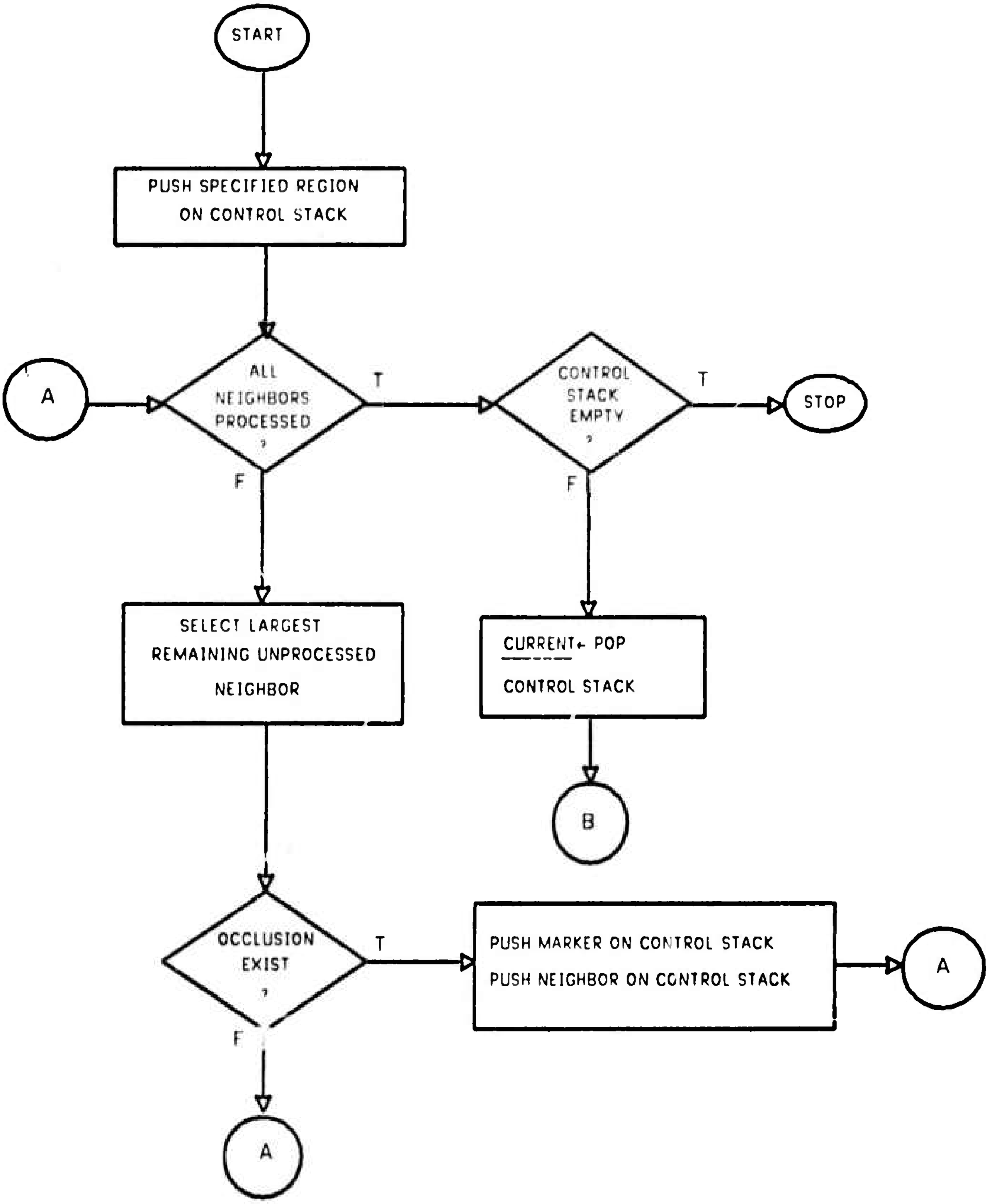

START
PUSH SPECIFIED REGION ON CONTROL STACK
A
ALL NEIGHBORS PROCESSED ?
T
F
CONTROL STACK EMPTY ?
T
STOP
F
SELECT LARGEST REMAINING UNPROCESSED NEIGHBOR
CURRENT← POP CONTROL STACK
B
OCCLUSION EXIST ?
T
PUSH MARKER ON CONTROL STACK
PUSH NEIGHBOR ON CONTROL STACK
A
F
A

Figure 4.24. Modified algorithm for detection of occlusions.

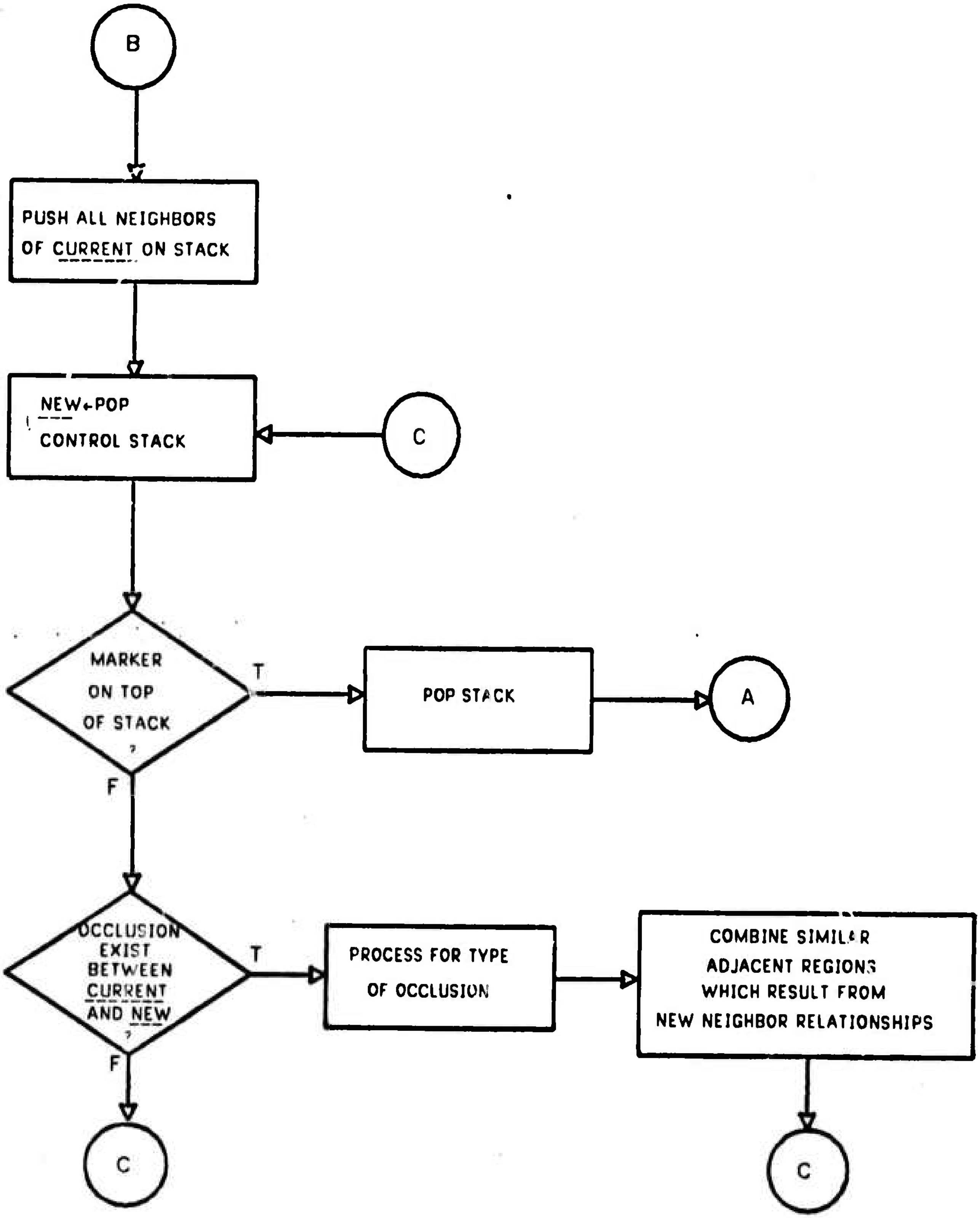

B
PUSH ALL NEIGHBORS OF CURRENT ON STACK
NEW←POP CONTROL STACK
C
MARKER ON TOP OF STACK ?
T
POP STACK
A
F
OCCLUSION EXIST BETWEEN CURRENT AND NEW ?
T
PROCESS FOR TYPE OF OCCLUSION
COMBINE SIMILAR ADJACENT REGIONS WHICH RESULT FROM NEW NEIGHBOR RELATIONSHIPS
F
C
C

Figure 4.24 (continued). Modified algorithm for detection of occlusions.

Notice that the recursive process has an additional beneficial side effect. It helps provide a solution to the fourth subcase listed under a type 2 occlusion. This was the situation that arose in figure 4.11 when an irregular occluding object extended beyond the bounds of the occluded region. When it comes time to check for additional regions which may have also been occluded and which may determine better boundaries, we find they have already been restored and are ready at hand.

When fully implemented the procedure just described should perform all the necessary functions for scenes of fairly regular construction. Specifying any region will direct the program to remove all occlusions and restore that area to its original form, or at least to a form which better fits the model

Implementation of Shadow and Highlight Knowledge

Implementation of shadow and highlight knowledge has not received as much attention as has been devoted to occlusion. Fortunately, many of the mechanisms necessary for the investigation of the subject are the same as those provided for occlusion removal. The subroutine which detects type 3 occlusions on the basis of similarity of properties was easly adjusted to detect shadows or highlights. The recomputation of boundaries is achieved by the same general merging procedure which eliminates the common portion of the border and reconnects along the uncommon part.

For low level detection of shadows and highlights we have not yet required all the conditions specified earlier. Our most critical check is for adjacent regions with values for hue and intensity meeting the criteria specified in the previous subsection. If this condition is met and if we have a shadow, we require that the affected region be adjacent to the image border or to an additional region that could be the cause of the shaded surface. For outdoor scenes the shadow causing area must be in the vertical upward direction. This heuristic is used because the sun is almost directly overhead for our scenes.

The final comments in this subsection address sequencing, i.e., when shadow and highlight restoration should be performed. In a completely asynchronous system, the knowledge source could make its contribution whenever sufficient evidence to evoke a response was present. Practically speaking, it is best to investigate for shadows or highlights before checking for possible occlusions. One reason for this is that the shadow check is usually simpler and less time consuming than an investigation for all types of occlusion. A second reason is that most of the instances of shadows or highlights are connected along a single border. Successful detection of a shaded surface would eliminate the troublesome type 5 occlusion from further consideration. Therefore, until the issues are better understood, we have decided to initiate shadow checks prior to occlusion checks in the algorithm presented in figure 4.24. Note that we must still recursively investigate possible occlusions of any detected shadow or highlight area.

Results

In this section we shall endeavor to more precisely lay out the capabilities of the interactive occlusion, shadow, and highlight subsystem. As implied earlier we have not yet completely implemented the final recursive control structure. We have constructed a detection mechanism which determines the proper type of occlusion for most regular surfaces of the type found in the room scene of figure 2.4.b. The detection is based upon estimation of directional neighbors and simulated relative range information. The neighbor calculation is based on the MBR (minimum bounding rectangle) technique described earlier.

In addition to the detection process there are a large number of complex subroutines which allow us to compute boundaries for the types of occluded objects found in the room scene. At this point in time the procedures are evoked by the user specifying the two regions involved in the occlusion. The kinds of occlusion presented in figures 4.7, 4.14, and 4.17 can all be corrected with the given mechanisms. The type of occlusion shown in figure 4.10.a can be corrected to the extent shown in figure 4.10.b but not as completely as shown in figure 4.10.c.

The last fundamental requirement needed to provide the basis for the implementation of an automatic subsystem is a subroutine which recomputes the properties (other than boundaries) of an occluded region. By far the most difficult requirement is recompution of two-dimensional and three-dimensional relationships. We have a program which effects the desired results for the initial control structure depicted in 4.22. We have yet to complete modifications which adapt it to the more complex control structure shown in figure 4.24.

Using the tools described above, and following the control structure of figure 4.24 we are able to derive the series of occlusion restorations shown in the series of figures, 4.25. The first picture shown in figure 4.25 is a slightly idealized result of the actual segmentation process. The legs of the chair are missing; they were simply not differentiated from the rug. The first action is to remove the shadowed area on the table caused by the vase (the segmentation process separated the edge of the table and the shadow as one piece). Notice also, that when the table is removed no problem will arise in reformulating the baseboard under the sofa. This is true even though the right sofa leg is restored first. The correct result, however, requires that strict attention be paid to neighbor recomputation. After reconstruction of the leg the baseboard only bears a three-dimensional relation to it (it is behind). Speaking two-dimensionally, only the table still lies between the two baseboards. This example underlines the need and complexity involved in maintaining proper relationships. As a final word, let us remark that the indentation remaining in the wall for the final result is due to an error in segmentation which did not separate the rear left leg of the sofa from the baseboard.

The capabilities and restrictions applicable to shadow and highlight detection and restoration were described in the last section. Utilizing the kinds of checks listed there (similarity, proximity, presence of shadow causing region) and the same control structure as for occlusions, we have been able to produce the shadow removals for the house scene shown in figures 4.26.

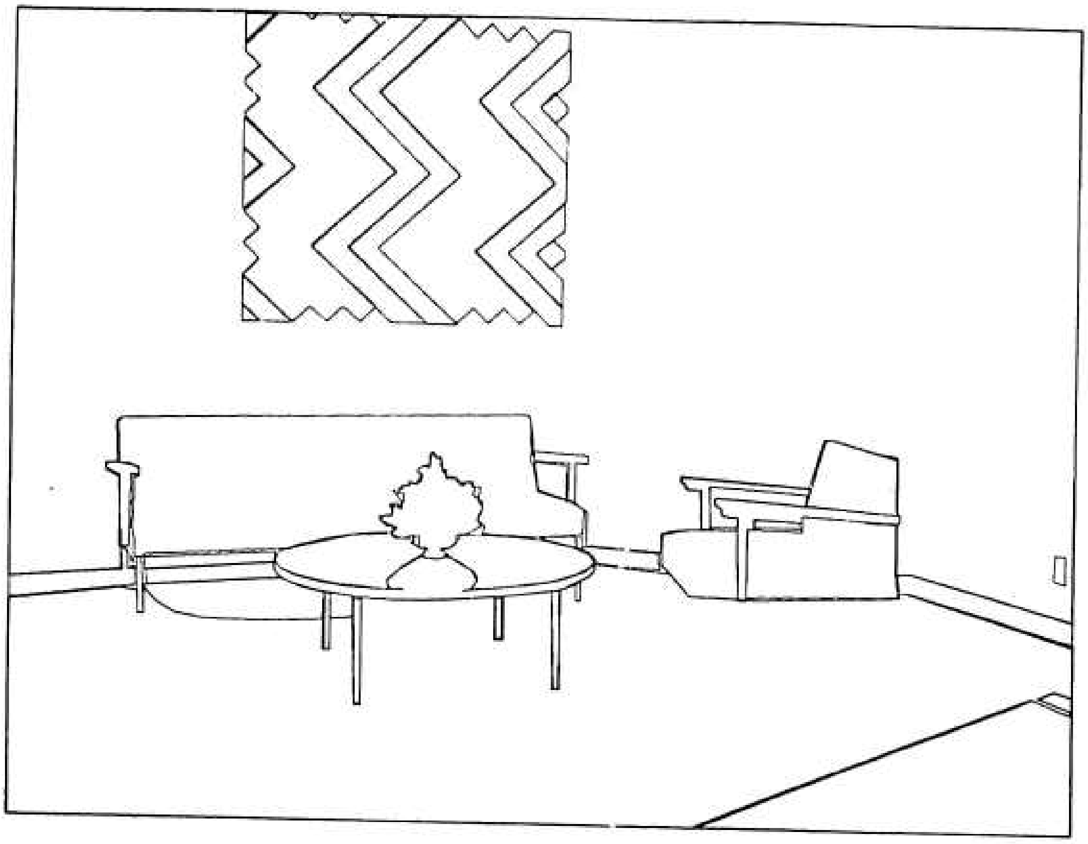


Figure 4.25. Series demonstrating complete removal of all occlusions in the room scene.

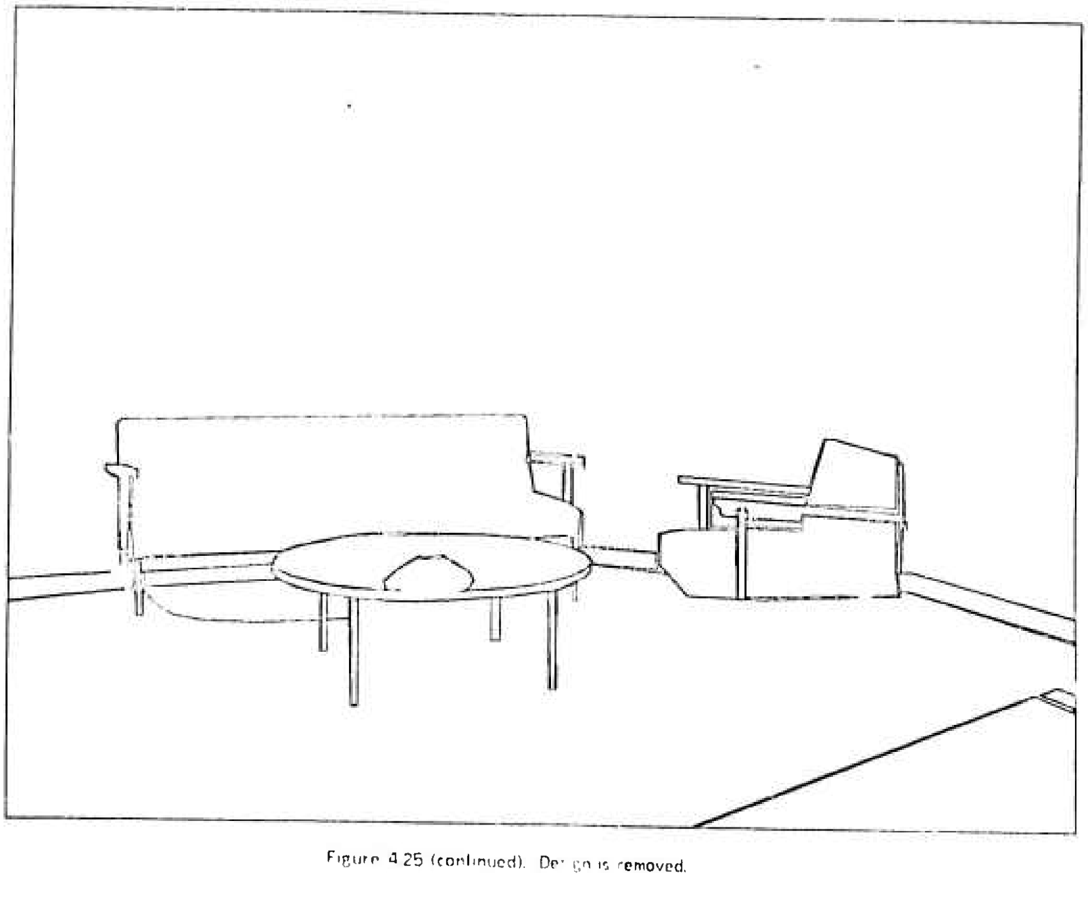


Figure 4.25 (continued). Design is removed.

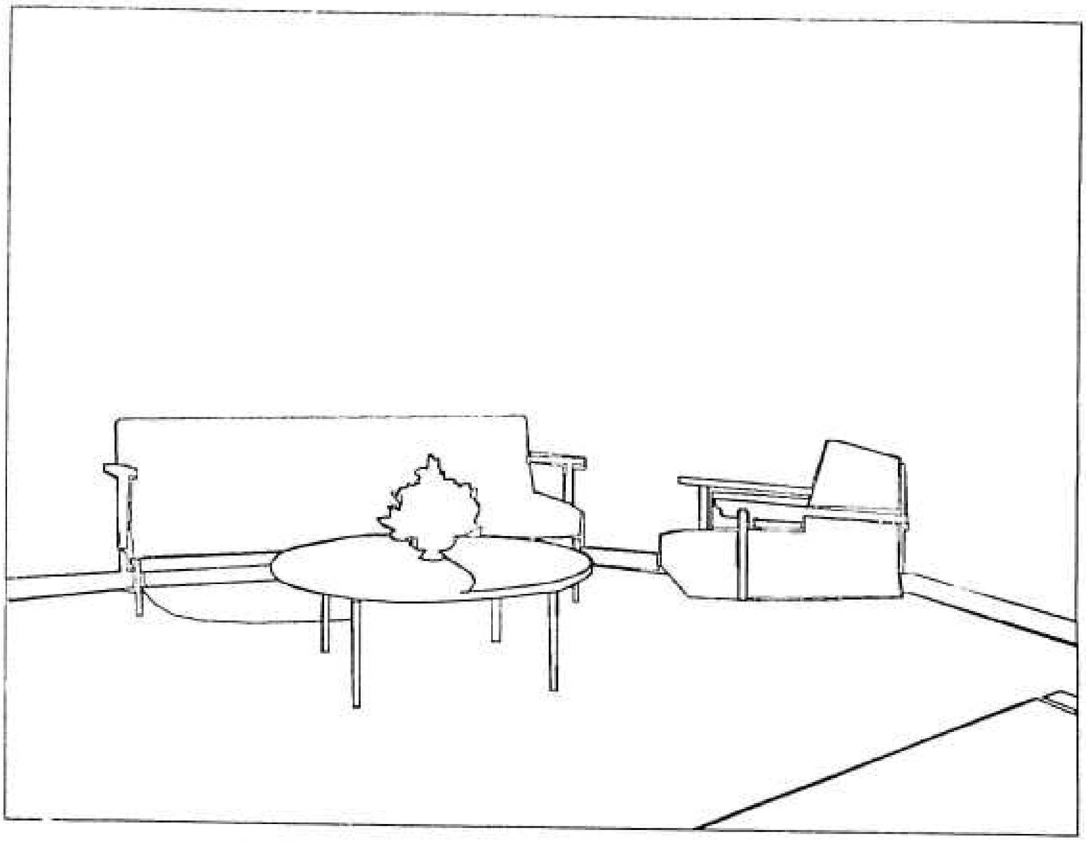


Figure 4.25 (continued). Shadow of vase is merged with table.

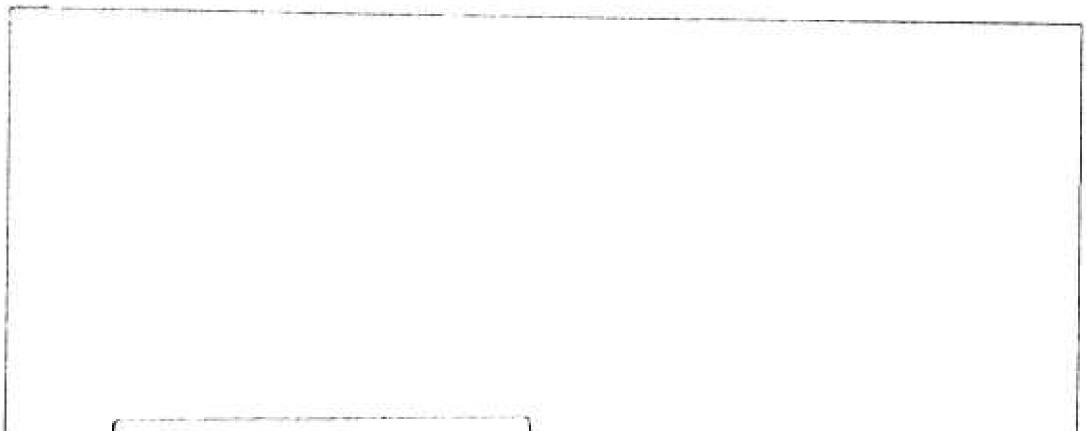


Figure 4.25 (continued). Vase is removed.

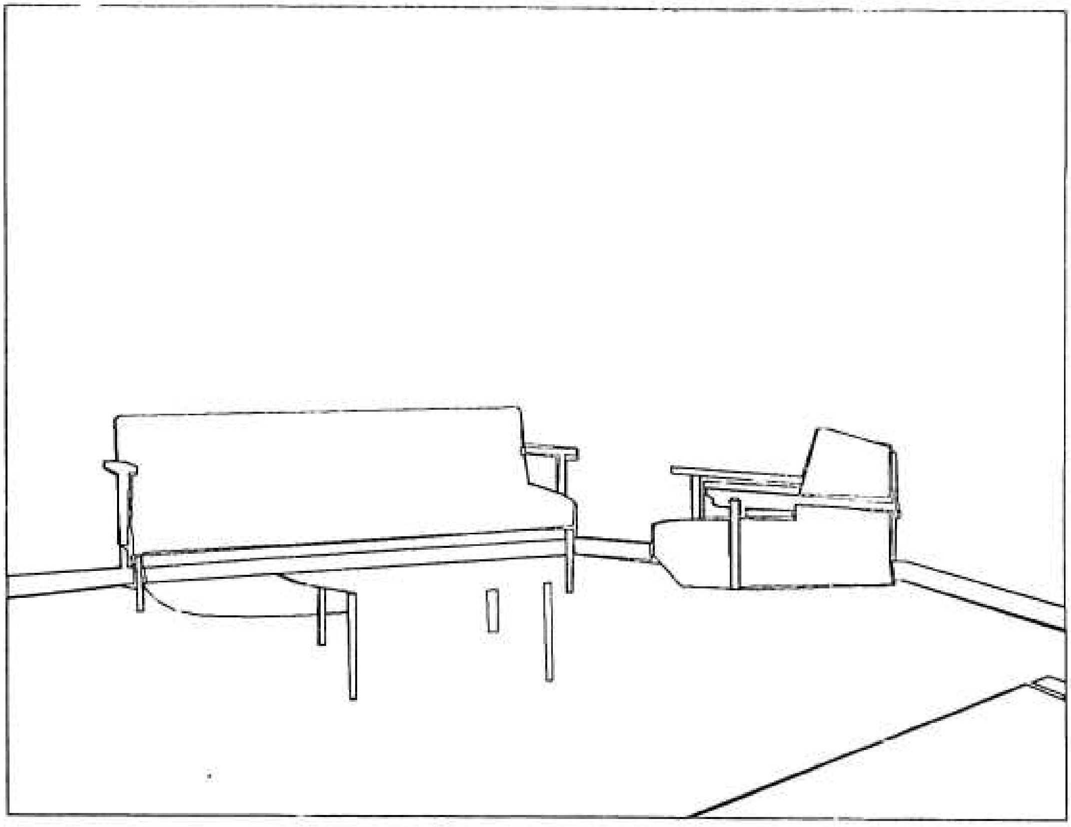


Figure 4.25 (continued). Table top is removed.

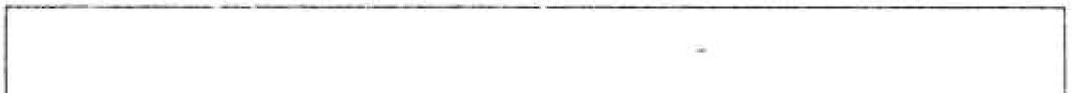


Figure 4.25 (continued). Upholstered part of chair is removed.

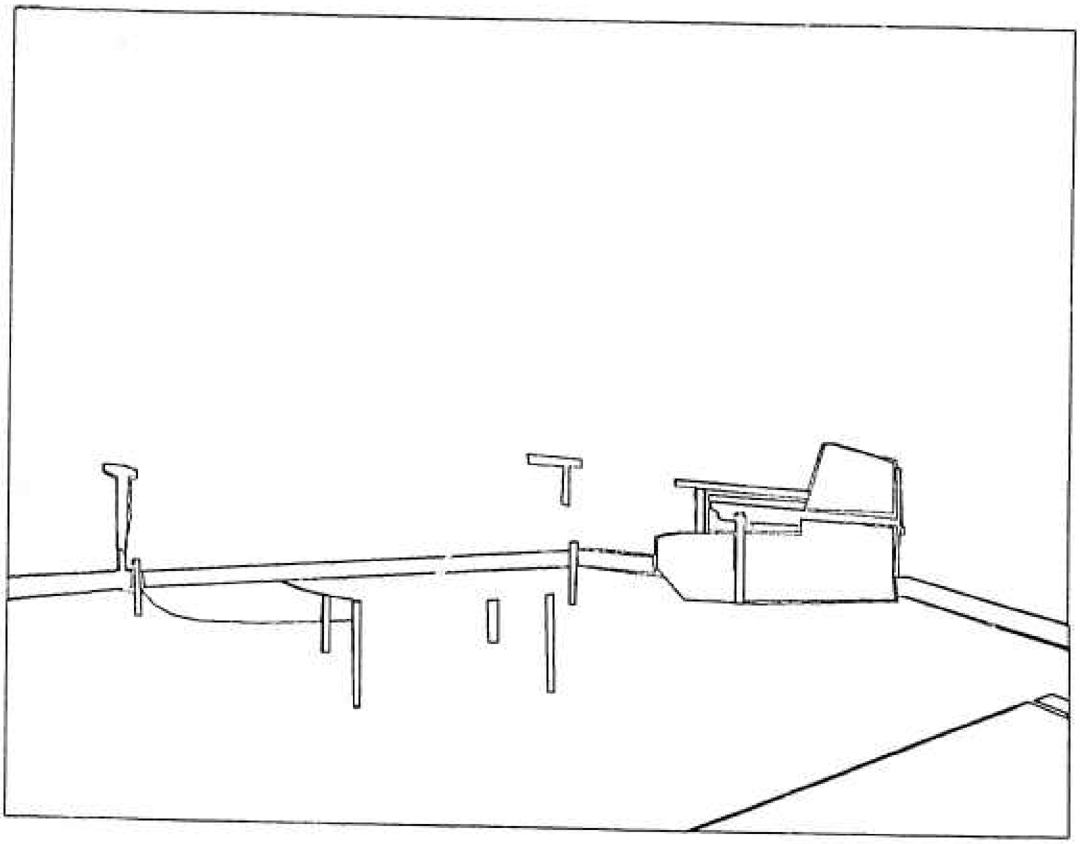


Figure 4.25 (continued). Strip of wall under sofa is merged.

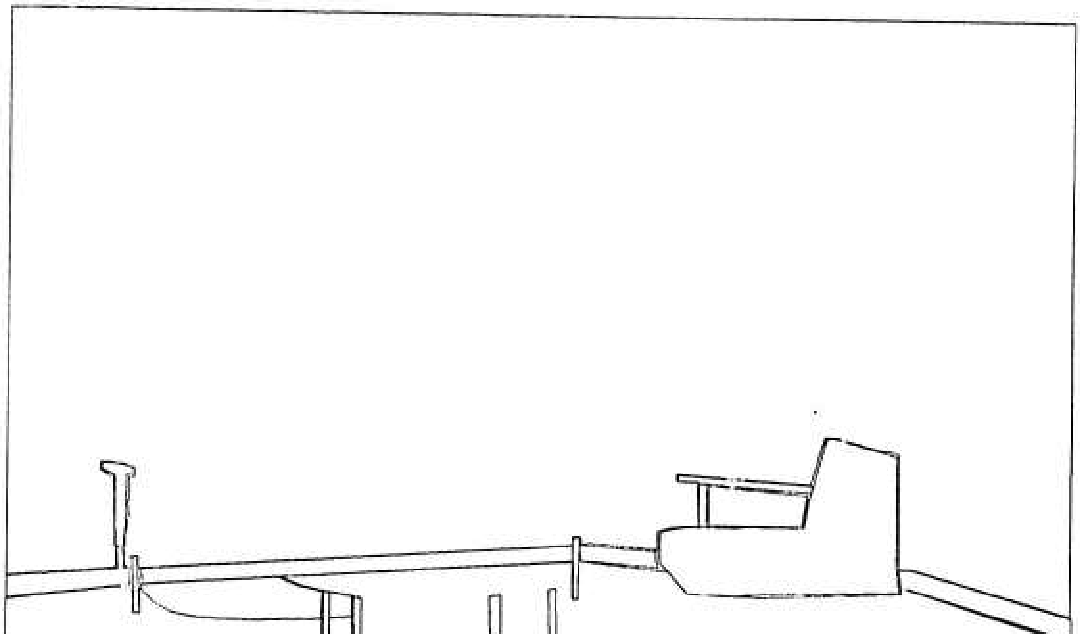


Figure 4.25 (continued). Right arm of sofa and front arm of chair are removed.

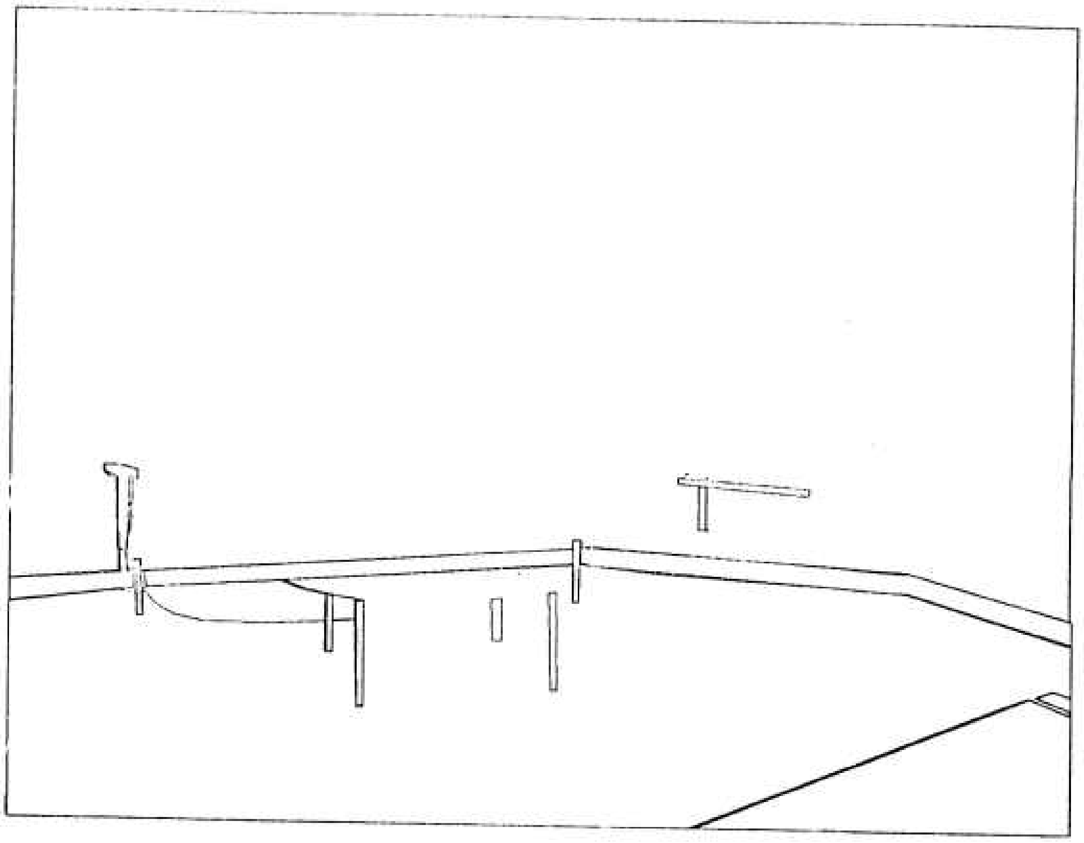


Figure 4.25 (continued). Section of wall under chair arm is merged.

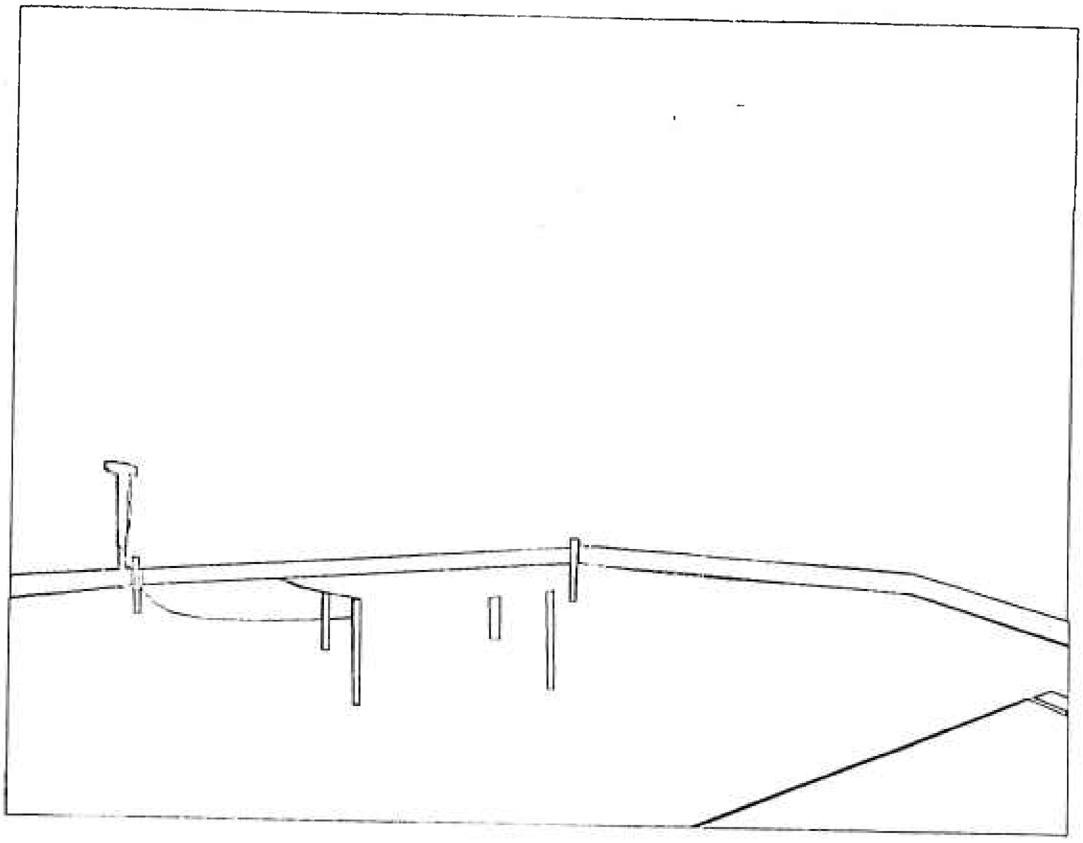


Figure 4.25 (continued). Chair arm is removed.

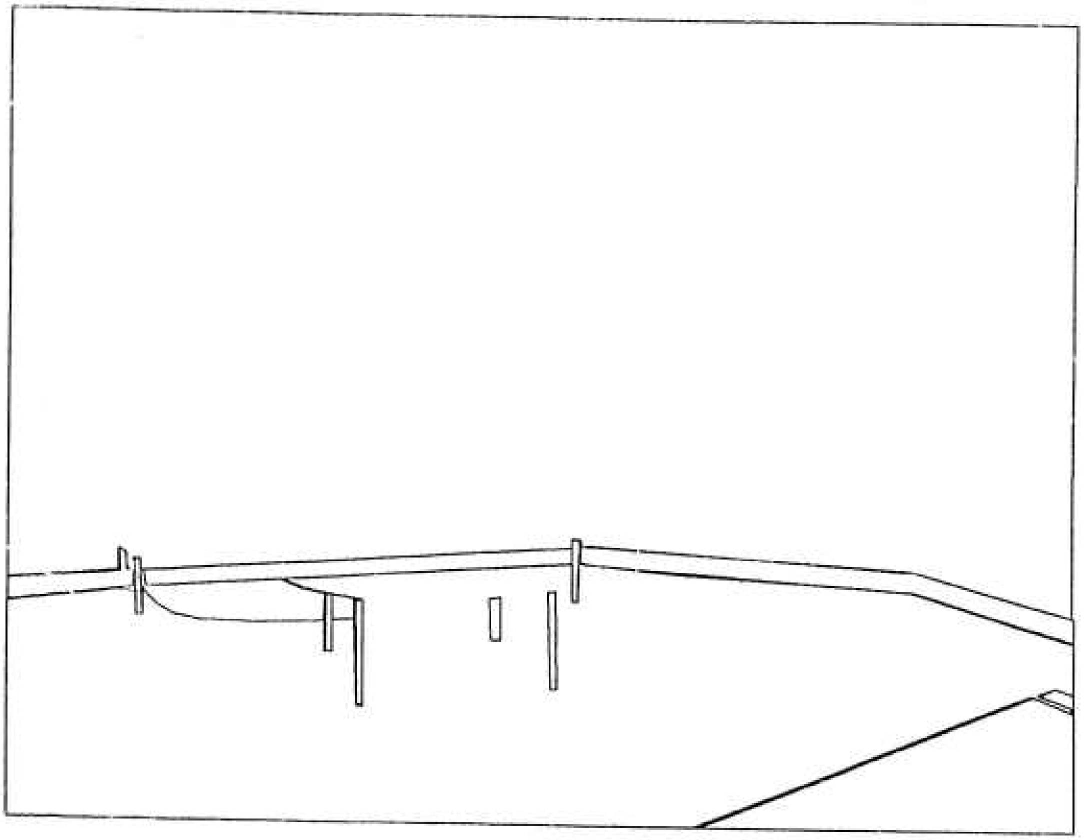


Figure 4.25 (continued). Sofa arm is removed.

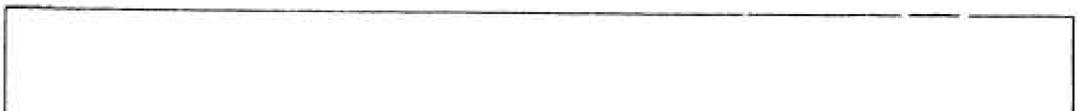


Figure 4.25 (continued). Sofa leg is removed.

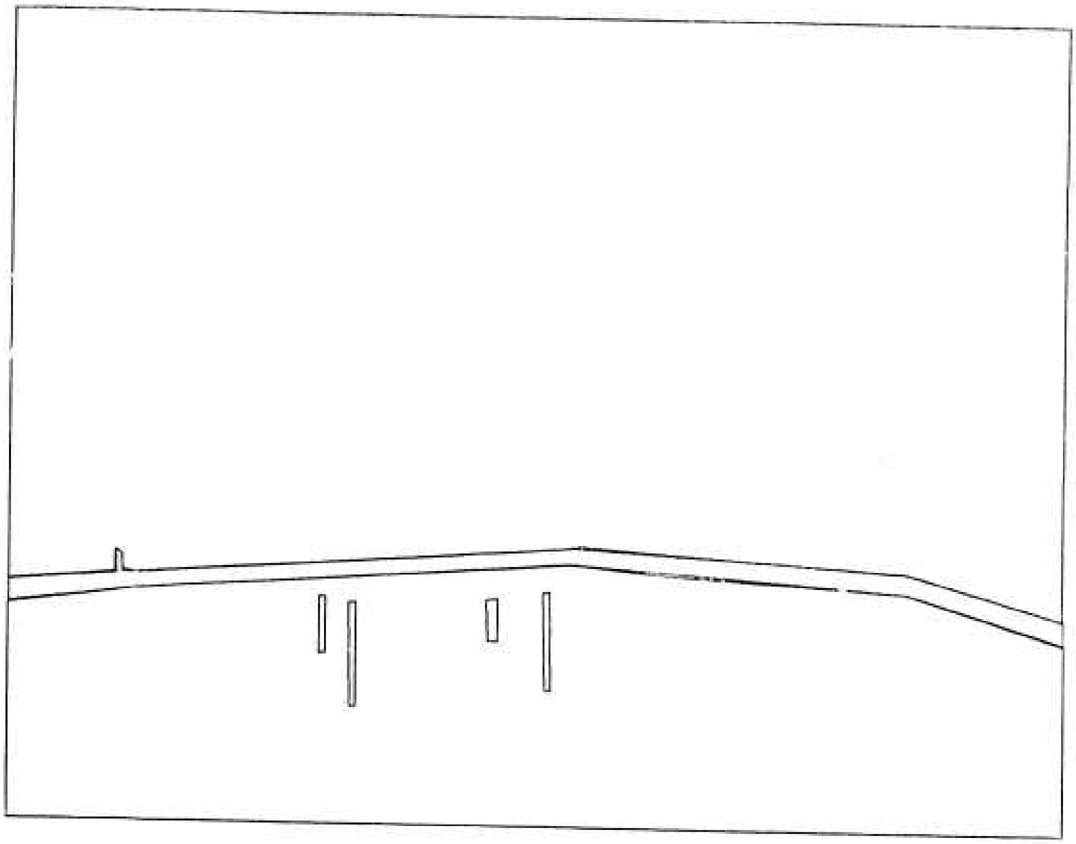


Figure 4.25 (continued). Shadow and highlight on rug are merged.

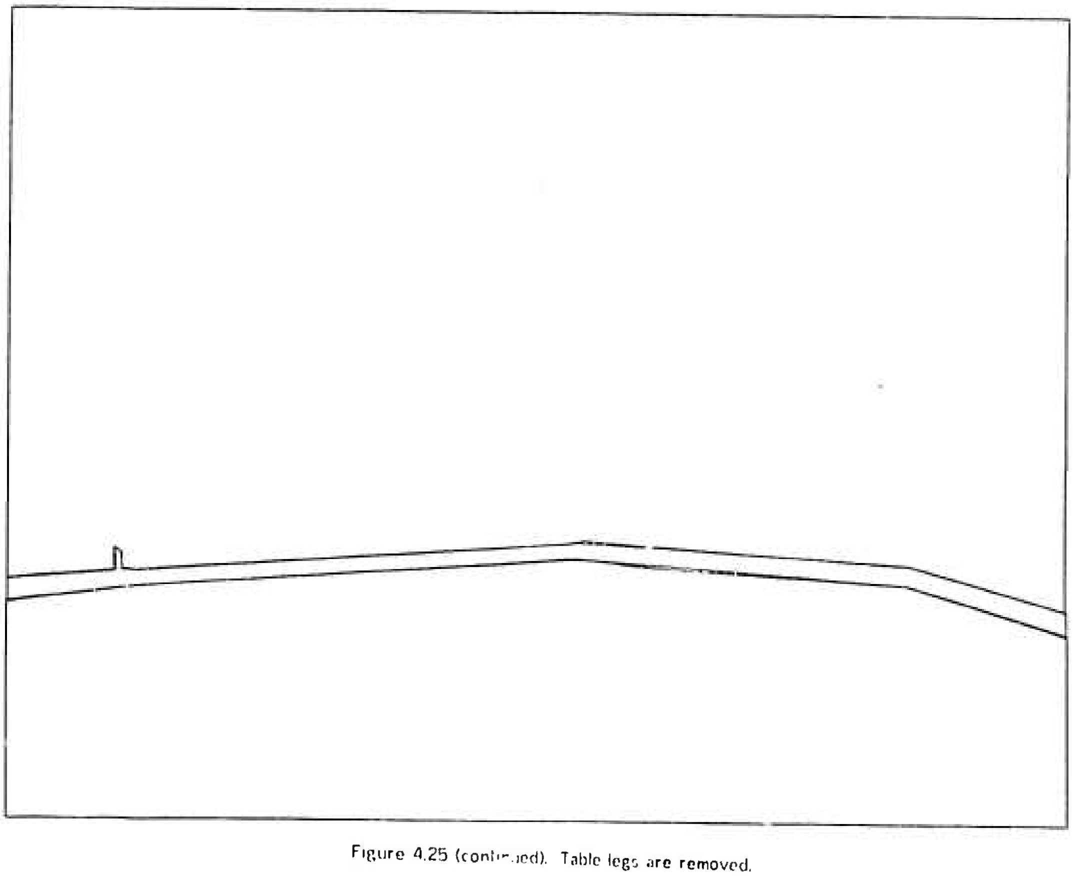


Figure 4.25 (continued). Table legs are removed.

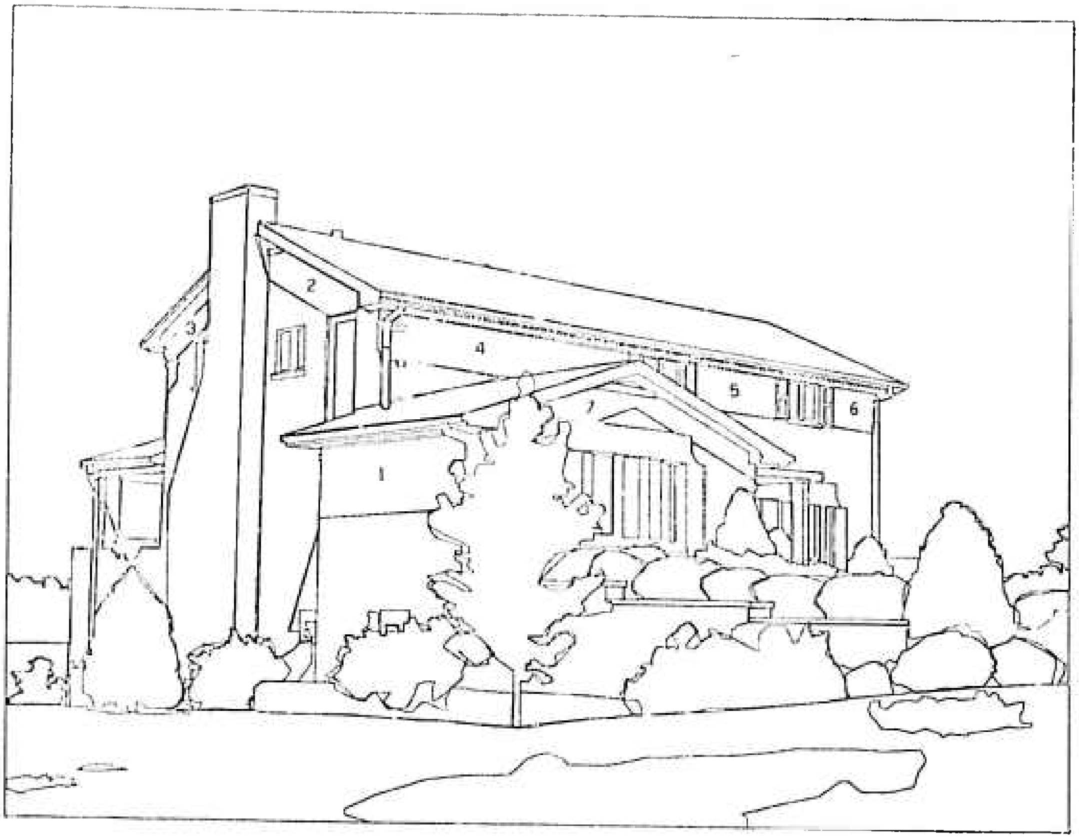


Figure 4.26. House with shadows under the eaves.

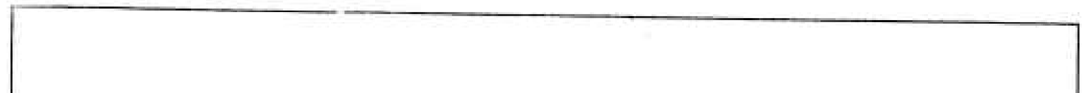


Figure 4.26. House with shadows under the eaves removed.

5 AVENUES FOR INCREASING SYSTEM PERFORMANCE

If practical use is to be made of the general segmentation scheme presented in this report, the real time processing figures (see the section on results in chapter 3), will have to be improved substantially (1 or 2 orders of magnitude). While we were developing the algorithm and investigating the feasibility of the segmentation process, we did not greatly concern ourselves with time and space issues. Now that the algorithm has proved to be of some worth, these issues assume more significance. One of the immediate goals of the vision group at CMU is to segment the remaining scenes in our data base. In addition to this task, we are undertaking analysis of Earth Resources Technology Satellite pictures, which are 16 times larger than the scenes we have processed. If this research effort is to be maintained the space and time issues assume paramount importance. In the remainder of this chapter we intend to discuss how gains in both domains can be made.

System Speed-up

Software Improvements

We shall first discuss improvement of performance in terms of realizeable goals within the existing system. The obvious optimizations that can be made by streamlining code and converting higher level programs to assembly language will not be discussed. The heart of the system, the picture accessing mechanism, is written in assembly language and has been optimized for the task it is designed to perform.

The first series of changes that we propose are in the area of improving the algorithm. This can be accomplished on two levels: improvement to the operator subroutines, and application of additional heuristics to the basic segmentation algorithm. For example, consider the smoothing operations which have been shown to take 66% of the total processing time (see Results, chapter 3). A change in the subroutine algorithm has successfully reduced the number of additions performed in the inner loop from 2n (n is the size of the window), to 4.[1] What is more, where formerly the number of additions grew linearly with the size of the smoothing window, now they remain constant. This has improved the time for the operations by a factor of 3, and reduced its share of the processing load to 43%. Another improvement being made along these lines involves ways of combining the smoothing, contraction, and expansion operations so that they can be performed with only one access to the data instead of three. This will not affect the CPU time for the process, but will reduce the Input/Output time to one third of its present value.

A third possible improvement that can be made in this area would affect the computation time for histogram calculations. The proposed change would not be to the histogram subroutine itself, but rather to the data structure of the segmentation

[1] This work has been accomplished by Keith Price who is a graduate student of the computer science department at CMU.

procedure. The majority of time required for the histogram computation is taken up in the calculation of the frequency with which the different density levels for the given parameter occur. This requires a complete scan of the parameter matrix. Currently the array which contains the frequencies of the density values for each set of histograms is discarded after use. We are suggesting that this array be associated with its corresponding template. Histograms could then be calculated for the processed segments extracted at a given level of recursion and the resulting frequency counts subtracted from the array associated with the template on that level. Frequency arrays would also need to be computed for all regions (except the largest one), which remained after masking out the processed segments. Since these regions will become templates, the arrays would have to be calculated eventually anyway. Subtracting these counts from the associated frequency array will now furnish the proper data for the calculation of the histogram for the largest remaining region in the template.

As mentioned previously, it is possible to employ additional heuristics within the structure of the basic algorithm which should produce increased performance. The first heuristic to command attention involves a priority of selection of sensory parameters in the computation of possible threshold limits. This step would be predicated on the fact that certain parameters embody more useful features than others. In the house scene, for example, we found the hue dimension to provide about 90% of the cutoff values during the processing of the picture. On the other hand, we have discovered that the "Y", "I", and "Q" parameters contribute very little to the processing of the entire range of scenes. What we are proposing, is that not all histograms for all sensory data be computed at one time. A precedence should be established for the order in which parameters are considered. If a histogram is found which can provide a mode meeting certain conditions, the search will proceed no farther. The precedence could be established on overall picture properties which might indicate the most helpful parameters. This could be done at the first level of extraction by examining the histograms for the entire scene to see which dimensions supply the most information. The adoption of the proposed heuristic might not produce as well defined segments as the current process does, but the careful establishment of adequate criteria for selecting the histogram peaks should produce acceptable results.

A second heuristic which should improve system performance is the use of "planning". We are speaking of planning in the sense used by Kelly (1970) in his face recognition program. Suppose that we reduce our picture by a factor of four in each dimension. This will leave a digital image of 150x200 pixels to process. If we employ the same procedure of recursive descent on this new construct we should extract a number of useful segments from the picture. The question is what detail will be lost and what will be the effect upon texture. In many cases heavily textured areas of the large scale pictures will have been smoothed and will fall out in their entirety. On the other hand, new heavily textured areas will have been created because of the compacting of detail. The issues with respect to texture are not entirely clear and will have to explored in great detail. The full scale picture is always available for close scrutiny if needed.

Processed segments that result from analysis of the reduced image can now be mapped into the corresponding portions of the full scale image to see if further refinement is necessary or desireable. Histograms of the area should detemine the matter. Uniform regions of medium to large size which possess light to medium texture should correspond rather well to what would be extracted by the segmentation process acting on the large scale picture. For these regions a single histogram check will suffice. Note that the time consuming smoothing operations of the high resolution image will not be necessary. In cases where further refinement is indicated we would probably want to enlarge the area of focus somewhat to be sure of maintaining integrity of structure. These areas would be processed in the normal way. In cases where the processed segments map back into busy areas some means will have to be devised for checking the preciseness of extracted boundaries. This might be done by correlating the histogram of the masked area with histograms of heavily textured portions of the high resolution image.

The last changes to software that we will discuss are directed towards improving I/O response. The two avenues to explore involve the picture accessing system and the PDP-10 monitor. The picture accessing mechanisms were designed to provide complete random access to any pixel in the matrix. The cost for such generality is always high. When the image representation does not fit in core (which is usually the case), the cost is paid in I/O operations. The access system operates like a paging system. For a number of reasons the page "unit" decided upon was one row of the picture. To process an entire image requires at least one disk access for each row. Since we are processing the picture sequentially, in most cases we do not need the full generality. By modifying the system we can input larger buffers of data and reduce our disk accesses significantly. The changes would not be extensive and should pay good dividends.

The modification proposed in the preceding paragraph will still require sequential input of data. Since the picture representational format is laid out in row major order, to access one pixel of a row means the row in its entirety must be read into core. A major modification of the picture format can be made to alleviate this situation. In this scheme page units will correspond to some window of the picture. This will allow us to treat specific areas of the picture without having to input irrelevant portions of the image. The savings in I/O should be substantial, but the implementation cost will be correspondingly high.

The final change that we propose concerning I/O improvement involves the buffering system of the PDP-10 monitor. Without going into the details of why, we were not able to take advantage of full buffered I/O. A modification to the monitor or a more extensive restructuring of the picture access mechanism would remedy this defect and allow some I/O operations without swapping of the program.

Hardware Improvements

The software modifications discussed above would gain us a speed-up factor of 10 to 30. This is unlikely to be sufficent, in the long run, for practical processing of large-scale pictures. To get the increases in performance which are needed, we must turn to functionally specialized architecture. A dedicated machine is one improvement that is immediately obvious. Large scale computer systems, however, are too expensive to serve in this capacity. The answer lies in smaller computer systems with highly specialized arithmetic units and multiprocessing capabilities. There must also be provisions for high bandwidth memories and secondary storage devices. Coupled with optimal software, such a system should provide speed-ups of two to three orders of magnitude.

If full realization of specialized machine architecture is not possible, there are some improvements which can be achieved with reasonable expenditures. The addition of a cache memory would speed up computation time and increase bandwidth significantly. The small inner loops and sequential memory access that is characteristic of the picture operations is made to order for a cache. Addition of an I/O processor with adequate buffering provisions could effectively eliminate I/O time.

Space Reduction

We can consider space reduction in terms of an outright decrease in storage requirements, as well as a decrease in bandwidth requirements. Reducing the bandwidth is an important adjunct to the speed-up in performance discussed in the last section. Typical data rates are 2 to 8 megabits per second, depending on the equipment. Practically speaking, time-sharing systems will reduce this by up to two orders of magnitude. Opportunities to reduce bandwidth are not as plentiful as opportunities for system speed-up. Some of the proposals of the last section would also have the effect of decreasing both bandwidth and storage requirements. Planning, for instance, would have this effect. Restrictive selection of histograms would also decrease the bandwidth requirement. Other than that, space reductions seem to require an outright decrease in size of the sensory data base. This can be accomplished by allowing smaller resolution, eliminating sensory parameters, or compacting the data. We could probably cut the pixel size from 8 to 6 using histogram equalization without any serious effect on the outcome. Before taking the other courses of action, however, we would need to know more about the effect upon the segmentation process.

6 CONCLUSIONS AND DIRECTIONS FOR FURTHER RESEARCH

Conclusions

The Perceptual Model

One thing that has been reaffirmed by this body of research is the viability of the model that we proposed earlier. It has proved flexible enough to permit both independent construction of knowledge modules and provide a guiding framework for the development of a general system. It has also been demonstrated that the features that provide for processing data of an errorful nature will become critical in the time to come. The use of imperfect mechanisms compound errorful sensory data to give even more errorful output. We can expect the problem to become even greater as more knowledge sources become available. A recognition knowledge source, for instance, can make improper identification on the basic of incomplete or erroneous attributes. Multiple representations which reside in a global data base thereby providing alternate paths of analysis, indeed offer an attractive solution.

Methodology

We believe it has also been shown, at least implicitly, that independent development of knowledge sources within a specific framework offers a reasonable way of coming to grips with very large problems. It is quite interesting to note the different paths that this process took in the development of the two different knowledge sources. In the case of segmentation there was already available a large body of knowledge. We knew what kinds of effects to expect from each operator. We had only to provide a picture processor and a number of these image data operators to the experimenter. The main research effort lay in extending the range of these operators and combining them in ways that would produce new results. On the other hand, almost no previous work has been done concerning the role of occlusion, shadows, abd highlights in natural scenes. In this case the human had to provide all of the initial phases of analysis. Invariants had to be isolated that could classify a number of types of occlusion. Principles had to be extracted which permitted boundary restorations for specific classes of occlusions and shadows. As the problem became better understood primative routines were developed which could manipulate the data structure. Eventually a large interactive graphic subsystem became available for a wider range of experimentation. The common factor to note in both these cases is that the methodology provides a starting point and method of development to what oftentimes seems an insoluble problem.

Segmentation

By utilizing multiple sources of sensory data and combining existing techniques, we have been successful in achieving a reasonable first level segmentation for some very complex natural scenes. To the best of our knowledge, the range of pictures that we have successfully dealt with is greater than that attempted by any previous system. The major factors in the segmentation process that have contributed to this success are: use of multiple sources of sensory data, use of the thresholding operator, adequate handling of the texture problem, effective integration of existing picture processing techniqes, and progressive isolation of unprocessed portions of the image. Multiple sources of data are inportant because one parameter may offer an indication of discontinuities when the others all appear uniform. The thresholding process has proven to be the most versatile of the region isolation techniques. It is more accurate than region growing, more robust than edge detection, and has the additional feature that it produces closed regions for easy extraction. Textured regions have to be isolated for special treatment. A crude yet effective method of establishing high frequency, high amplitude edge points per unit area fulfills this requirement. The difficulty of the segmentation task requires use of many picture operators. The system must not only utilise threshold and texture operators, but also make effective use of smoothing, contraction, expansion, following, and masking techniques. The most critical step in the segmentation process, at least in terms of effecting a reasonable degree of segmentation, is progressive isolation of unprocessed portions of the image. This allows accurate analysis of a relatively small area without interference of sensory data from unrelated portions of the image. The basic algrothm provides this when enough uniform regions are extracted from the picture to leave unconnected unprocessed sections behind. Some pictures, however, do not provide sufficiently rich variations in sensory input to isolate more that one or two areas by thresholding along some dimension of uniformity. In these cases we have shown the necessity of pursuing other means of extracting parts of the image for further analysis. This is accomplished by estimating homogeneous and heavily textured sections of the picture which are then further refined with the basic algorithm.

Occlusions, Shadows, and Highlights

No one questions the importance of adequately handling the effects of occlusions, shadows, and highlights upon natural scenes if reasonable recognition on a regular basis is ever to be achieved. We have made a first effort to treat some of the issues involved with occurrences of these conditions. We feel that one of our most important contributions has been the formalization of the knowledge by case analysis of several different types of occurrences of these phenomena. This has allowed us to identify certain invariants which help in the detection of the conditions. The invariants or local clues are: proximity, discontinuity, and dissimilarity in the case of occlusions; and proximity and similarity in the case of shadows and highlights. We have also identified invariants of continuity within the different types that have permitted us to reconstruct boundaries of hidden, shadowed or highlighted surfaces in some simple cases.

Conclusions

Directions for Further Research

The General System

The basic requirements of a general vision system have only begun to be explored. There remain, of course, all the previously proposed knowledge sources which need to be constructed, but there is the recognition module, especially, which should receive the most emphasis. Implementation of an identification process will complete the skeleton required for a minimal functioning automatic system. There are a number of issues involving representation of objects and model construction that have never, to our mind, been answered satisfactorily for large image understanding systems. There exist important questions concerning problems of how to correct erroneous segmentations. Procedures need to be constructed which can trim regions which extend beyond actual boundaries of objects. On the other hand, regions will often have to be joined to effect correct identification. Another critical issue is the construction of a matching procedure which will compare structures in the data base with prespecified knowledge contained in object and world models. Many of these issues have already been investigated to a limited extent and will be the subject of a forthcoming report.

Methodology

Additional methods of knowledge acquisition are necessary for future research. One path that we have begun to explore along these lines utilizes an experimental system which allows the study of the protocols of humans as they try to indentify scenes and objects which are not visible to them.[1] The system consists of two graphic terminals and an interfacing program. The subject is able to ask various simple questions concerning properties of the scene. The experimenter sees these questions repeated on his own screen and can provide answers from his own analysis of the scene which he has in front of him. The entire process is recorded for later analysis. We are hopeful that this line of investigation will serve a twofold purpose. In the first place, we hope that the process used by humans in determining unobserved scenes will be useful in providing knowledge which can be generalized to machine use. Secondly, we expect the experiment to furnish us with some insight for extending the experiment to capture other types of knowledge.

Segmentation

There are a number of aspects of the segmentation process that require further investigation. In chapter 5 we proposed directions for research to improve performance of the system. We also need to gain some appreciation of the range of the algorithm. It should be determined just what types of pictures the process will successfully deal with. If the procedure fails for certain images, different means

[1] This work was performed in conjunction with Omer Aygun of the Department of Architecture at CMU. A report of initial findings is now in progress.

should be employed to partition them into portions which are, perhaps, more amenable to analysis.

Another aspect of the process which could stand a good deal of improvement Is the texture analysis. It must be determined just what information can be provided by the various operators that are available. Once this is established, an obvious avenue of investigation suggests itself, i.e., implementation of texture as one of the parameters for the recursive descent segmentation process. It should be treated just like any other source of sensory information.

Occlusions, Shadows, and Highlights

There are a number of lines of research that can be pursued in the area of occlusions, shadows, and highlights. Refinement of the case analysis is needed to more closely isolate invariants of occlusion properties. The issues concerning restoration of boundaries for the irregular shapes found in outdoor scenes require a better understanding before substantial improvement in this area can be expected. The line of investigation which seems most promising at this time involves the implementation of heuristics that can speed up the detection process for shadows and occlusions. If we extend the dissimilarity requirement for occlusion so that we require two regions to be dissimilar in all properties for an occlusion relationship to exist, we can reduce the number of candidates. For example, if range is available, the walls, design, baseboards, and rug of the room scene would all be considered as one region because of the similarity of range attributes. This leaves only the chair and sofa as candidates for occlusion. They could then be removed with the standard "restoration" mechanism. The simplification of the house scene would be even more striking.

7 REFERENCES

Bajcsy, R. (1972), Computer Identification of Textured Visual Scenes, AIM-180, Ph. D. Thesis, Stanford University.

Barrow, H. G., and Popplestone, R. J. (1971), Relational descriptions in picture processing, Machine Intelligence 6. Meltzer, B., and Michie, D. (eds.), University Press, Edinburgh, pp. 377-396.

Brice, C. R., and Fennema, C. L. (1970), Scene analysis using regions, Artificial Intelligence, 1, pp. 205-226.

Clowes, M. B. (1971), On seeing things, AI Journal, Spring 1971.

Duda, R. O., and Hart, P. E. (1973), Picture Processing and Scene Analysis. Wiley, New York.

Ejiri, M., Uno, T., Yoda H., Goto, T., and Takeyasu, K. (1971), An intelligent robot with cognition and decision-making ability, In: Proc. IJCAI-2. London: British Computer Society, pp. 350-358.

Erman, L. D., and Lesser, V. R. (1975), A multi-level organization for problem solving using many, diverse cooperating sources of knowledge, Department of Computer Science, Carnegie-Mellon University, March, 1975.

Falk, G. (1970), Computer interpretation of imperfect line-data as a three-dimension scene, AIM-132, Stanford University, August 1970.

Feldman, J. A., Feldman, G. M., Falk, G., Grape, G., Pearlman, J., Sobel, I., and Tenenbaum, J. M. (1969), The Stanford hand-eye project, In: Proc. IJCAI. Washington D. C., pp. 521-526a.

Grape, G. R. (1973), Model Based (Intermediate-Level) Computer Vision, AIM-201, Ph. D. Thesis, Stanford University, May 1973.

Guzman, A. (1968), Computer Recognition of Three-Dimensional Objects in a Visual Scene, MAC-TR-59, Ph. D. Thesis, MIT Project MAC.

Hayes, K. C., and Rosenfeld, A. (1972), Efficient edge detectors and their applications, TR-207, University of Maryland, November 1972.

Hueckel, M. H. (1973), A local visual operator which recognizes edges and lines, JACM 20, 1973, pp. 634-647.

Huffman, D. A. (1971), Impossible objects as nonsense sentences, In: Machine Intelligence 6. Meltzer, B., and Mitchie, D. (eds.), Edinburgh: University Press, pp. 295-323.

References

Kelly, M. D. (1970), Visual Identification of People by Computer, AIM-130, Ph. D. Thesis, Stanford University.

Kriz, S. (1973), Hardware for high-speed digital vector drawing, 1973 SID International Symposium Digest, May 1973, New York, pp. 52-53.

Lesser, V. R., Fennel, R. D., Erman, L. D., and Reddy, D. R. (1974), Organization of the Hearsay II speech understanding system, Proc. IEEE Symp. Speech Recognition. Pittsburgh, Pa., pp. 11-21. Reprinted in: IEEE Trans. on Acoustics, Speech, and Siginal Processing, ASSP-23, no. 1, Feb., 1975, pp. 11-23.

Levine, M. D., O'Handley, D.A., and Yagi, G. M. (1973), Computer determination of depth maps, Computer Graphics and Image Processing, 2, pp. 131-150.

Lieberman, L. (1974), Computer Recognition and Description of Natural Scenes, Moore School Report 74-08, Ph. D. Thesis, University of Pennsylvania.

MacLeod, I. D. G. (1972), Comments on techniques for edge detection, Proc. IEEE 60, March 1972, p. 344.

Mendelsohn, M. L., Mayall, B. H., and Prewitt, J. M. S. (1968), Approaches to the automation of chromosome analysis, Image Processing in Biological Science, Ramsey, D. M. (ed.), University of California Press, Berkely and Los Angeles, 1968, pp. 119-136.

Montanari, U., and Reddy, R. (1971), Computer processing of natural scenes: some unsolved problems, Artificial Intelligence, AGARD Conference Proceedings No. 94-71, London, paper no. 12.

Nilsson, N. J. (1969), A mobile automaton: an application of artificial intelligence tecniques, In: Proc. IJCAI. Washington D. C., pp. 509-521.

O'Handley, D. A. (1973), Scene analysis in support of a Mars rover, Computer Graphics and Image Processing, 2, pp. 281-297.

Pingle, K. K., and Tenenbaum, J. M. (1971), An Accommodating Edge Follower, In: Proc. IJCAI-2. British Computer Society, London, pp. 1-7.

Reddy, D. R., Davis, W. J., Ohlander, R. B., and Bihary, D. J. (1973), Computer analysis of neuronal structure, Intracellular Staining in Neurobiology. Kater, S. B., and Nicholson, C. (eds.), Springer-Verlag, New York.

Reddy, D. R., Erman, L. D., and Neely, R. B. (1973a), A model and a system for machine perception of speech, IEEE Trans. Audio and Electroacoustics, AU-21, 3, pp. 229-238.

Reddy, D. R., Erman, L. D., Fennell, R. D., and Neely, R. B. (1973b), The HEARSAY speech understanding system: an example of the recognition process, In: Proc. IJCAI-3, Stanford, California, August, 1973.

References

Roberts, L. G. (1963), Machine Perception of Three-Dimensional Solids, *Optical and Electro Optical Information Processing*. Tippett, J. T., et. al. (eds.), MIT Press, Cambridge, pp. 159-197.

Rosen, B. (1973), The architecture of a high-performance graphic display terminal, *1973 SID International Symposium Digest*, May 1973, New York, pp. 50-51.

Rosenfeld, A., (1969), Picture processing by computer, *Computing Surveys 1*, September 1969, pp. 147-176.

Rosenfeld, A. (1969a), *Picture Processing by Computer*. Academic Press, New York.

Rosenfeld, A. (1970), A nonlinear edge detection technique, *Proc. IEEE 58*, May 1970, pp. 814-816.

Rosenfeld, A., and Troy, E. B. (1970a), Visual texture analysis, In: *Proceedings of the UMR-Mervin J. Kelly Communications Conference*, Rolla, Mo., October 1970, paper no. 10-1.

Rosenfeld, A., and Thurston, M. (1971), Edge and curve detection for visual scene analysis, *IEEE Trans. on Computers C-20*, May 1971, pp. 562-569.

Rosenfeld, A. (1972), Picture processing: 1972, *Computer Graphics and Image Processing 1*, December 1972, pp. 394-416.

Rosenfeld, A., Thurston, M., and Lee, Y. H. (1972a), Edge and curve detection: further experiments, *IEEE Trans. on Computers*, *C-21*, pp. 677-715.

Rosenfeld, A. (1973), Progress in picture processing: 1969-71, *Computing Surveys 5*, June 1973.

Shirai, Y. (1972), A heterarchical program for recognition of polyhedra, AI Memo No. 263, MIT AI Laboratory, June 1972.

Sutton, R. N., and Hall, E. L. (1972) Texture measures for automatic classification of pulmonary disease, *IEEE Trans. on Computers*, *C-21*, July 1972, pp. 667-676.

Strand, R. C. (1972) Optical-image recognition experiments in the track chambers of high-energy physics, *Proc. IEEE 60*, July 1972, pp. 1122-1137.

Stevens, M. E. (guest ed.) (1970), Special Issue on optical Character Recognition, *Pattern Recognition 2*, September 1970, pp. 145-214.

Tenenbaum, J. M. (1973), Object Recognition in multi-sensory scene analysis, SRI Technical Report 84.

Tenenbaum, J. M., Garvey, T. D., Weyl, S., and Wolf, H. C. (1974), An Interactive Facility for Scene Analysis Research, SRI Technical Report 87.

References

Tomita, F., Yachida, M., and Tsuji, S. (1973), Detection of homogeneous regions by structural analysis, In: Proc. IJCAI-3, Stanford, California, August, 1973, pp. 564-571.

USC (1973), Image Processing Research, USCEE report 444, University of Southern California, 1973.

Waltz, D. L. (1972), Generating semantic descriptions from drawings of scenes with shadows, AI TR-271, MIT AI Laboratory, November 1972.

Winston, P. H. (1970), Learning Structural Descriptions From Examples, MAC-TR-76, Ph. D. Thesis, MIT Project MAC.

Woods, W. A., Makhoul, J. (1973), Mechanical inference problems in continuous speech understanding, In: Proc. IJCAI-3, Stanford, California, August, 1973, pp. 200-207.

Yakimovsky, Y. (1973), Scene Analysis Using a Semantic Base for Region Growing, AIM-209, Ph. D. Thesis, Stanford University.