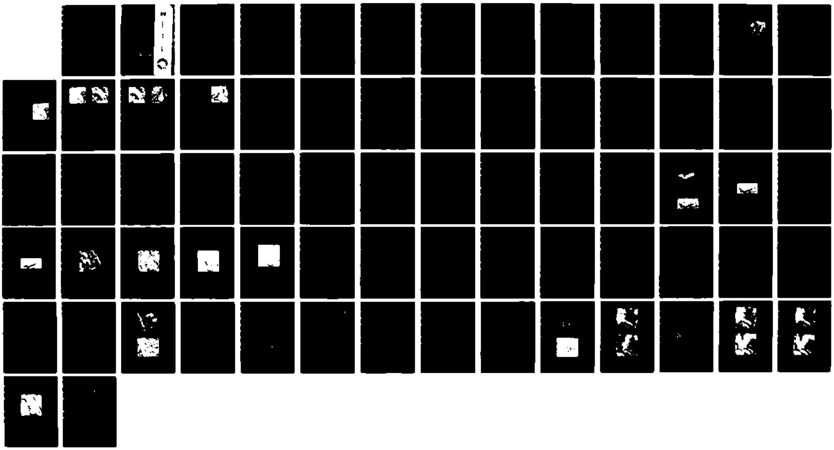
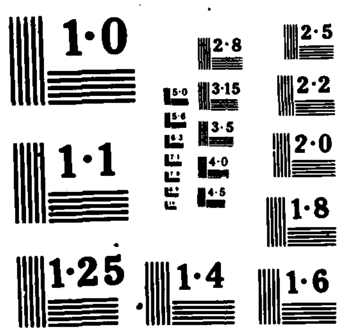
AD-A168 337 RESEARCH IN EXPERT INTERACTIVE CARTOGRPPHIC SYSTEMS(U) 1/1
SRI INTERNATIONAL MENLO PARK CA ARTIFICIAL INTELLIGENCE
CENTER A J HANSON 11 MAY 86 ETL-0417 DACA72-85-C-0008

UNCLASSIFIED F/G 9/4 NL




1·0
2·8
2·5
3·15
2·2
3·5
2·0
1·1
4·0
4·5
1·8
1·25
1·4
1·6

ETL-0417

AD-A168 337

Research in expert interactive cartographic systems

Andrew J. Hanson

SRI International
333 Ravenswood Avenue
Menlo Park, California 94025

May 1986

Prepared for
U.S. ARMY CORPS OF ENGINEERS
ENGINEER TOPOGRAPHIC LABORATORIES
FORT BELVOIR, VIRGINIA 22060-5546

ETL

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188
Exp. Date: Jun 30, 1986

1a. REPORT SECURITY CLASSIFICATION: Unclassified	1b. RESTRICTIVE MARKINGS: n/a
2a. SECURITY CLASSIFICATION AUTHORITY: n/a	3. DISTRIBUTION/AVAILABILITY OF REPORT: Approved for public release; distribution unlimited
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE: n/a	
4. PERFORMING ORGANIZATION REPORT NUMBER(S): 8645 Final Report	5. MONITORING ORGANIZATION REPORT NUMBER(S): ETL-0417
6a. NAME OF PERFORMING ORGANIZATION: SRI International Artificial Intelligence Center; 6b. OFFICE SYMBOL (If applicable):	7a. NAME OF MONITORING ORGANIZATION: U.S. Army Engineer Topographic Laboratories
6c. ADDRESS (City, State, and ZIP Code): 333 Ravenswood Avenue, Menlo Park, California 94025	7b. ADDRESS (City, State, and ZIP Code): U.S. Army Corps of Engineers, Fort Belvoir, Virginia 22060-5546
8a. NAME OF FUNDING/SPONSORING ORGANIZATION: U.S. Army, Engineer Topographic Labs; 8b. OFFICE SYMBOL (If applicable): EK239	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: DACA72-85-C-0008
8c. ADDRESS (City, State, and ZIP Code): U.S. Army Corps of Engineers, Fort Belvoir, Virginia 22060-5546	10. SOURCE OF FUNDING NUMBERS: PROGRAM ELEMENT NO.: ; PROJECT NO.: 8645; TASK NO.: ; WORK UNIT ACCESSION NO.:

11. TITLE (Include Security Classification)
Research in Expert Interactive Cartographic Systems (U)

12. PERSONAL AUTHOR(S)
Andrew J. Hanson

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Final	FROM 5/13/85 TO 5/13/86	May 11, 1986	67

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Spatial Reasoning, Cultural Object Detection, Image Segmentation

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

We have carried out a research program in the application of interactive environments to the development of knowledge-based methods for image understanding. As a sample domain, we chose to work mainly on the problem of locating generic cultural objects in aerial imagery. The discovery of such objects was accomplished by defining a generic model for rectilinear objects, along with rules for parsing the image geometry and correcting probable errors of the segmentation algorithm. These tools permit the semantic resegmentation of an initial syntactic scene partition to yield well-delineated buildings. The method owes its success to the combined utilization of both high-level and low-level knowledge about the target object context and the image.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT: ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION: Unclassified	
22a. NAME OF RESPONSIBLE INDIVIDUAL: Andrew J. Hanson	22b. TELEPHONE (Include Area Code): 415/859-4395	22c. OFFICE SYMBOL: MS EK239

DD FORM 1473, 84 MAR — 83 APR edition may be used until exhausted. All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED

Contents

PREFACE

This document was generated under contract DACA72-85-C-0008 for the U.S. Army Engineer Topographic Laboratories, Fort Belvoir, Virginia, by SRI International, Menlo Park, California, and was submitted as SRI Project 8645. The Contracting Officer's Representative was Mr. John Benton.

Research in Expert Interactive Cartographic Systems

By: Andrew J. Hanson, Project Leader

1 Introduction

The goal of this investigation was to carry out a program of basic research studying applications of interactive expert system techniques to the domain of automated cartography. We have ~~performed this work~~ in the context of the SRI Image Understanding Testbed, ~~which has been~~ extended to include Symbolics Lisp Machines and the SRI ImagCalc(TM) software support system.

As our sample domain, we chose the problem of locating rectilinear cultural objects in aerial imagery. This is an interesting research subject because each of the obvious object-location methods, edge-based or region-based, has significant problems.

A low-level image partition will *always* contain errors with respect to the task of object delineation, no matter how much the process is refined, because knowledge of the object model and context are missing. Algorithms based on edges alone, on the other hand, lack the strong constraints and context information provided by segmentation regions; furthermore, edge-based approaches have significant problems with sign-changes in the figure-ground relationship.

We therefore decided that the most effective approach to the object-delineation problem would be a knowledge-based architecture that used semantic knowledge about edge geometry to correct an initial segmentation.

During the course of the effort, we evolved through several stages while working to consolidate and broaden the rule base for discovering cultural structures of increasing complexity. Our latest approach utilizes generic models for multiply-branched rectilinear structures to parse cultural objects in the image and carry out a model-based resegmentation.

We have also developed substantial interactive capabilities within the context of the SRI ImagCalc system. A demonstration environment with moderate explanation capabilities and flexible interactive control procedures now supports the cultural-object delineation activity.

Finally, we have also been working on three-dimensional analysis techniques that will be incorporated into the general context of the object-delineation system. This will become especially significant when we need to deal quantitatively with such aspects as perspective distortions in the objects being extracted from the source imagery. The special stereographic display hardware acquired in the course of this project is essential for pursuing the three-dimensional analysis capability.

We feel that we have made significant progress towards our objective of investigating the mechanisms supporting rule-based solution of image understanding problems, and the means by which such problems can be solved by exploiting the cooperative strengths of a human operator and a knowledge-based computer system.

2 Summary of Technical Results

Our work on this project concentrated on the detection of building-like cultural objects in aerial imagery. This is both a useful domain in terms of potential practical applications and one that has clear geometric signatures that can be exploited. Furthermore, the accuracy of a result is easily checked for the purposes of evaluating the success of the paradigm.

Three different technical reports, characterizing the early, intermediate, and late stages of the research undertaken in the course of this project are provided in Appendices A, B, and C.

The major results that have been achieved are the following:

- **Model-based resegmentation:** Image segmentation is typically a syntactic process involving context-free operations on image intensity data. The resulting segmentations produce regions that do not correspond reliably to a given target because the specific high-level context of the target cannot be taken into account. By adding detailed knowledge about the context and by using specific models for the target object geometry, as well as knowledge about the probable failure modes of the segmentation itself, we are able to make intelligent corrections to the region shapes provided to us in the original segmentation. The result is a model-based resegmentation of the image that incorporates significant semantic knowledge about the object domain, and corresponds very closely to the regions containing target objects.

- **Identification of generic cultural objects:** Many techniques have been used to model specific objects and to discover them in an image. However, the most challenging problem arises when one knows only a general class, such as the class of rectilinear buildings, but nothing about the specific objects one might encounter. We have solved the problem of reliably finding instances of generic rectilinear buildings without any recourse to the use of rigid templates. Instead, we define a model for arbitrarily complex rectilinear structures of any size or shape and use this model to carry out the resegmentation process.

- **Parsing rules to extract model instances from real, noisy imagery:** There are many theoretically interesting approaches to model-based image interpretation. When applied to typical, noisy image data, however, most approaches require significant human intervention to parse the raw image data into the format required as input for the modeling systems. Our system includes carefully constructed parsing rules that take unprocessed image data and parse the information into the generic model representations needed for resegmentation. The whole approach derives its unique flexibility from the extensive use which is made of interaction between low-level and high-level knowledge.

The technical foundation that is now in place is expected to serve as the basis for further work. In particular, we plan to extend the explanation abilities of the interactive system and to develop rules for additional object domains, as well as incorporating some of the system components into an interactive, three-dimensional cartographic analysis, sketching, and display system.

Appendix A

Locating Cultural Regions in Aerial Imagery Using Geometric Cues

By: Pascal Fua and Andrew J. Hanson

Artificial Intelligence Center, SRI International

Published in Proceedings of the Image Understanding Workshop,
pp. 271–278 (December 1985).

LOCATING CULTURAL REGIONS IN AERIAL IMAGERY USING GEOMETRIC CUES

Pascal Fua and Andrew J. Hanson

Artificial Intelligence Center
SRI International, Menlo Park, California

ABSTRACT

To locate cultural regions in aerial imagery, we merge pixel-level techniques with geometric reasoning and generic (as opposed to specific or template-like) object descriptions. We utilize discrepancies between the generic models and the image data to refine an initial low-level segmentation and produce a more accurate delineation of cultural regions.

1 Introduction

Detecting and labeling scene objects is one of the more demanding tasks in automated image analysis. In the typical case of a high-altitude aerial image, there are no existing segmentation techniques that can reliably produce regions that have a one-to-one correspondence with objects of interest. Most segmentation procedures produce a wide mixture of undersegmented objects, where the object is merged with other data, and oversegmented objects, where the object is broken up into a "jigsaw puzzle" of indistinct parts. Furthermore, such segmentations are normally unstable with respect to minor changes in the program parameters, digitization methods, viewpoint, scene lighting, and film-processing methods.

We therefore propose to explore the application of knowledge-based methods to the problem of correcting an initial segmentation so it coincides with recognizable objects. Other related efforts include those of Ohta et al [1979], Nagao et al [1980], Reynolds et al [1984], Nazif and Levine [1984], McKeown et al [1984], and Hwang et al [1985]. Our work relies upon contextual geometric reasoning and generic, template-free models of the features to be extracted from the image. We overcome some of the limitations of previous approaches by providing powerful facilities for utilizing generic shapes and spatial context to resolve undersegmented objects.

The work reported here was supported by the Defense Advanced Research Projects Agency under Contract MDA903-83-C-0027 and by the U.S. Army Engineer Topographic Laboratories under Contract DACA72-85-C-0008.

For the purposes of our current work, we have imposed the following constraints:

- **Object type:** We restrict ourselves to the identification of cultural structures in aerial imagery, thereby providing the opportunity to use such observations as the presence of straight lines to focus attention on regions likely to be components of a target object [see, e.g., Shirai, 1978].
- **Image data:** We assume that we are given a digitized aerial image that is essentially a straight-down view, along with lighting and camera-model parameters. Typical images used in our experiments have scales of 1 to 2 feet per pixel on the ground.
- **Initial segmentation:** We assume we are provided with a syntactic partition of the image computed by an Ohlander-style segmenter [Ohlander et al, 1978; see also Laws, 1982, 1984].
- **Knowledge characteristics:** We assume that no precise templates of the target cultural objects are available, and thus we are required to deal with complex objects having only general, semantic descriptors.

Our results to date may be summarized as follows:

- **Undersegmented Regions Are Correctly Refined.** The identification of cultural portions of a region on the basis of groups of parallel and perpendicular lines leads to a very reliable splitting of undersegmented regions when combined with other contextual knowledge.
- **Templates Are Eliminated.** Many traditional systems for discovering buildings use relatively rigid rectangular templates, possibly with an allowable range of constraints on dimensions [e.g., Binford, 1982; Hwang et al, 1985]. Instead, we employ *generic* knowledge of the object geometry. By generalizing the concept of a "side" to include a large class of rectilinear zig-zag shapes and searching for rectangular geometric relationships among these compos-

ite shapes, we can accept and identify very complex polygonal structures with rectilinear components. No assumptions whatsoever are made about *specific* shapes, and thus we avoid the restrictions of the template approach while gaining substantial power.

- **Semantic Knowledge Supports Correction and Labeling of the Initial Segmentation.** We have linked domain knowledge with image-level operations in several ways to improve overall system behavior. We utilize knowledge of how the segmenter is likely to misplace region boundaries relative to desirable edges to recover such edges in the resegmentation, as well as to reject improbable geometries. Predicting the way shadows may be separated or incorrectly merged in the original segmentation leads to the correct parsing of shadow evidence required for identification of raised structures.

In the succeeding sections, we first describe our general approach to the design of an object-recognition system, and then present some initial results. We conclude with our plans for future refinement of the system.

2 Approach to the Object Recognition Problem

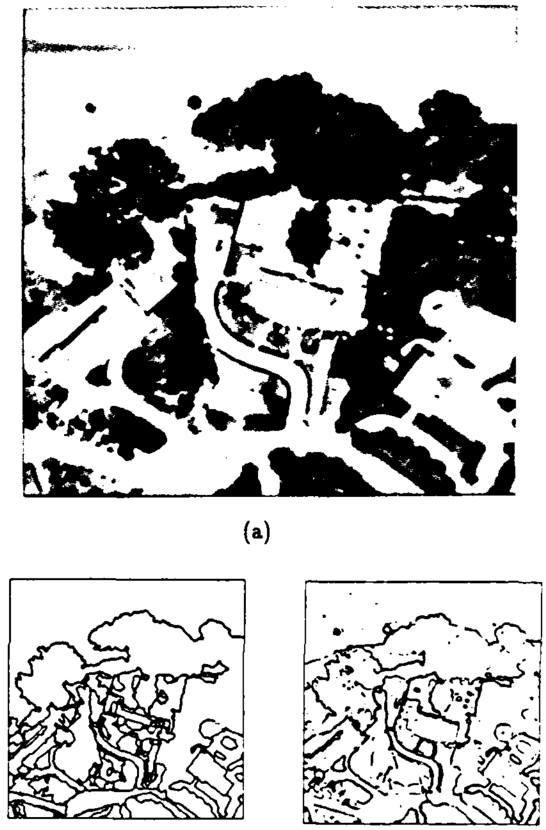
Several observations and theoretical concepts form the basis for our approach to the object recognition problem.

Recursive segmentation guarantees strong derivatives. An Ohlander-style segmentation of an image is recursive. A set of pixels in a given value range is selected on the basis of the shape of a frequency-of-occurrence histogram; these pixels are then labeled as belonging to one of several regions on the basis of spatial contiguity. The histogram of a region derived in this way will often have a shape entirely different from the parent histogram. The procedure is applied recursively until regions with no significant histogram structure are obtained.

Neighboring regions thus will often belong to *noncontiguous* value ranges of the histogram; the deeper the level of recursion, the more likely it is to find regions widely separated from their neighbors with respect to the range of pixel values in their histograms. *Region boundaries tend to lie on discontinuities in the pixel values and, therefore, strong derivatives occur between regions.*

In Figure 1, we verify these observations for a grey-scale image by showing the qualitative correspondence between segmentation region boundaries and the pixels in the image with high Sobel derivative strengths.

Sobel directions align with region boundaries. Edge direction can be determined in two ways. One is to fit a line to a set of points in an edge sequence, and the other is to compute the Sobel direction at a point. Because of the high correlation between Sobel derivatives and region boundaries shown in Figure 1, the latter will be quite reliable (see also Burns et al, 1984, for another approach).

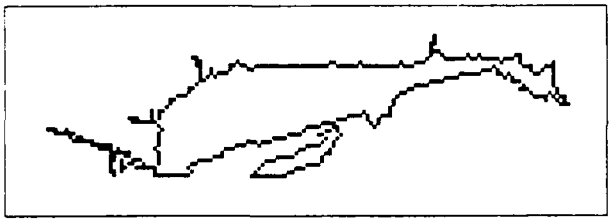


(a) (b) (c)

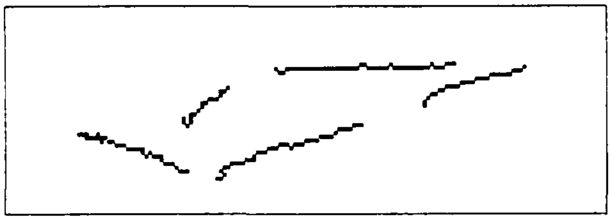
Figure 1: (a) An example of an aerial image containing houses. (b) The boundaries of the regions resulting from a segmentation of the image. (c) A binary image showing those pixels with strong magnitudes of the Sobel derivative.

In Figure 2, we show a typical region boundary obtained from the SRI SLICE segmenter [Laws, 1984], together with the long, straight lines obtained by an algorithm that looks only for consistency in the Sobel directions of a contiguous set of boundary points. The sets of points with compatible Sobel directions and the apparent linear boundary pieces are in good agreement.

Lines are classified by geometric direction. Semantically significant clusters of lines are often collinear, but *laterally displaced*. The direction that we assign to a cluster of two or more collinear or parallel lines is a



(a)



(b)

Figure 2: (a) Typical region boundary taken from the bottom center of Figure 1. (b) Long, straight lines in the region boundary derived only from requiring consistency of the Sobel directions in sets of contiguous points.

weighted average of the directions of each individual line, rather than the direction produced by fitting a line to the complete collection of points. This distinction is illustrated in Figure 3.

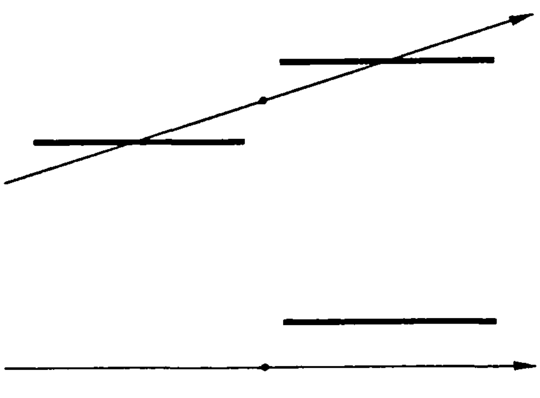


Figure 3: (a) The result of fitting a line to all the points in a pair of parallel, offset lines. The resulting direction is *incorrect* for the purposes of this work. (b) The composite direction of two lines computed from a weighted average of the direction of each line.

Shadows may be separated efficiently. Shadows form high-contrast regions with predictable geometric shape characteristics [see, e.g., Shafer, 1985; Medioni, 1983]. Our line-extraction methods are especially appropriate for extracting shadows that may have several broken segments aligned with the sun azimuthal angle.

Backtracking mechanisms are supported. Backtracking is accomplished in the current system using a library of reversible, rule-like procedures. An example of such a backtracking operation is shown in Figure 4; a composite line can be broken when a rule gives preference to the construction of a more complex structure, such as a U-shape.

We have previously expressed portions of our system in the framework of MRS [Genesereth et al, 1983] in an attempt to utilize the backtracking facilities provided in such a reasoning environment; in the current implementation, we have chosen for practical reasons to revert to procedural rule representation. Perhaps when a more complete understanding of this problem domain is achieved, we shall translate some of our procedurally represented rules into a more succinct declarative representation.

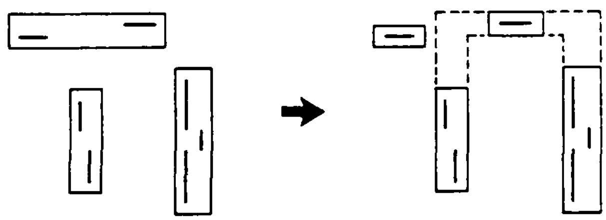


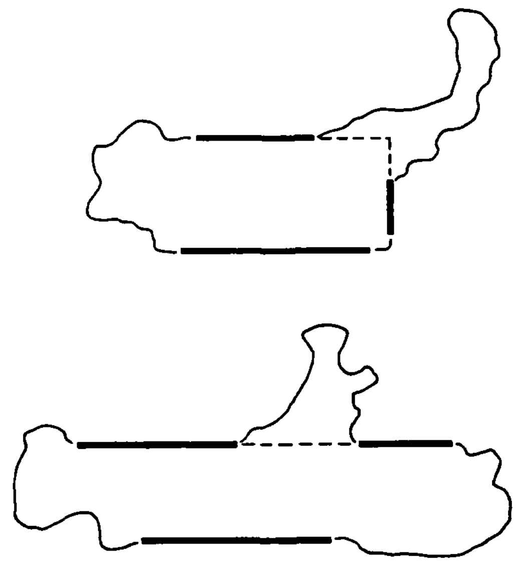
Figure 4: Backtracking by breaking a composite line to form a U-shaped structure. The U-shape is preferred because it provides strong evidence for a cultural object.

Geometric structure localizes semantically significant subregions. The current system relies upon general relationships such as perpendicularity and parallelness of composite line structures to single out portions of an arbitrarily shaped region that have suggestive polygonal substructures. This information is then used to correct the original segmentation.

We extract and use relationships such as *in front of, behind, between, beside, enclosed by, enclosing, at a certain angle from,* and *at a certain distance from* in both geometric and contextual reasoning processes. This vocabulary provides a basis for semantic reasoning, e.g., "look for dark areas in the direction of the solar azimuthal angle relative to a region boundary in order to confirm the hypothesis of a building wall."

Once interesting region portions are selected, a pixel-based line-linking procedure can be invoked to connect related lines, complete corners, and close open-ended **Parallels** or **U**'s. When the resulting links are satisfactory, the undesirable portions of the region are amputated, leaving clean cultural structures as the residue. Figure 5 illustrates linking processes that would be carried out when significant linear structures are present in an undersegmented region.

(a)



(b)

Figure 5: (a) Resegmenting a region with a good **U** by completing a corner. (b) Resegmenting a region with a good **Parallel** by linking the elements of a composite line.

3 Examples and Results

The current implementation of the system consists of two main sequences of operations:

- Discovering the geometric features and relationships within each single region.
- Resegmenting some regions based upon geometric relationships within a region or among distinct regions, and grouping interesting regions based on context knowledge.

Resegmentation is currently carried out using the F^* algorithm of Fischler et al [1981]. We compute the required cost array by using the Sobel edge strength combined with geometric constraints on the directions in which edge completion is predicted to take place. As a result, when the Sobel strength near a boundary segment follows a desirable path different from the boundary, F^* will pick up that path.

The final result of the computation is a resegmentation of the image with explicitly identified cultural-region clusters. Below, we present three examples illustrating the general features of the approach.

3.1 Example 1: An easy region.

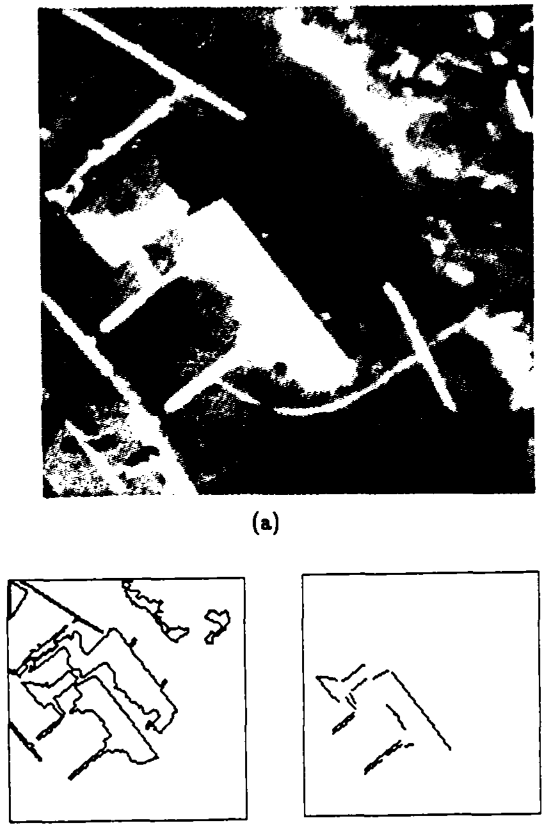
In the lower right-hand corner of the aerial image in Figure 1a there is a house whose outline corresponds exactly to one of the regions produced by the segmentation. The good lines found in the region boundary are shown in Figure 6. This house is characterized by the two sets of parallel lines that close to form a **Box**; an appropriately located shadow is also present.

Figure 6: The long, straight lines belonging to a distinct house region. These lines form a **Box** structure, indicating very strong evidence for a cultural object and distinguishing the region from its surroundings.

Even when the segmentation of an image is effectively perfect, locating the cultural correspondences can be non-trivial. Our method immediately focusses on this structure without *a priori* knowledge of its shape and singles it out because of its exceptional geometric structures.

In this case, no resegmentation is performed because there is no significant difference between the paths found by linking the lines and the region boundary itself. The result is a single, identifiable house-region, as shown in Figure 7.

Figure 7: The single identified house region boundary overlaid on the image. No resegmentation was necessary in this ideal case.



(a)

(b) (c)

Figure 8: (a) Another image containing a complex house structure. (b) Region boundaries. The upper L-shaped structure is a shadow; the lower L-shaped structure arises from two juxtaposed pieces of sunlit roof joined into a single region by a narrow neck. (c) Long, straight lines in the boundary of the sunlit roof region.

3.2 Example 2: Repartitioning a complex region

The next example, with the image, region boundaries, and elementary line segments shown in Figure 8, contains a heavily shadowed, approximately L-shaped, composite building. The segmentation confuses complex porches with roof tops, inappropriately combines sidewalks with the roof, and merges a significant shaded roof portion with background vegetation. The sunlit portions of the composite roof are contained in a single region; the two main lobes of this region are joined by a narrow neck. We observe that, given only the good edges of this roof region as shown in Figure 8c, the roof structure is confusing to parse even for a human.

We first search for basic geometric relationships within the roof-containing region. Two distinct **U**'s are found that support the identification of a cultural object, one in each lobe of the region. Both of these **U**'s require the breaking of a composite **T**, a type of backtracking, for their construction.

Next, the system attempts to link composite lines and to close the open ends of the **U**'s to form boxes using the line-linking algorithm. This procedure amputates the porch and sidewalk appendages and leaves two **Boxes**, outlined in Figure 9, that provide a clear semantic context. Applying knowledge of shadows here generates the hypothesis that both boxes are associated with the same large shadow region, so we label the group as a composite 3-dimensional structure.

We have thus succeeded in taking a single, confusing region and using its geometric structure to break it up into manageable parts. We note that the area enclosed between the pair of **Box** structures and the shadow is a heavily shaded, peaked-roof portion whose region features are so poor that it could not have been recognized by our basic methods; the labeled enclosing regions now provide the required semantic context to support this identification.

3.3 Example 3: Multiple region clustering

In Figure 10, we show another portion of the image of Figure 1a and its segmentation. This image is typical of cultural scenes that are difficult to parse using pattern-

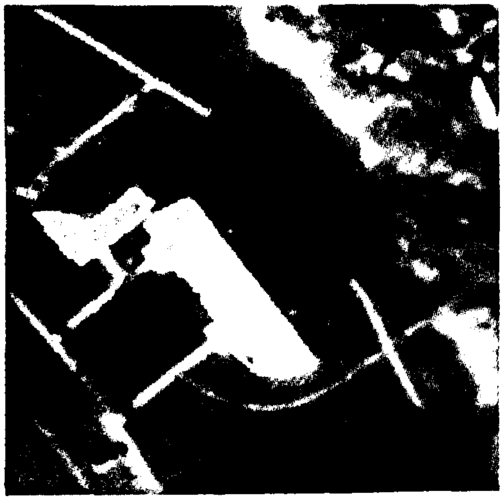


Figure 9: Final results of the splitting. The initial segmentation is split several times to give two subregions with good **Box** structures whose boundaries are outlined in the figure. The large shadow region is recognized as common to both subregions. The area between the shadow and the **Box** structures is now identifiable by its semantic context as a heavily shaded roof portion.

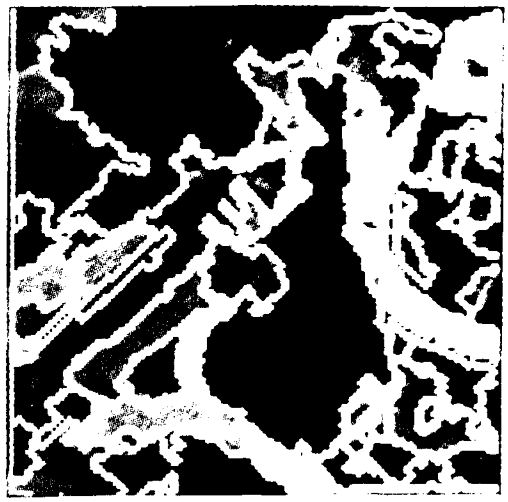


Figure 10: Left portion of the image of Figure 1a with segmentation boundaries from Figure 1b overlaid.

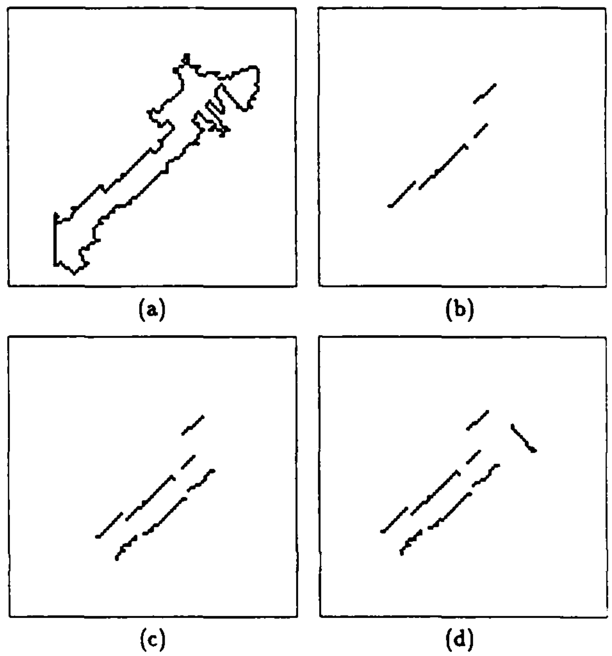
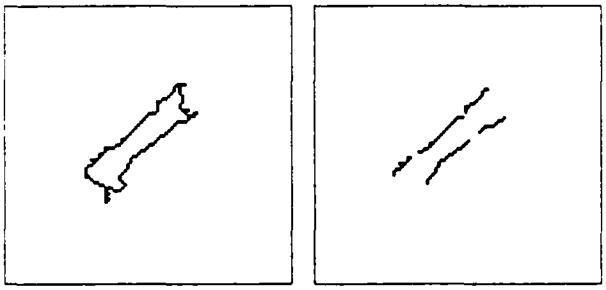


Figure 11: An illustration of the procedure by which geometrical relations are constructed within a single region. (a) Boundary of a typical region including portions of a house. (b) An example of a composite line with many elementary components extracted from the boundary. (c) A pair of parallel lines formed within the boundary by two composite lines. (d) The U structure constructed by finding a line in the boundary that closes off one end of the parallels. In this example, all good line segments belong to the U.

matching techniques because the terrain and roads are highly irregular and the houses have very complex shapes. Figures 11 and 12 show typical regions resulting from the segmentation of a house-containing area, along with illustrations of the process by which geometric structures are discovered. The first region contains an excellent **U**, while the second has a **Parallel**.

When we repeat the analysis for each region in Figure 10, we find only these two regions that have suggestive structures and appear to be geometrically related. Since an appropriate shadow region is present, we deduce that these regions probably belong to a single cultural cluster.

The geometric relations among lines in the boundaries of these regions are now used to predict the locations of the resegmentation boundaries to be constructed by linking. The results of the linking and resegmentation operations, depicted in Figure 13, show clearly the successful extraction of this complex building. We note that three different types of repartitioning were carried out to achieve this: (1) linking a corner formed by two lines belonging to a single region, thereby splitting off an irrelevant appendage; (2) linking a corner whose lines belong to two separate regions, thereby splitting yet a third region lying between them (this completes a U whose sides are the parallel lines in Figure 12); and (3) closing off the bottom of the U's formed by each of the two major roof segments.



(a) (b)

Figure 12: (a) Border of a second region belonging to the same house. (b) These parallel lines are the best structure that can be built. The Sobel directions of the short left edge are not sufficiently consistent to allow us to accept it as a closing line for a **U** structure.

The roof segments are labeled as belonging to a 3-dimensional, raised structure with a peaked roof, since they correspond to a "sunny side" and a "shady side" of the roof, with a narrow shadow adjacent to the "shady side."

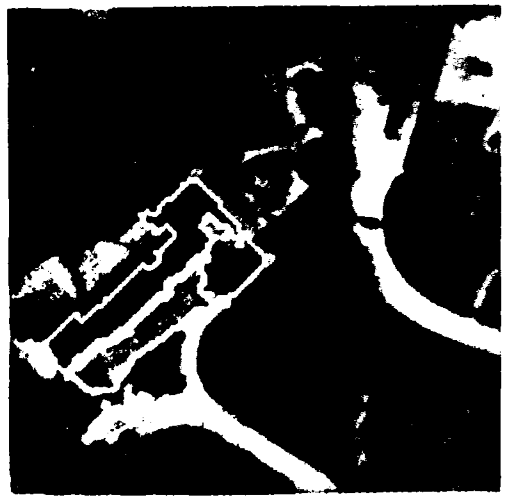


Figure 13: The results of computing linking lines and cutting regions accordingly. A third region comes into play when the linker completes the right-hand corner. The resulting three regions contain the area that one would visually associate with a house.

4 Directions for Future Work

We plan to add the following enhancements to the current system during the next stage of development:

- Generate interactive explanations of various actions to facilitate user understanding and debugging of domain rules; support user input of domain knowledge and corrections of the labeling.
- Merge "jigsaw puzzles" of objects that have been badly oversegmented.
- Extend the domains of expertise to include "explainable anomalies," of which the current shadow analysis is one example.
- Support additional classes of target objects.
- Incorporate additional geometric information such as perspective distortion of target shapes present in oblique views and nonplanarity of the underlying land.
- Support exploitation of multiple images covering the same scene.

The investigation described here explores a number of promising theoretical directions for knowledge-based partitioning and object identification, and produces satisfying experimental results for particular classes of images. Our next task will be to extend these ideas while incorporating support for explanatory interactions with the user.

References

T.O. Binford, "Survey of Model-Based Image Analysis Systems," The International Journal of Robotics Research, **1**, No. 1, pp. 18–64 (Spring, 1982).

J.B. Burns, A.R. Hanson, and E.M. Riseman, "Extracting Straight Lines," Proceedings of the Image Understanding Workshop, pp. 165–168 (1984).

M.A. Fischler, J.M. Tenenbaum, and H.C. Wolf, "Detection of Roads and Linear Structures in Low-Resolution Aerial Imagery Using a Multisource Knowledge Integration Technique," Computer Graphics and Image Processing **15**, pp. 201–223 (1981).

M. Geneserith, R. Greiner, and D.E. Smith, "MRS — A Meta-Level Reasoning System," Stanford University Heuristic Programming Project Report No. HPP-83-27 (1983).

V. Hwang, L. Davis, and T. Matsuyama, "Hypothesis Integration in Image Understanding Systems," University of Maryland Report CAR-TR-130 (1985).

K.I. Laws, *The PHOENIX Image Segmentation System: Description and Evaluation*, Technical Note 289, Artificial Intelligence Center, SRI International, Menlo Park, California (September 1982).

K.I. Laws, *Goal-Directed Texture Segmentation*, Technical Note 334, Artificial Intelligence Center, SRI International, Menlo Park, California (September 1984).

D. McKeown, W.A. Harvey, and J. McDermott, "Rule Based Interpretation of Aerial Imagery," Proceedings of IEEE Workshop on Principles of Knowledge-Based Systems, pp. 145–157 (1984). See also IEEE Trans. PAMI, *in press.*

G.G. Medioni, "Obtaining 3D Information from Shadows in Aerial Images," Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 73–76 (1983).

M. Nagao and T. Matsuyama, *A Structural Analysis of Complex Aerial Photographs*, Plenum Press (New York, 1980).

A.M. Nazif and M.D. Levine, "Low Level Image Segmentation: An Expert System," IEEE Trans. PAMI **6**, pp. 555–577 (1984).

R. Ohlander, K. Price, and D.R. Reddy, "Picture Segmentation Using a Recursive Region Splitting Method," Computer Graphics and Image Processing **8**, pp. 313–333 (1978).

Y. Ohta, T. Kanade, and T. Sakai, "A Production System for Region Analysis," Proc. 6th Inter. Joint Conf. on Artif. Intell., pp. 684–686 (1979).

G. Reynolds, N. Irwin, A.R. Hanson, and E.M. Riseman, "Hierarchical Knowledge-Directed Object Extraction Using a Combined Region and Line Representation," Proceedings of the Image Understanding Workshop, pp. 195–204 (1984).

S.A. Shafer, *Shadows and Silhouettes in Computer Vision*, Kluwer Press (1985).

Y. Shirai, "Recognition of Man-Made Objects Using Edge Cues," in *Computer Vision Systems*, Ed. A.R. Hanson and E.M. Riseman, Academic Press (New York, 1978).

Appendix B

Using Generic Geometric Knowledge to Delineate Cultural Objects in Aerial Imagery

By: **Pascal Fua** and **Andrew J. Hanson**

Artificial Intelligence Center, SRI International

SRI Technical Note 378
April, 1985

Using Generic Geometric Knowledge to Delineate Cultural Objects in Aerial Imagery

ABSTRACT

We present a paradigm for discovering the outlines of arbitrarily complex cultural objects in aerial imagery. The approach starts with a low-level image partition and and generic (as opposed to specific or template-like) object descriptions. We then use geometric reasoning and context knowledge to suggest corrections to the discrepancies between the segmentation boundaries and the object models. Finally, when the corrections appear consistent with the generic cultural object model, we resegment the partition to produce new labeled regions with clear semantic interpretations. The general features of our approach appear to be applicable to a number of other domains, including the delineation of vegetation areas.

1 Introduction

We describe a knowledge-based approach to the construction and labeling of regions corresponding to cultural objects in aerial imagery. Such a paradigm is necessary because typical low-level scene segmentation techniques cannot reliably generate regions that have unambiguous correspondences with object labels. The regions produced by a syntactic image segmentation method are typically either undersegmented, with cultural objects merged into background features, oversegmented, with semantically distinct objects broken into many confusing pieces, or both.

A low-level image partition will *always* contain errors with respect to the task of object delineation, no matter how much the process is refined. Algorithms based on edges alone, on the other hand, lack the strong constraints and context information provided by segmentation regions. We therefore suggest that the most effective approach to the object delineation problem is a knowledge-based architecture that uses semantic knowledge about edge geometry to correct an initial segmentation.

The current work concentrates on the detection of building-like cultural objects in aerial imagery. This is both a useful domain in terms of potential practical applications, and one that has clear geometric signatures that can be exploited [see, e.g., Shirai, 1978]. Furthermore, the accuracy of a result is easily checked for the purposes of evaluating the success of the paradigm.

Among the previous efforts relevant to our approach, we note the work of Tavakoli [1980] and Hwang et al [1985], which incorporates primitive concepts of generic shapes; Binford [1982], which surveys model-based object recognition methods; Burns et al [1984], and Reynolds et al [1984], which employs innovative edge segmentation techniques; McKeown et al [1985], which utilizes knowledge-based region-growing and sophisticated geometrical context knowledge; Shafer [1985] and Medioni [1983], which studies evidence available from shadows; Nazif and Levine [1984], which attempts a conventional production-rule approach to low-level segmentation; Nagao et al [1980] and Ohta et al [1979], which gives ambitious approaches to the region-labeling problem; and Nevatia and Huertas [1985], which explores geometric primitives similar to ours and makes extensive use of shadows.

Improved performance in difficult and ambiguous scenes has been attained in the current work because of the following features of our approach:

- Introduction of a significant generalization of the notion of a rectangular structure to support the concept of a *generic* cultural object model.
- Support for models of composite objects having arbitrary intensity characteristics relative to the background.
- Choosing corrective strategies based on explicit knowledge about the behavior of the segmentation process.
- Exploitation of knowledge about the interaction of edges and the segmentation regions to which they belong.
- Incorporation of rules and goal-directed edge-finding procedures that handle the splitting of regions containing undersegmented objects.
- Incorporation of rules that support the knowledge-driven grouping of oversegmented object parts.

The next section gives an overview of our system design philosophy. We then discuss the rules and geometric reasoning methods that underlie the approach. Finally, we show the results that we obtain on a complex cultural scene.

2 System Design

We have found that simple edge-parsing methods are too ambiguous to be generally effective for our work. We therefore provide a strong initial context for edge-based geometric reasoning by choosing an Ohlander-style segmentation as the starting point of our system design [see Ohlander et al, 1978, as well as Laws, 1982, 1984]. The main characteristic of such a segmentation is that it groups together contiguous pixels belonging to a particular intensity range in a histogram that has been derived from recursive splitting of histograms of parent regions. As a result, region boundaries tend to lie on contours with high intensity derivatives; it is thus appropriate to use simple operators such as the Sobel derivative to study the characteristics of Ohlander-style region boundaries.

We have made no special effort to tune the segmentation parameters to our application in the images we have studied; our objective is to prove that, in the presence of the inevitable errors produced by segmentation processes, knowledge and geometric reasoning can be used effectively to overcome the segmentation anomalies and produce meaningful object delineations.

A significant characteristic of edges belonging to region boundaries is that they may be assigned a topological direction that provides additional consistency constraints on edge combination processes. Such constraints continue to be useful even for edges belonging to distinct neighboring regions or islands (interior boundaries assigned to large regions that completely enclose a smaller region).

One of the unique properties of our design is the use of composite edge structures to compensate for the fact that semantically meaningful straight lines bordering cultural objects tend to be zigzagged as well as broken up by photometric anomalies. Even more critical for the achievement of building recognition is the fact that, when a building "side" is allowed to be one of our composite edge structures, a "box" built of four such mutually-perpendicular structures can in principle correspond to any object composed of adjoined rectangles. Thus, what our rule system treats as a "box" semantically encompasses objects that are perceived as boxes, L's, T's, crosses, U's, zigzags, and so on.

Our basic system architecture for identifying and labeling objects in a scene using knowledge-based resegmentation is the following:

- **Compute Single-Region Structures.** Given a segmentation and the values of the Sobel derivative, we first accumulate atomic edges composed of adjacent region-boundary pixels that satisfy particular semantic criteria for the problem at hand. To identify buildings, we use a straight line extractor.

  Next, we collect together sets of atomic edge elements belonging to a single region to form composite edges. For buildings, we choose sets of straight atomic edges that share a geometric direction; the weighted average direction of the straight edges is the direction of the composite.

  Finally, we construct semantically-meaningful geometric structures. Generic models for object features are used to produce geometric structures that characterize the presence of a cultural object. Typically, there is a hierarchy of such geometric evidence, with the different levels giving increasing confidence that an object is indeed present. Boxes and U's built of composite edges give strong generic supporting evidence for the presence of buildings. These structures work equally well in the context of multiple regions and islands, except that additional semantic constraints are usually required to replace the strong intrinsic constraints present in the single-region context.

- **Group Structures Across Regions.** Cultural objects are typically broken up in predictable ways by the segmentation process. Thus, we must check for evidence of such fragmentation and attempt to verify the existence of reasonable links among structures that might have arisen from a single object. The system checks for common edges in structures belonging to adjacent regions, and groups the structures together if they pass various consistency tests. In this way, multiple region information provides support for composite structures that would be neglected if we restricted ourselves to the single-region domain.

- **Use Model-Driven Prediction to Correct the Segmentation.** Comparing the geometric structures with their underlying models in the context of the segmentation now provides predictions about the probable locations of missing structure segments. These are fed into an edge-finding procedure, and the resulting new boundaries remove extraneous structures from undersegmented regions. Conversely, knowledge of the object model permits regions belonging to an object that has been broken up by the segmentation to be grouped into a more meaningful composite structure. Among the methods that might be used to test hypotheses about correcting the segmentation in order to better match the object models we note:

  - path finders such as F* [Fischler et al, 1981]; this is the method utilized in the current system to determine the probable location of missing segmentation boundaries.
  - region growers [e.g., McKeown et al., 1985].
  - path predictors and extrapolators, such as would be required to deal with occlusion.
  - reiterating the original segmentation process (or another selected for its special properties) over the region or a particular subregion that is known to be of interest. In this case, scoring functions evaluating any of several levels of semantic content could be used to make segmentation iterations effectively "goal-directed."

  Finally, when all meaningful clustering and partitioning has been carried out, we attach semantic labels that could be used by abstract, image-independent query processes.

Each step of the processes described above makes use of our system's library of general geometric reasoning tools. In our experience, new bodies of semantic information can be easily added to the system by developing procedural rules based upon the power and flexibility of these fundamental tools.

3 Rules for Geometric Reasoning about Cultural Structures

3.1 General Issues

The first step in constructing a system to reason about generic cultural structures in aerial imagery is the introduction of a spatial vocabulary. The next step is to accumulate knowledge and heuristics derived from a wide variety of experiments and empirical observations and use that information to construct viable rules.

We list below some of the observed geometric features that characterize buildings, and thereby influence the form of the rules we use:

- Cultural objects such as buildings are characterized at the lowest level by straight edges. However, region edges are often ambiguous, broken by photometric anomalies, and zig-zagged due to the existence of multiple structural parts.

- In order to accommodate edge ambiguities, we construct *composite edges*. These edges are the key to making the shape model more truly generic. Semantically significant clusters of edges are often collinear, but *laterally displaced.* The direction that we assign to a cluster of two or more collinear or parallel edges is a weighted average of the directions of each individual edge, rather than the direction produced by fitting a line to the complete collection of points. We illustrate the construction in Figure 1.

- Complex cultural objects are formed from many adjoined rectangular sections, so looking for simple rectangles and L-shapes will not be sufficient. Generalized rectangles made from *composite edges*, however, can describe any shape in this generic category.

The basic vocabulary of geometric entities relevant to building extraction, ranked in order of precedence for the purposes of backtracking and redefining a structure, are:

- atomic edge – a statistically-determined contiguous set of pixels making a straight line in a region boundary.

- composite edge – a set of atomic edges with mutually consistent directions, along with a composite direction derived from the directions of the edges, not from the union of the set of edge points.

- corner, T-corner – two perpendicular composite edges; an ordinary corner has the two closest ends arranged so that their head-to-tail directions in the region boundary agree, and so that neither intersects the other (with some tolerance) when extrapolated; T-corners have a significant intersection upon extrapolation.

- parallel – two parallel composite edges.

- U – a parallel structure each of whose elements form a corner or a T-corner with the same end element.

- box – a structure built from two perpendicular sets of parallel structures.

In our system as it is currently implemented, rules are procedurally encoded in a set of 50 or 60 functions. The basic structure of each function is

$$\textbf{IF}\ \boxed{\textit{Pattern Match}}$$
$$\textbf{THEN}\ \boxed{\textit{Operate on Data Structure}}.$$

The pattern-matching procedure is typically so complex that it has proven much easier to obtain reasonable performance and control using procedurally-encoded rules rather than declarative rules. The data structures that are manipulated by a rule consist mainly of the trees of associations that build semantically meaningful statements from atomic edges.

We have followed a customary "expert system development" philosophy to evolve the capabilities of the software. There is a basic set of rules and capabilities that are fully automated, plus appropriate junctures at which the operator can be asked to supply a judgement currently beyond the capabilities of the automated rule base. By noting such judgements and their semantic explanations, we acquire the information required to add corresponding rules to the fully automated system.

3.2 Rule Examples

We now present several examples of the rules and reasoning processes that must be carried out for our application — the discovery of building outlines.

Avoiding a Composite Edge. One simple example of a rule is illustrated in Figure 2. The knowledge upon which the rule is based is the fact that regions whose boundaries "double back" on themselves almost inevitably behave that way because a piece of yard or sidewalk adjacent to a building has been included in the segmentation, but semantically is an appendage to the region representing the building sought. Thus, if two line segments appear to overlap, they should not be joined into a composite edge.

Motivating a Composite Edge Using a Neighboring Parallel. Next, we look at a typical rule involved in the construction of parallels. In Figure 3, we show the case where the three edges of Figure 2 have a common parallel edge in the same region. Using the knowledge that spatial proximity of the two parallel elements may be used to recognize the existence of the unwanted region appendage, probably resulting from a yard or sidewalk, the procedure eliminates the more distant parallel, assuming it is an appendage, and merges the two nearer edges into a single composite line to complete the parallel structure.

Making a Better Structure by Breaking a Composite Edge. An existing composite edge should be broken when doing so results in the successful construction of a more complex structure, such as a U-shape. In Figure 4, we illustrate such an action in the case of a region whose interpretation is that of a building segment merged with an adjacent irrelevant structure. By breaking off the extraneous structure, we recover a **U** that is more consistent with the geometric expectations of a structure belonging to a building.

Resegmenting by Prediction of Border Completion. Another form of rule involves recognizing where a missing segment of a geometric structure should lie, and feeding the predicted location to a likelihood-based edge finder. In Figure 5, we show how such a process would rediscover a weak edge missed in the original segmentation. The same basic rule works both for structures in a single region and for structures whose elements are spread across multiple regions or island regions, as illustrated in Figure 6. The tight constraints available in the single-region case must of course be supplemented in the multiple-region case by knowledge of probable scales and domain-dependent features.

Completing a U in an Associated Region. In Figure 7, we illustrate a multiple-region splitting rule. The parallel at the bottom may suffer from noisy edges that prevent the component lines from extending to the true end of the building; the upper **U** structure provides an improved context for predicting the path to be used to close one end of the lower parallel.

Grouping Using Sun Angle. In Figure 8, we illustrate the process that checks for regions on the shady side of atomic edges comprising a good high-level structure such as a **U** or a **Box**. Once a good structure belonging to the sunny portion of the roof is recognized, an hypothesis for the location of the shaded roof portion and the shadow itself is formed and tested. Then the structures belonging to the tentative shaded roof are examined, and other applicable rules invoked to close off relevant structures to make good boxes delineating the roof portions. An important feature of the shaded roof location process is the fact that only regions on the shady side of edges belonging to structures with strong cultural indications are examined. One should not examine *all* of the region border, since irrelevant sidewalk appendages would find darker grassy regions on their shady side, and so forth.

4 Using Generic Models to Discover Buildings

In this section, we illustrate both the general power of the paradigm presented in Section 2, and the effectiveness of the particular set of rules that are used within this context to discover and label buildings.

This work is currently in progress, with significant additions still being made to the rule base. We have therefore chosen illustrations that reflect a combination of totally automated rule structures such as those illustrated above in Section 3 with interactively-guided heuristic choices. The use of human interaction is in fact an essential step in acquiring the knowledge necessary to build such a system – by making judgements and choices that are quickly reflected in the resulting segmentation, the human user develops the intuitive knowledge necessary to state and encode rules that embody general principles of the problem.

Virtually all of the interactively-guided choices made in the examples presented here will be translated into automated rule invocations in the near future.

4.1 Example: The Structure of a Single Building

Our first example is an image containing a single, complex building shown in Figure 9. It contains a heavily shadowed, approximately L-shaped, composite building. The segmentation shown in Figure 10 mixes roofs and sidewalks, and has a large, confused region that contains both vegetation and shaded roof portions. Figure 11 shows the atomic edges extracted from the boundaries of the image partition, and Figure 12 shows the significant geometric structures that are built from the edges.

The system next invokes a set of rules that take the observed geometric structures and search for neighboring regions that are semantically consistent with the identification "building with sunny roof plus shady roof." The structure-completion rules then run the edge-finder and complete the delineation of the sunny and shady roof portions shown in Figure 13.

4.2 Example: A Cluster of Buildings

We now let the system run on a large image, shown in Figure 14, which contains a cluster of buildings. Examining the initial segmentation boundaries shown in Figure 15, we note a large region that is virtually unsegmentable, with shaded rooftops, grass, roads, and other vegetation indiscriminately merged into the region. Thus one needs semantic knowledge to distinguish relevant structures within this region.

In an image such as this with low sun elevation, several very simple criteria such as intensity, size, and the existence of edge structures parallel to the sun azimuth serve to identify uniquely the shadow-like regions shown in Figure 16. For the three buildings with sunlit roofs in the central part of the image, shadow information is superfluous due to the existence of strong geometric evidence. However, the shadow information may be used to predict the presence of the other, noisier, buildings. Alternatively, a procedure may be invoked to generate hypotheses about the locations of other sunlit roof regions by comparing the intensity signature of the clean sunlit roofs to other unlabeled regions.

Using the shadow identifications and probable directions of shaded roofs relative to sunlit roofs and shadows, we apply our usual rules to construct and resegment the building-like groups shown in Figure 17.

5 Conclusions and Remarks

We have described a framework for a knowledge-based system to delineate and label objects in an image when supplied with a reasonable but highly erroneous partition. Choosing as an example the domain of cultural structures in aerial imagery with shapes corresponding to generalized rectangles, we have derived and tested a series of rules that successfully implement the proposed framework.

Given our fundamental model for carrying out geometric reasoning about the features of cultural objects within the context of a low-level image partition, we have found it straightforward to extend the hierarchy of knowledge to include the implications of higher-level concepts such as shadows, peaked roofs, and backyards. While considerable effort may be involved in developing the necessary additional rule bases, we believe that this approach can be applied to at least the following domains:

- **Raised rectangular cultural objects.** This includes primarily buildings of the kind the current system already handles successfully.

- **Circular cultural objects.** Various kinds of storage structures have circular shapes. To account for possible obliqueness of the camera angle, such a system would need to deal with ellipses as well as circles.

- **Linear cultural structures.** This category includes roads, sidewalks, and parking lots.

- **Natural linear structures.** Streams, rivers, canyons, dry guileys, and eroded areas should be recognizable by the *non*-cultural signature of their region edges.

- **Natural irregular objects.** Vegetation, individual trees, and forest boundaries should be recognizable also by the irregular signature of the edges of their regions. Preliminary work with characteristics of vegetation boundaries indicates that requiring either good fractal measures or large variances in edge directions (indicating chronic crookedness) are extremely effective in ranking scene regions according to the amount of vegetation in the region boundaries. Replacing straightness of edges in the house-delineation paradigm by fractal crookedness of edges and appropriately readjusting the rest of the resegmentation algorithm appears to produce reasonable vegetation regions.

We hope in future work to extend the basic object delineation approach we have presented here and to develop a broad, knowledge-based scene segmentation and labeling tool. We would like to develop rule bases for a selection of the domains noted above, and to install a general interactive architecture and explanation system to support the existence of such multiple contexts. The output of such a system would then provide a firm basis upon which to build much more abstract intelligent systems, such as planners, that need detailed symbolic knowledge extracted from imagery before they can function.

References

T.O. Binford, "Survey of Model-Based Image Analysis Systems," The International Journal of Robotics Research, **1**, No. 1, pp. 18–64 (Spring, 1982).

J.B. Burns, A.R. Hanson, and E.M. Riseman, "Extracting Straight Lines," Proceedings of the Image Understanding Workshop, pp. 165–168 (1984).

M.A. Fischler, J.M. Tenenbaum, and H.C. Wolf, "Detection of Roads and Linear Structures in Low-Resolution Aerial Imagery Using a Multisource Knowledge Integration Technique," Computer Graphics and Image Processing **15**, pp. 201–223 (1981).

V. Hwang, L. Davis, and T. Matsuyama, "Hypothesis Integration in Image Understanding Systems," University of Maryland Report CAR-TR-130 (1985).

K.I. Laws, *The PHOENIX Image Segmentation System: Description and Evaluation,* Technical Note 289, Artificial Intelligence Center, SRI International, Menlo Park, California (September 1982).

K.I. Laws, *Goal-Directed Texture Segmentation,* Technical Note 334, Artificial Intelligence Center, SRI International, Menlo Park, California (September 1984).

D. McKeown, W.A. Harvey, and J. McDermott, "Rule Based Interpretation of Aerial Imagery," Proceedings of IEEE Workshop on Principles of Knowledge-Based Systems, pp. 145–157 (1984). See also IEEE Trans. PAMI, *in press.*

G.G. Medioni, "Obtaining 3D Information from Shadows in Aerial Images," Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 73–76 (1983).

M. Nagao and T. Matsuyama, *A Structural Analysis of Complex Aerial Photographs,* Plenum Press (New York, 1980).

A.M. Nazif and M.D. Levine, "Low Level Image Segmentation: An Expert System," IEEE Trans. PAMI **6**, pp. 555–577 (1984).

R. Nevatia and A. Huertas, "Building Detection in Simple Scenes," Univ. of Southern California Intell. Systems Group Technical Report, (September, 1985).

R. Ohlander, K. Price, and D.R. Reddy, "Picture Segmentation Using a Recursive Region Splitting Method," Computer Graphics and Image Processing **8**, pp. 313–333 (1978).

Y. Ohta, T. Kanade, and T. Sakai, "A Production System for Region Analysis," Proc. 6th Inter. Joint Conf. on Artif. Intell., pp. 684–686 (1979).

G. Reynolds, N. Irwin, A.R. Hanson, and E.M. Riseman, "Hierarchical Knowledge-Directed Object Extraction Using a Combined Region and Line Representation," Proceedings of the Image Understanding Workshop, pp. 195–204 (1984).

S.A. Shafer, *Shadows and Silhouettes in Computer Vision,* Kluwer Press (1985).

Y. Shirai, "Recognition of Man-Made Objects Using Edge Cues," in *Computer Vision Systems*, Ed. A.R. Hanson and E.M. Riseman, Academic Press (New York, 1978).

M. Tavakoli, "Toward the Recognition of Cultural Features," Proceedings of the Image Understanding Workshop, pp. 33–57 (1980).

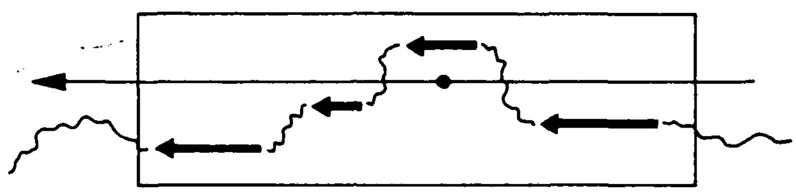


Figure 1: Each thick arrow represents one of a set of straight edge segments lying in a region boundary. This set of atomic edges forms a composite edge for geometric reasoning purposes. The long arrow denotes the semantically correct direction of the composite edge, computed from a weighted average of the directions of each atomic edge.

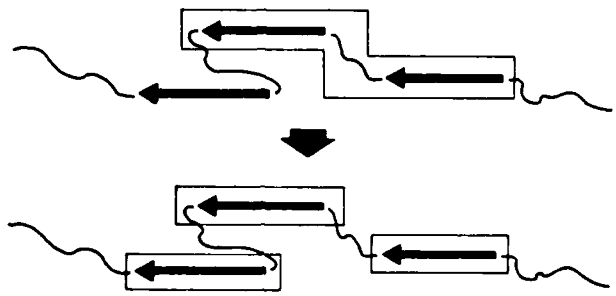


Figure 2: In the first stage of composite edge accumulation, the two contiguous edges enclosed in the box at the top are associated. However, a second stage checks the consistency of the geometry and discovers that the next edge in this region boundary lies to the right of the leftmost end of the tentative composite line. This is the signal to dissociate these atomic edges from the composite structure, as shown at the bottom.

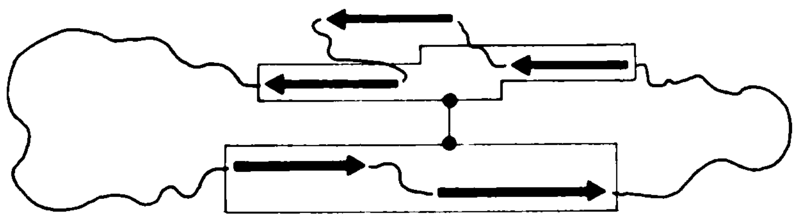


Figure 3: Here there are three short edges that might be logically linked with the bottom long edge, except that two short edges overlap because one belongs to an appendage. Using the knowledge that such an appendage is probably due to a neighboring part of a yard or patio, rather than the building itself, we choose to merge *only* the closest short edge into the composite line, forming the final parallel structure shown.

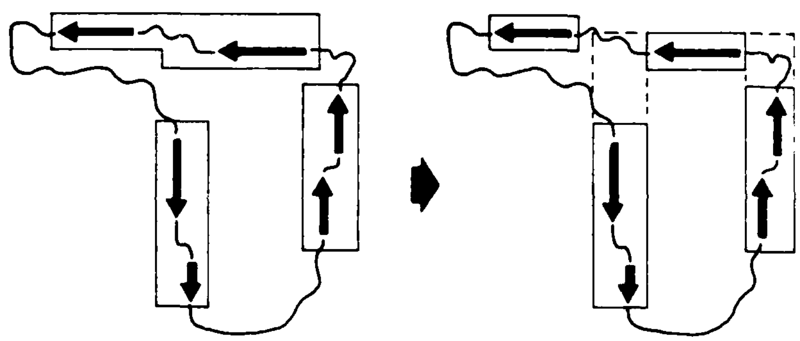


Figure 4: Backtracking by breaking a composite line to form a U-shaped structure. The U-shape is preferred because it provides strong evidence for a cultural object.

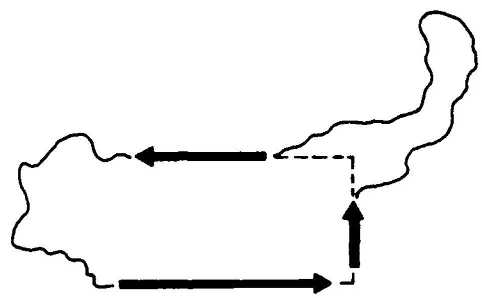


Figure 5: The existence of a good **U** structure here serves to predict that the missing portions of the corner should be constructed if possible. If the line finder successfully finds a good path in the predicted geometric vicinity, the erroneous appendage is removed and the is region split in two along the resulting linking path.

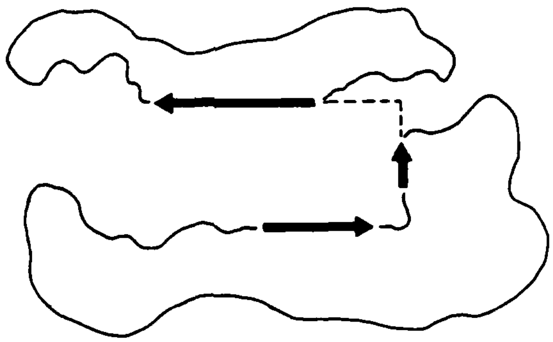


Figure 6: One may use the same geometric rules as for single regions when dealing with multiple interior boundaries of regions with holes because the orientation of edges in these "island" regions is reversed. In the case shown here, two neighboring island regions have edges that can be combined to form a **U**, and the enclosed region is resegmented along the predicted path to close off the **U**.

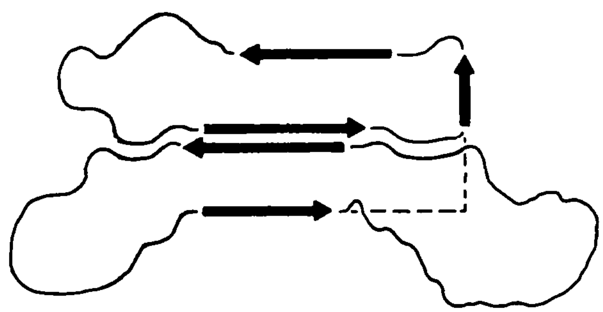


Figure 7: The upper U closure determines the path predicted for a meaningful closure of the lower parallel, both of whose ends are open.

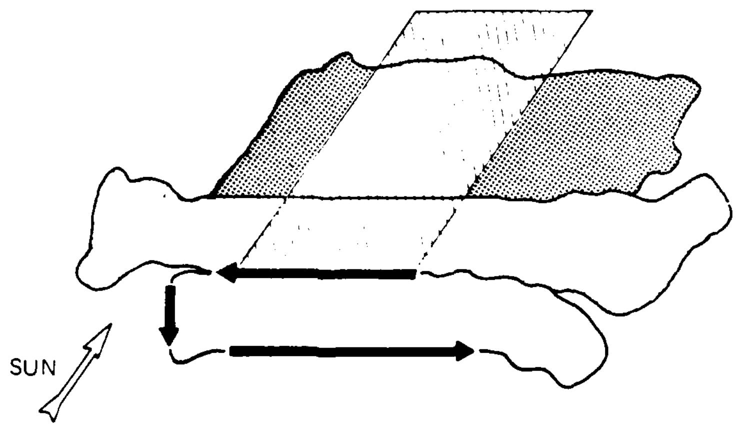


Figure 8: A sunlit roof portion with a U structure. The edge elements on the shaded side of the structure are used to look for regions that might be the shaded portion of a peaked roof.

Figure 9: Image of complex building, showing shaded roofs, shadows, sidewalks, and roads.

Figure 10: Initial segmentation of the building-containing image.

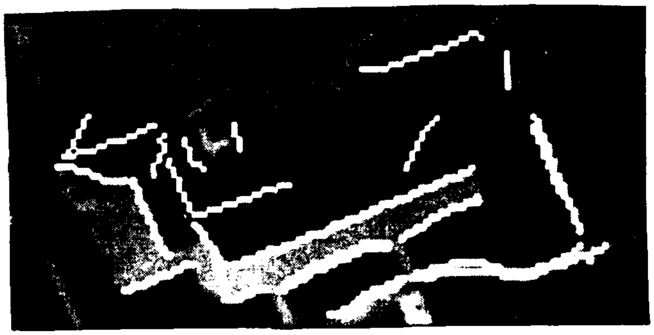


Figure 11: The straight edges used to produce the geometric structures characteristic of the cultural object.

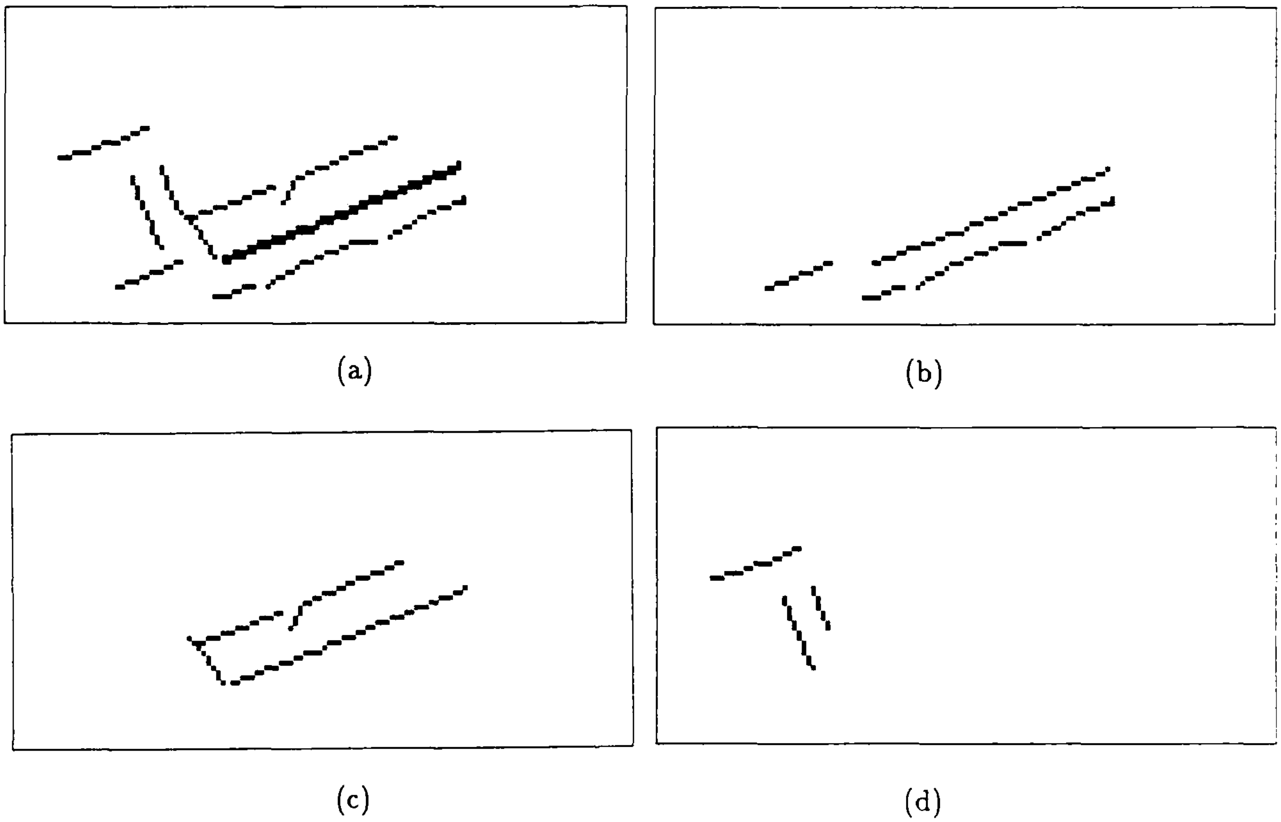


Figure 12: The geometric structures used to parse the regions belonging to the building. (a) All the edges belonging to structures. (b) A parallel belonging to the lower right sunny roof. (c) A **U** belonging to the upper right shady roof. (d) A **U** belonging to the upper left shady roof. Each of these structures can be used to predict where missing pieces of the object boundary should fall.

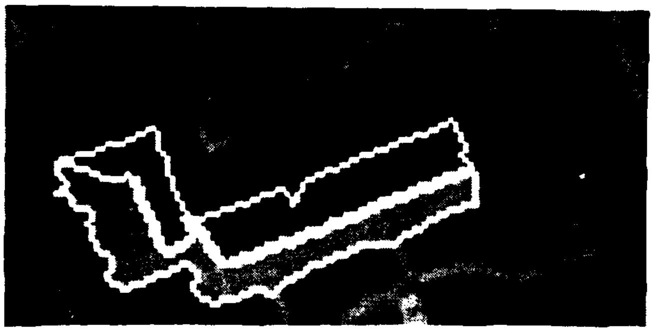


Figure 13: Final results of splitting the regions and closing off the cultural structures. Structures such as narrow sidewalks are split off to produce a cluster of regions corresponding precisely to a building with sunny and shady sides of the roof.

Figure 14: A large image containing the previous example as a subimage.

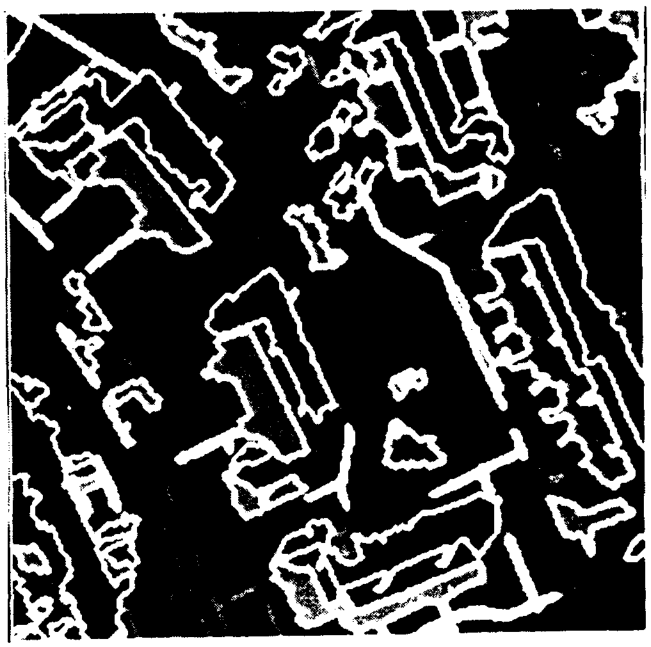


Figure 15: The segmentation boundaries of the large image.

Figure 16: Shadow region boundaries extracted from the large region by applying simple criteria based on alignment with the sun, intensity, and size.

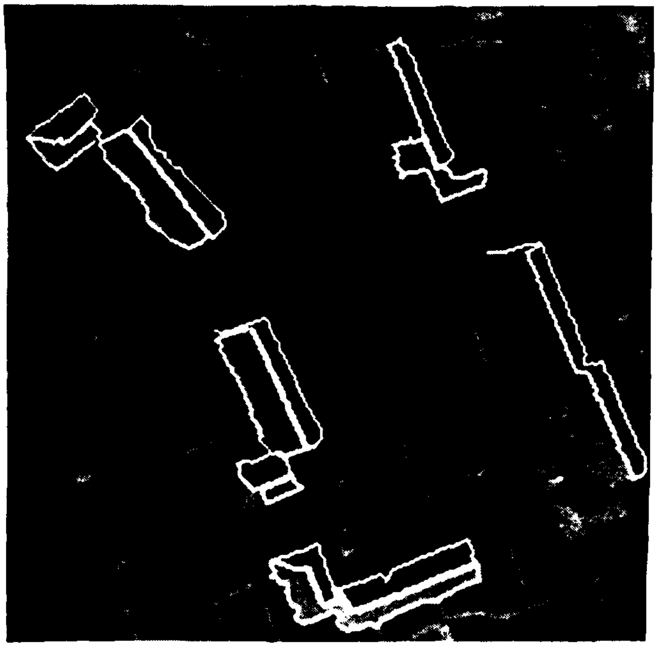


Figure 17: Final results of running the system on the entire image. The initial segmentation produces good candidates for three sunlit roof portions and all shadows. The sunlit roofs, or, conversely, the shadows, then predict the location of the shaded roof portions in the large unsegmentable region.

Appendix C

Resegmentation Using Generic Shape: Locating General Cultural Objects

By: **Pascal Fua** and **Andrew J. Hanson**

Artificial Intelligence Center, SRI International

Submitted to AAAI-86,
The National Conference on Artificial Intelligence,
August 11–15, 1986,
Philadelphia, Pennsylvania.

TRACK: Science Track.
TOPICS: Perception and Signal Understanding (Vision),
Automated Reasoning (Spatial Reasoning).

March 31, 1986

Resegmentation Using Generic Shape: Locating General Cultural Objects[1]

Pascal Fua and Andrew J. Hanson

Artificial Intelligence Center, SRI International
333 Ravenswood Avenue
Menlo Park, California 94025

Submitted to AAAI-86,
The National Conference on Artificial Intelligence,
August 11–15, 1986,
Philadelphia, Pennsylvania.

TRACK: Science Track.
TOPICS: Perception and Signal Understanding (Vision),
Automated Reasoning (Spatial Reasoning).

ABSTRACT

We locate and outline cultural objects in aerial scenes by performing a model-based resegmentation of an initial low-level scene partition. To accomplish this, we define generic data structures for two-dimensional rectilinear shapes, along with robust rules for parsing the image geometry and performing a semantic resegmentation. We apply the system successfully to aerial imagery containing complex cultural objects.

[1]The work reported here was partially supported by the Defense Advanced Research Projects Agency under Contract MDA903-83-C-0027 and by the U.S. Army Engineer Topographic Laboratories under Contract DACA72-85-C-0008.

1 Introduction

People can often perceive and label objects they have never seen before using generic functional and structural concepts. Automating the ability to locate generic object instances in a scene is a fundamental problem of image understanding. In this work, we suggest a promising approach to a portion of this problem and use it to extract rectilinear cultural objects from aerial imagery.

An example of a cultural object that might be identified using a generic shape model is shown in Figure 1, along with a typical low-level partition providing rough shape information. Low-level partitioning methods will always make substantial errors in the object delineation task because they lack knowledge of the object model and its context. Higher-level object-modeling approaches (e.g., GHOUGH [Ballard, 1981] or ACRONYM [Brooks, 1981; Binford, 1982]) possess knowledge of specific model templates, but cannot deal either with generic shapes or with anomalies in the image or its partition. Other approaches, such as the building finder of Nevatia and Huertas [1985] and the airport-extraction system of McKeown, et al. [1984, 1985], still impose strong conditions on allowed shapes and context and have insufficient ability to compensate for inaccurate segmentations and incomplete edge maps.

Our goal is to start with information like that in Figure 1 and locate instances of generic objects. We will accomplish this for the case of cultural objects in aerial imagery by doing the following:

- **Use generic models** for cultural objects instead of rigid templates.
- **Define noise-tolerant parsing rules** that build generic object models from a partition of the image.
- **Resegment the image** by combining generic model-based predictions of shape, knowledge of probable segmentation anomalies, and image-based path-finding operations.

Our approach derives its effectiveness from the unique interaction between high-level and low-level knowledge about the image in the resegmentation process.

We begin by defining data structures and a notation to support our task. We then formulate a set of rules used to instantiate these data structures starting from an image and to obtain information needed for the resegmentation procedures. Finally, we show some examples of the results obtained when our approach is applied to real images containing buildings with complex shapes, followed by a brief discussion of how the methods can be extended to other domains.

2 A Vocabulary for Generic Rectilinear Shape

The elementary geometric "phoneme" upon which we build higher-level rectilinear structures is the *edge*, a set of contiguous image points lying approximately on a straight line and having a significant, uniform image derivative and direction. Our preferred approach to computing atomic edges is to start from *region boundaries* generated by a syntactic partitioning algorithm (see Fua and Hanson [1985] for more motivating details of this approach).

We replace the standard definition of edge orientation [Nevatia and Babu, 1978, 1980] by a more semantically significant orientation based on image regions. This orientation may *differ* from the definition of orientation based on the sign of the derivative across the edge when the sign of the figure-ground intensity difference changes around the object boundary. Region-based orientations support spatial reasoning tasks that are difficult using derivative signs alone.

The data structures that form the basis for our approach to generic rectilinear shape recognition are summarized graphically in Figure 2, and are defined as follows:

- **Pixels.** Image data, perhaps including derived data such as that produced by convolving the image with various operators.

- **Atomic Edges.** Elementary, contiguous sets of pixels satisfying a straight-edge criterion and having an assigned orientation.

- **Edges.** Sets of collinear atomic edges that appear semantically related. The edge data structure may include atomic edges perpendicular to the edge itself; these perpendicular edges are used in the delineation process as linking path predictors.

- **Edge Pairs.** Pairs of edges are associated when possible into rectilinear geometric structures such as parallels, corners, and T's. The structure that forms the basis for the bulk of the geometric reasoning process is the *parallel*, shown in Figure 2 as two adjacent parallel lines with counterclockwise relative orientation.

- **U Structure.** U structures result when the ends of a parallel are joined by a mutual interior corner.

- **Box Structure.** Box structures result when both ends of a set of parallel lines are closable by corners.

In Figure 3, we list the relationships that may be formed among the elementary structures. There is a very strong resemblance to the structures that are formed by elementary edges. All open circles denote a relationship of some kind among basic structures, with the different letters within the circles signifying the type of boundary-closing rules that should be used to complete the particular topology. Below is a summary of the meaning of each structure in Figure 3.

- **Line Relationship.** Two sets of parallels that obey a rough collinearity criterion are joined, much as a set of collinear atomic edges are merged as components of a composite edge.

- **Corner Relationship.** Two sets of perpendicular parallels form a corner, just as edges do.

- **T Relationship.** A T formed from parallels is usually evidence that the enclosed area is to be merged together; this contrasts with the case for T-shaped edge structures, where the T may be evidence for breaking apart composite edges.

- **Parallel Relationship.** A pair of parallel structures that are parallel to one another may be independent, or may be evidence for a missing parallel structure linking a set of ends.

- **Cross Relationship.** The cross relationship is actually a shorthand for a circular list of four corner relationships that also form four T's.

- **Shared-Edge Relationship.** Structures sharing edges occur often in complex objects with multiple semantic pieces or significant noise sources in the middle of a single semantic structure. Shared edges can consist of a single physical edge with opposite orientation interpretations in two adjacent structures, or two distinct parallel edges that are interpretable as arising from a single physical edge. We denote shared edges in Figure 3 by a filled-in circle tangent to the common edges of the parallels.

This basic vocabulary can now be used to construct a language of rectilinear structures, which, in turn, characterize cultural objects (see, e.g., Shirai [1978] and Tavakoli [1980]). Our representation is closely related to the *generalized cone* concept [Blum, 1973; Binford, 1971; Brooks, 1981; Rosenfeld, 1986], except that it emphasizes enclosable associated areas rather than single areas swept out along a skeletal core.

In Figure 4, we give the symbolic representations that would result from error-free parses of a number of common cultural shapes. A very important point to note is that the *depiction* of the structure must be thought of as a *symbol* for an internal *data structure*, and not as a literal *picture* of the edges in the image.

In Figure 5, we illustrate the behavior of the representation of a U with a rectangular bump as the data become increasingly noisy or undergo successive coarsening of the image resolution. We see that the kinds of noise and confusion that may result from resolution-dependent effects are handled correctly. In particular, it is often very difficult to distinguish an almost-invisible protruding structure from noisy line data. The process of grouping related atomic edges (e.g., related by being parallel and in sequence on the same region boundary) into a composite edge is very effective in maintaining semantic consistency across scales and in the presence of noise.

In the next section, we present the construction rules needed to parse the image geometry.

3 Rules for Construction of a Geometric Object Representation

The theoretical approach that we used above to define a generic representation is not well-defined in isolation, but requires for its implementation a concrete description of a parsing mechanism. We therefore define a set of parsing rules that adequately circumvents the ambiguities and instabilities found in the usual skeleton-parsing procedures [Blum, 1973]. The necessity for providing such a specific, noise-tolerant prescription as an interface between a theoretical knowledge-based vision system structure and the real world is often overlooked.

In Figure 6, we present an abbreviated outline of the layers in the image parsing procedure. While the outline is superficially linear, in actual implementation its structure is much more like a rule-based, goal-satisfaction architecture. In particular, there is a considerable amount of backtracking – deleting previously hypothesized ways of satisfying the goal of making a consistent structure – and recursion. A particular association of an edge or parallel structure with another object may be made and broken a number of times as new geometric knowledge from neighboring structures is brought to bear. The early steps in the procedure may be understood as "stashed facts" that serve as preconditions for the satisfaction of the structure-building goals.

While many methods might be used to extract elementary edges from an image, we have found that a very effective approach is to begin with a rough image partition generated by an Ohlander-style segmenter [Ohlander, et al. 1978; Laws, 1984] and then use the strong correlation between the resulting region boundaries and strong image derivatives to extract edges [Fua and Hanson, 1985]. We note also that the computations below depend upon only a small number of parameters such as minimum edge length and angular tolerance for collinearity, and, optionally, upon such values as expected structure size. These numbers can, in principle, be roughly determined from the known resolution and camera model of an image.

The following steps describe the mechanism by which complex rectilinear data structures are extracted from an image.

Construction Rules

- Get atomic edges and orientations. We typically accumulate edge-elements from region boundary pixels satisfying minimum length and collinearity requirements.

- Build composite edges. Group atomic edges that individually have similar spatial orientations into composite edges.

- Build binary edge relationships. Group pairs of composite edges as follows:

  – Find all parallels, corners, T's.

  - Delete crosses.
  - Merge parallel edges – two edges parallel to the same edge are merged when consistent.
  - Break parallels if a part is semantically too narrow or wide.

- Build closures. Make area-bounding parallel structures as follows:

  - Break T's within local structures and recompute the structures. An edge pointing at a junction between two atomic edges in a composite edge is strong evidence that there is a semantic break in the composite edge at that point.
  - Find closing edges of all parallels and move the edges that are subparts of other distinct structures to the newly closed ones. (Remove a merged edge from a parallel if it is a U's closing edge.)

- Build networks of parallels. This is accomplished by performing a set of association operations very similar to those performed with low-level edges (see Figure 3). The relationship labels in the network have special meanings with respect to the kinds of linking operations that may be performed in the final delineation step.

  - Build composite collinear parallel relationships.
  - Build parallel, corner, and T relationships.
  - Identify crosses (four corners).
  - Identify relationships between structures sharing edges.

- Close areas by carrying out region closure predictions and computing closure paths. Simple models for the failure of the partitioning process (e.g., losing edges in the middle of an undersegmented region) are incorporated into the predictions. A typical low-level, prediction-based closure procedure used in this step is that of Fischler, et al. [1981].

We note that a number of the rules involve backtracking and reconstruction that characterize the making and refutation of hypotheses about the geometric structure. For particular semantic contexts, e.g., where assumptions about building structure and illumination models are justified, any section of these rules may be supplemented or modified by rules derived from knowledge of the problem domain.

To illustrate the function of the rules, we show in Figure 7 how sets of edges forming a stair-step are first merged into a composite parallel, then broken apart on the basis of a vertical line; this line and the corresponding gap in the composite parallel constitute strong evidence for the existence of two separate semantic entities within the original composite parallel.

We have designed in some redundancy as an important characteristic of the rule base; several different consistent paths of making and breaking associations will lead to the same or equivalent structures. This helps achieve our goal of making the result relatively stable in the presence of noise and lost edges.

4 Applications to Real Imagery

We now apply the entire procedure to a series of images. In Figure 1a, we show a relatively complex building scene. Starting from the Ohlander-style segmentation overlaid in Figure 1b, we extract atomic edges from the region boundaries. Carrying out our construction procedures for the entire scene, we find the network of associations depicted symbolically in Figure 8; recall that this symbolic network stands for a complete representation of the object using the internal data structures of the system. Applying the final rules for the extraction of a distinct group from a discovered network of rectangles and running the linking procedures, we find the final delineation of the complex building structure shown in Figure 9.

Next, in Figure 10, we show a pair of images of the same house digitized from different sources at different resolutions. Parsing these gives the two symbolic networks of Figure 11, and the resultant structure delineations in Figure 12. The consistent interpretation of the same structure in two very different images illustrates the noise and scale-insensitivity of our approach.

Finally, we examine the image in Figure 13a, which contains a large number of buildings with multiple parts and shaded roofs. The image partition in Figure 13b shows substantial problems with the segmentation due to the confusion of shaded roofs with the background. Parsing and resegmenting the entire scene, we find the buildings and building portions delineated in Figure 14.

We have run the system on a number of building-containing areas in a variety of images and have consistently achieved satisfactory building delineation when the resolution is sufficient to discriminate edges reliably.

5 Conclusions

Our main results are the following:

- **Generic Shape Extraction.** We defined a generic shape vocabulary for cultural objects along with geometry-parsing rules that extract noise-tolerant shape representations from real images.

- **Object Delineation Using Structure Linking and Resegmentation.** Knowledge of object geometry and expected anomalies of a given partitioning process were combined with low-level linking operators to produce delineated and labeled objects.

The paradigm for segmentation correction using generic shape models that we have described in detail here for the case of rectilinear structures is now being extended to other domains, such as the following:

- Cultural linear features, such as roads, canals, and paths.
- Natural linear features, such as rivers, streams, gullies, and canyons.
- Natural objects with fractal boundaries, such as forests, vegetation areas, lakes, and coastlines.
- Three-dimensional rectilinear objects with particular illumination and shadow characteristics seen from arbitrary viewing angles.

It seems feasible in each of these domains to identify certain geometric signatures that would allow object classification based on generic characteristics. In particular, the role of straight edge segments in buildings is played by curvilinear segments in roads and by fractal edge segments in vegetation areas; these segments can then be grouped into partial region-enclosing relationships similar to the corner and parallel edge relationships present in cultural structures.

In summary, we have proposed a framework based on generic structural models that can, in principle, permit an automated system to parse never-before-seen instances of certain object types, perform a model-driven resegmentation, and delineate the object in the image even though the original low-level segmentation may correspond very poorly to the object shape. We have implemented the entire paradigm for the case of rectilinear cultural objects in aerial imagery. The system achieves reliable identification of quite complex cultural objects in a variety of images, thus justifying our confidence in the paradigm.

References

D.H. Ballard, "Generalizing the Hough Transform to Detect Arbitrary Shapes," Pattern Recognition **13**, pp. 111-122 (1981).

T.O. Binford, "Visual Perception by Computer," Invited talk at IEEE Systems Science and Cybernetics Conference, Miami (December 1971).

T.O. Binford, "Survey of Model-Based Image Analysis Systems," The International Journal of Robotics Research, **1**, No. 1, pp. 18–64 (Spring, 1982).

H. Blum, "Biological Shape and Visual Science (Part I)," J. Theoretical Biology **38**, pp. 205–287 (1973).

R.A. Brooks, "Symbolic Reasoning Among 3-D Models and 2-D Images," Artificial Intelligence Journal **16**, (1981).

M.A. Fischler, J.M. Tenenbaum, and H.C. Wolf, "Detection of Roads and Linear Structures in Low-Resolution Aerial Imagery Using a Multisource Knowledge Integration Technique," Computer Graphics and Image Processing **15**, pp. 201–223 (1981).

P. Fua and A.J. Hanson, "Locating Cultural Regions in Aerial Imagery Using Geometric Cues," Proceedings of the Image Understanding Workshop, pp. 271–278 (December 1985).

K.I. Laws, *Goal-Directed Texture Segmentation,* Technical Note 334, Artificial Intelligence Center, SRI International, Menlo Park, California (September 1984).

D. McKeown, W.A. Harvey, and J. McDermott, "Rule-Based Interpretation of Aerial Imagery," Proceedings of IEEE Workshop on Principles of Knowledge-Based Systems, pp. 145–157 (1984); IEEE Trans. PAMI **7**, pp. 570–585 (1985)

R. Nevatia and K.R. Babu, "Linear Feature Extraction," Proceedings of the Image Understanding Workshop, pp. 73–78 (November 1978).

R. Nevatia and K.R. Babu, "Linear Feature Extraction and Description," Computer Graphics and Image Processing **13**, pp. 257–269 (1980).

R. Nevatia and A. Huertas, "Building Detection in Simple Scenes," Univ. of Southern California Intell. Systems Group Technical Report, (September 1985).

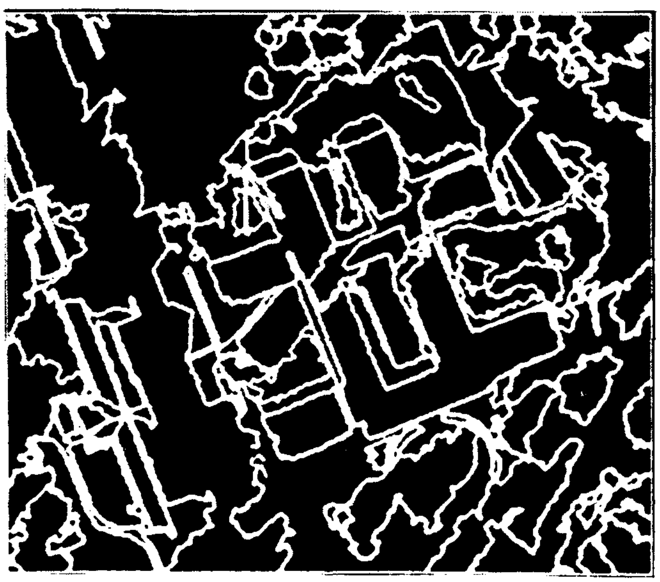
R. Ohlander, K. Price, and D.R. Reddy, "Picture Segmentation Using a Recursive Region Splitting Method," Computer Graphics and Image Processing **8**, pp. 313–333 (1978).

A. Rosenfeld, "Axial Representations of Shape," Computer Vision, Graphics and Image Processing **33**, pp. 156–173 (1986).

Y. Shirai, "Recognition of Man-Made Objects Using Edge Cues," in *Computer Vision Systems*, Ed. A.R. Hanson and E.M. Riseman, Academic Press (New York, 1978).

M. Tavakoli, "Toward the Recognition of Cultural Features," Proceedings of the Image Understanding Workshop, pp. 33–57 (1980).

(a)



(b)

Figure 1: (a) A typical aerial image containing a complex building. (b) A syntactic image partition overlaid on the image.

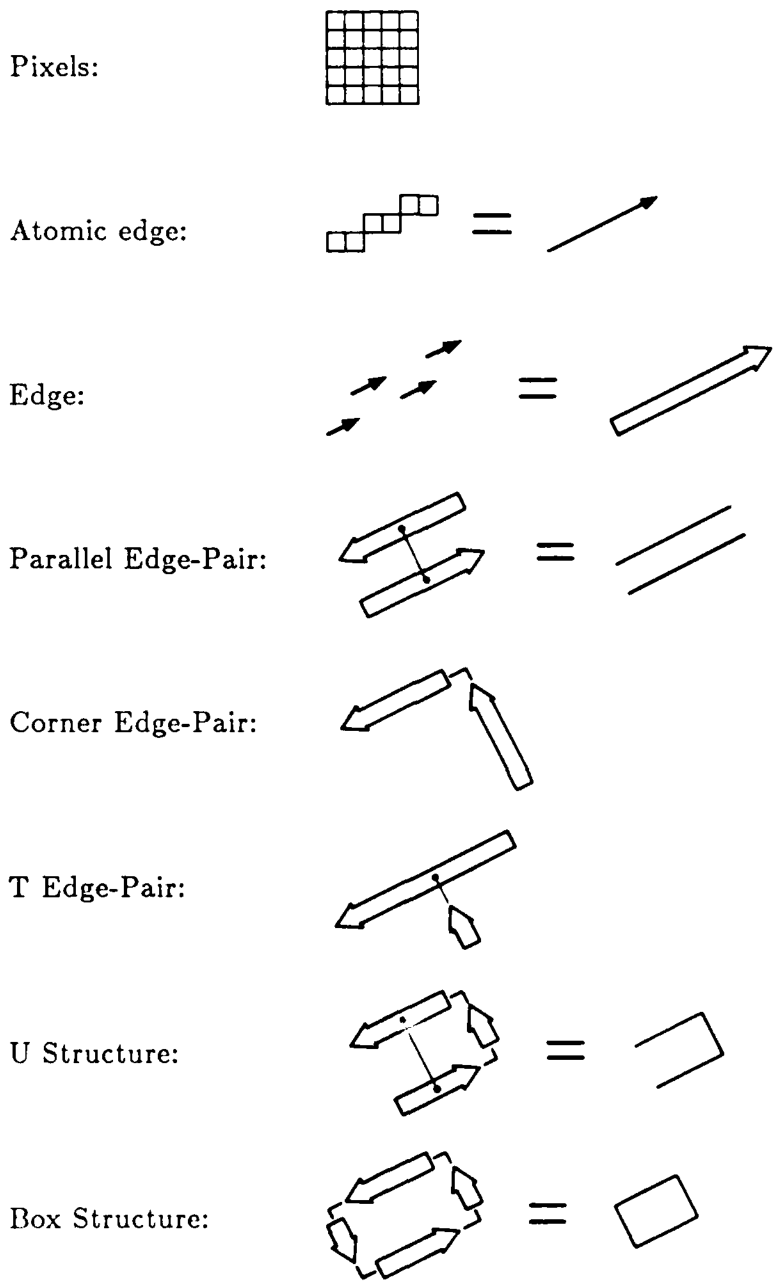


Figure 2: Summary of the definitions and notations used to represent the data structures denoting generic rectilinear objects.

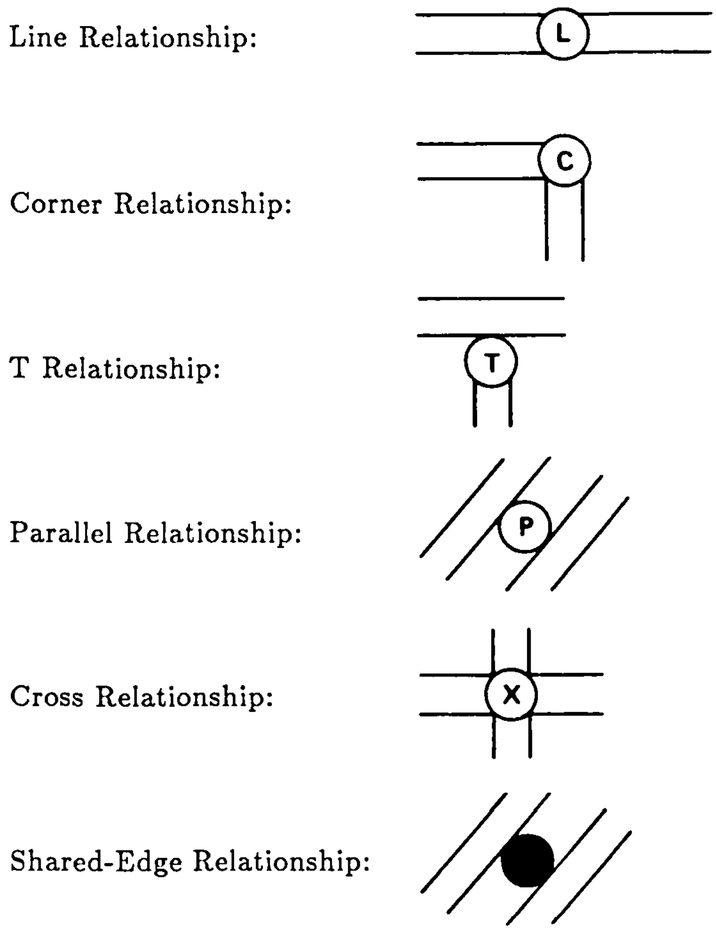


Figure 3: Summary of the relationships among geometric structures that serve as the links in the relationship network characterizing a complex geometric object.

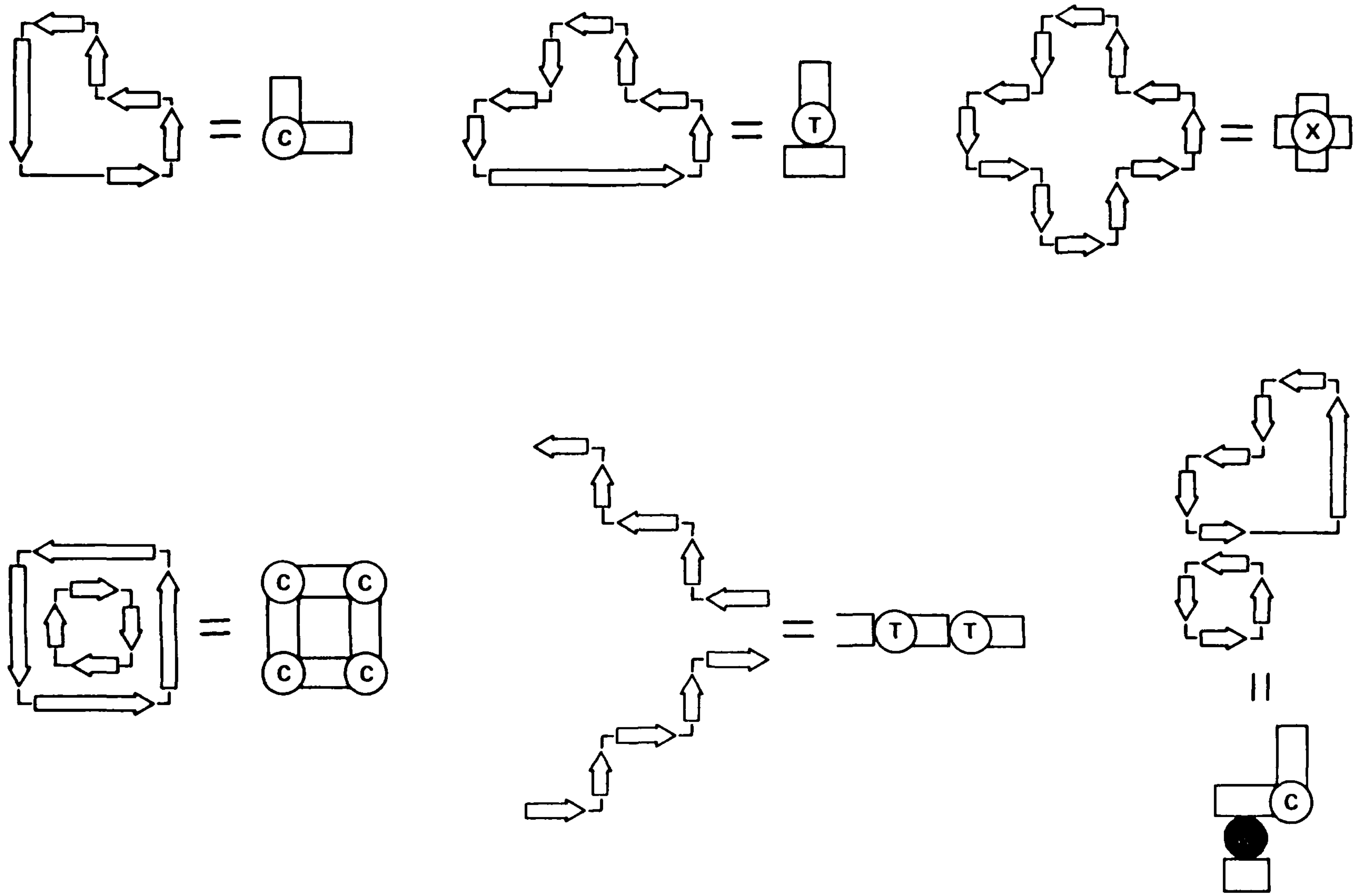


Figure 4: Examples of some typical simple structures that occur in real images and the symbolic notation for their parsed geometry. We show an L, a T, a cross, a courtyard, a set of nested T's, and a shared-edge.

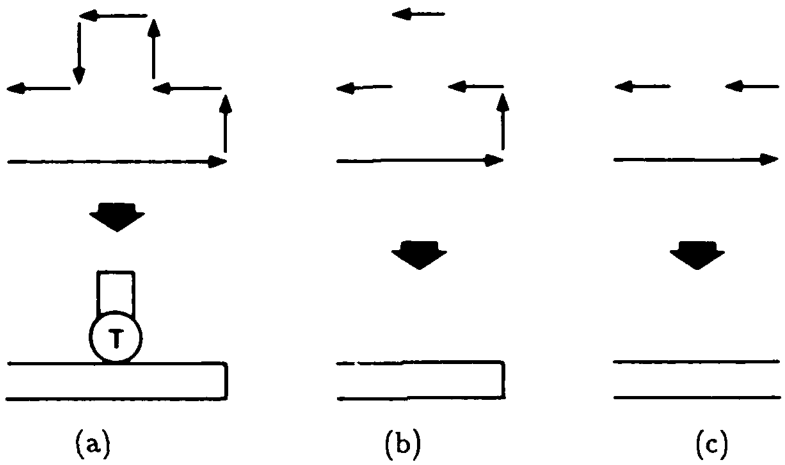


Figure 5: The evolution of interpretations that would be found in a U with rectangular substructure under several changes in image resolution or degradation due to noise, from good in (a) to poor in (c). Atomic edges are shown at the top, with their symbolic interpretation at the bottom.

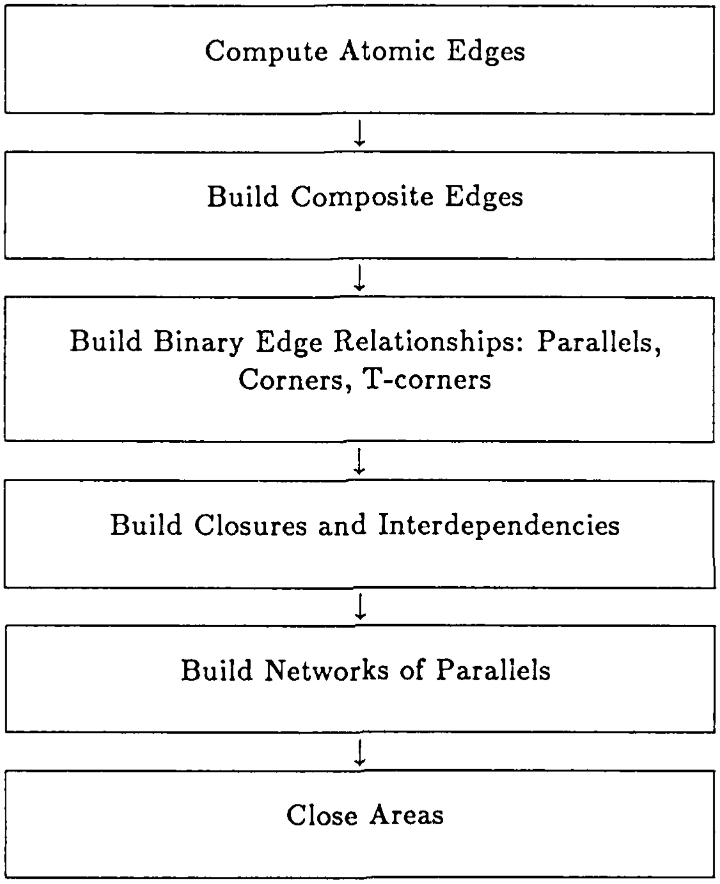


Figure 6: Outline of the rules for the extraction of cultural objects formed from generic rectilinear structures in aerial imagery.

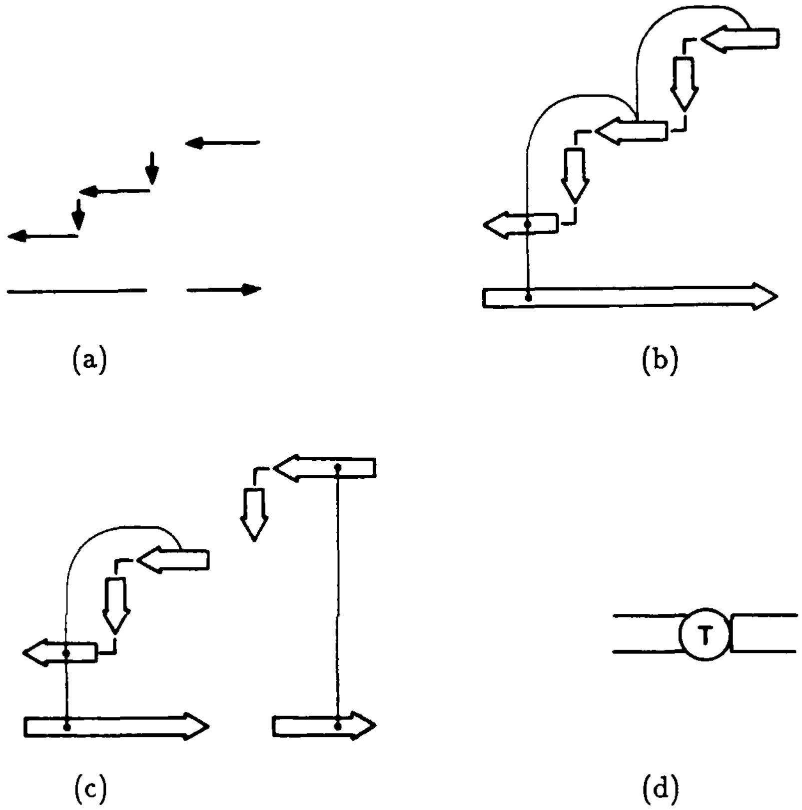


Figure 7: The parsing of parallel structure with step-like internal structure. (a) The atomic edges. (b) The composite edges merged into a composite parallel (resulting from merging parallels with a common member). (c) Result of T-breaking: a vertical line is evidence suggesting separation of two portions of an original composite line, thus breaking the parallel structure also. (d) Final symbolic parse: a parallel on the left forms the stem of a T structure; the T structure is joined to the linking line of the U on the right.

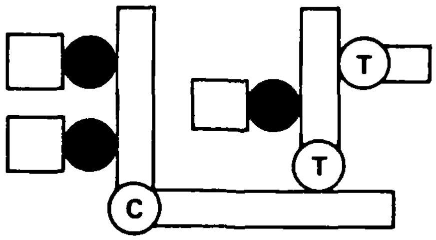


Figure 8: The resultant symbolic representation of the parse network of the entire building structure.

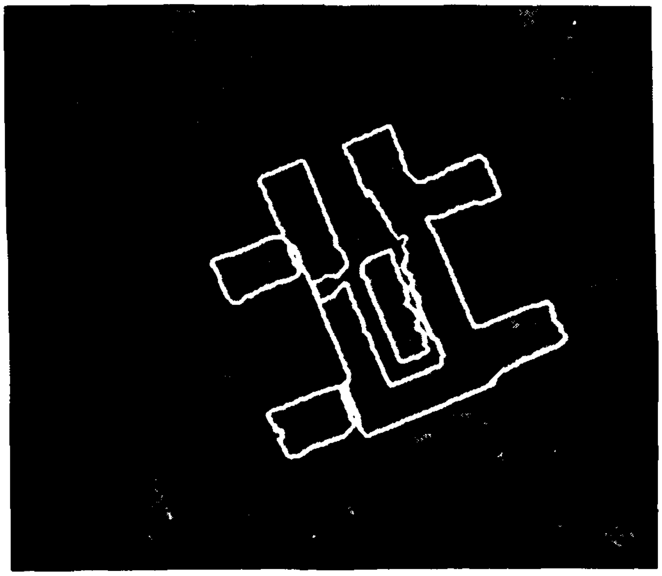
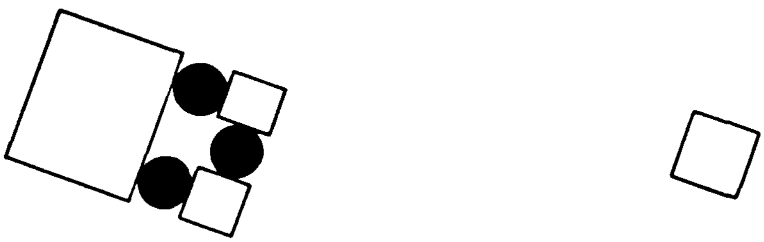


Figure 9: The result of using the parsed geometry to predict region closing paths and joining operations, yielding the final semantically-motivated building shape.

(a)

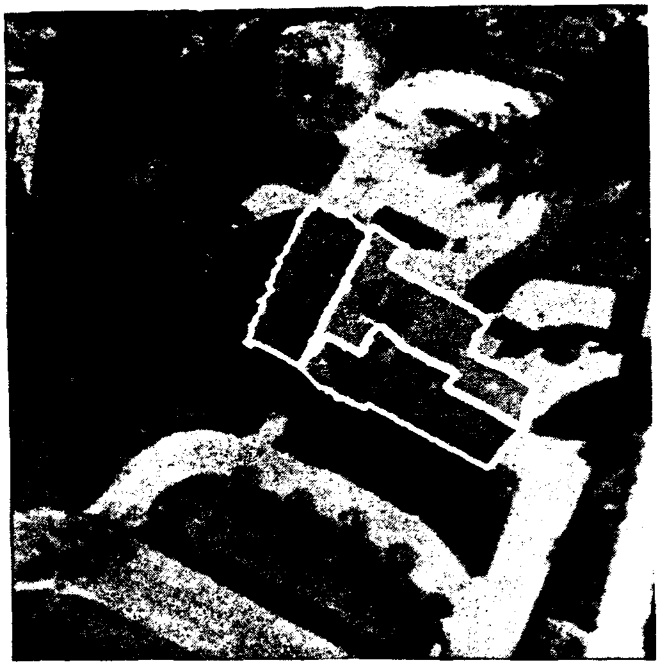
(b)

Figure 10: (a) A high-resolution aerial image containing a complex building. (b) Low resolution image of the same building on a different day.

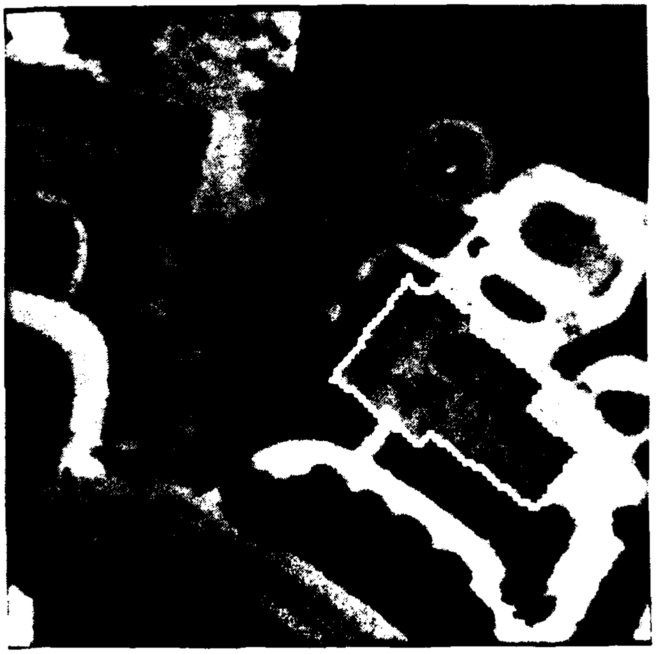


(a) (b)

Figure 11: The resultant symbolic representations of the parse network of the building structure at the two resolutions, (a) and (b).

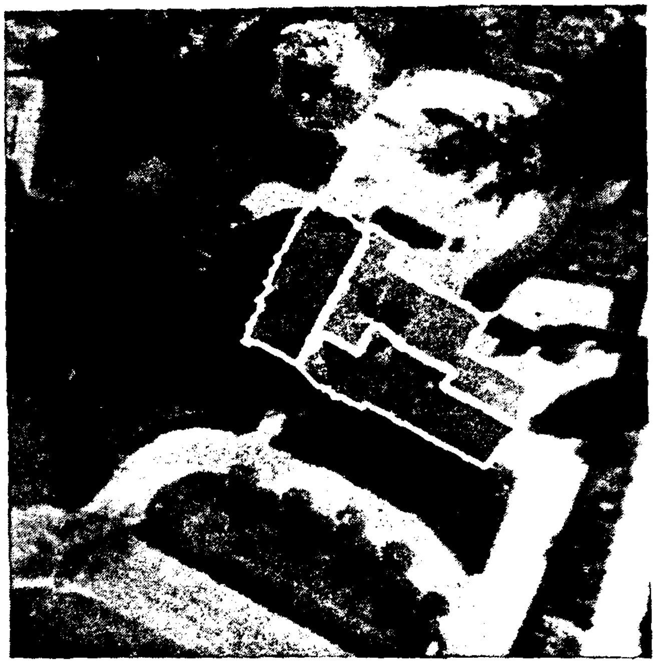


(a)



(b)

Figure 12: The result of using the parsed geometry to predict region closing paths and joining operations, yielding the final semantically-motivated building shapes at the two resolutions (a) and (b).



(a)

(b)

Figure 12: The result of using the parsed geometry to predict region closing paths and joining operations, yielding the final semantically-motivated building shapes at the two resolutions (a) and (b).

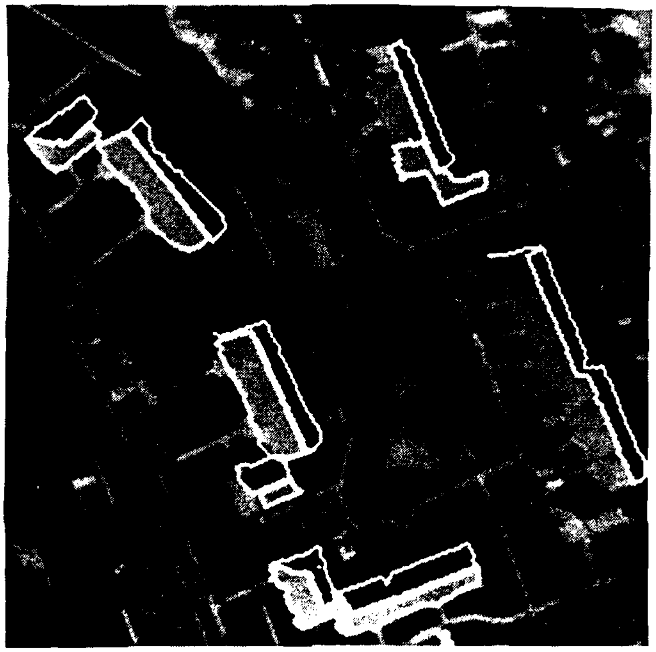


Figure 14: The result of using the resegmentation procedure to outline all the identifiable building areas. At this resolution, some of the more complex buildings at the right have unresolvable structure.

END

DTIC

7-86