SRI International

USING GENERIC GEOMETRIC KNOWLEDGE TO DELINEATE CULTURAL OBJECTS IN AERIAL IMAGERY

Technical Note 378

March 31, 1986

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

Artificial Intelligence Center

Computer Science and Technology Division

APPROVED FOR PUBLIC RELEASE

DISTRIBUTION UNLIMITED

SRI Projects 5355 and 8645

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.

SRI International

333 Ravenswood Ave. • Menlo Park, CA 94025

(415) 326-6200 • TWX: 910-373-2046 • Telex: 334-486

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.

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 gulleys, 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).

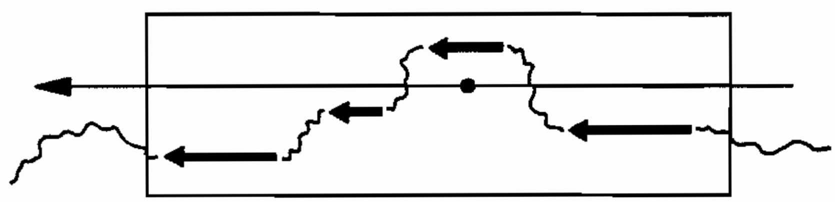


Figure 1: Each thick arrow represents one of a set of straight edge segments lying in a region boundary. This set of atomic edges forms a composite edge for geometric reasoning purposes. The long arrow denotes the semantically correct direction of the composite edge, computed from a weighted average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

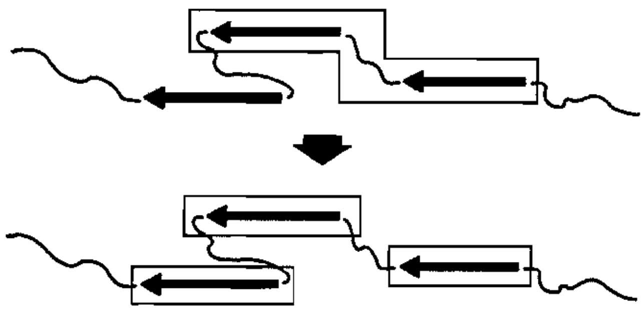


Figure 2: In the first stage of composite edge accumulation, the two contiguous edges enclosed in the box at the top are associated. However, a second stage checks the consistency of the geometry and discovers that the next edge in this region boundary lies to the right of the leftmost end of the tentative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

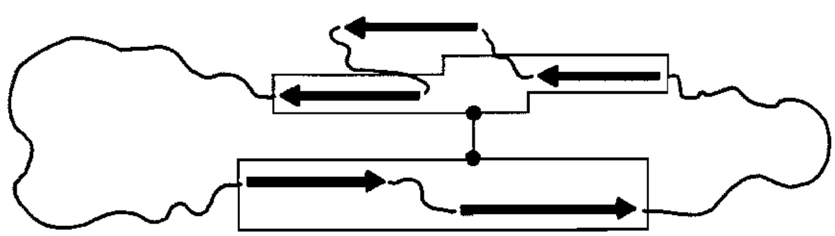


Figure 3: Here there are three short edges that might be logically linked with the bottom long edge, except that two short edges overlap because one belongs to an appendage. Using the knowledge that such an appendage is probably due to a neighboring part of a yard or patio, rather than the building itself,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

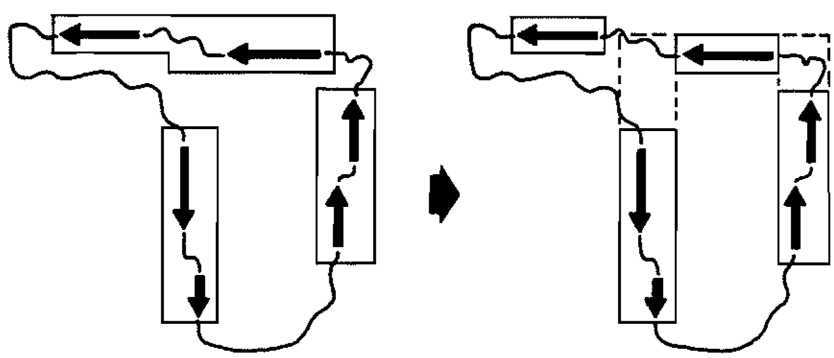


Figure 4: Backtracking by breaking a composite line to form a U-shaped structure. The U-shape is preferred because it provides strong evidence for a cultural object.

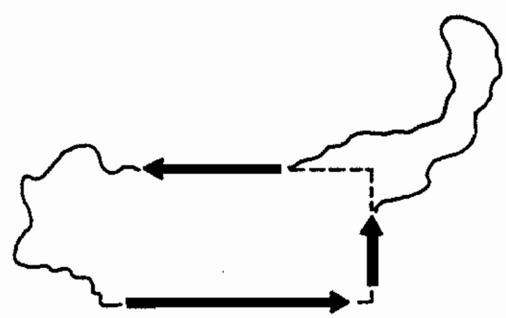


Figure 5: The existence of a good U structure here serves to predict that the missing portions of the corner should be constructed if possible. If the line finder successfully finds a good path in the predicted geometric vicinity, the erroneous appendage is removed and the is region split in two along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

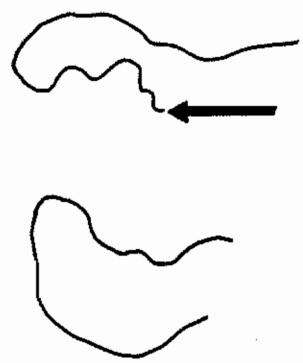


Figure 6: One may use the same geometric rules as for single regions when dealing with multiple interior boundaries of regions with holes because the orientation of edges in these "island" regions is reversed. In the case shown here, two neighboring island regions have edges that can be combined to form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

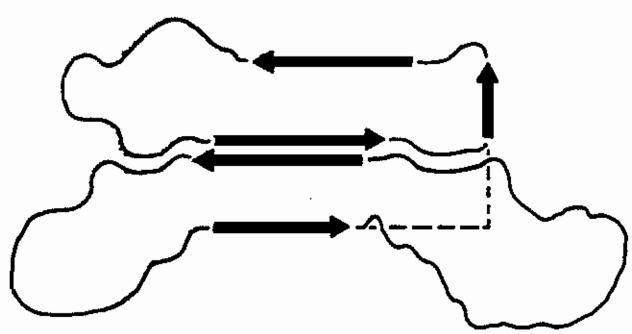


Figure 7: The upper **U** closure determines the path predicted for a meaningful closure of the lower parallel, both of whose ends are open.

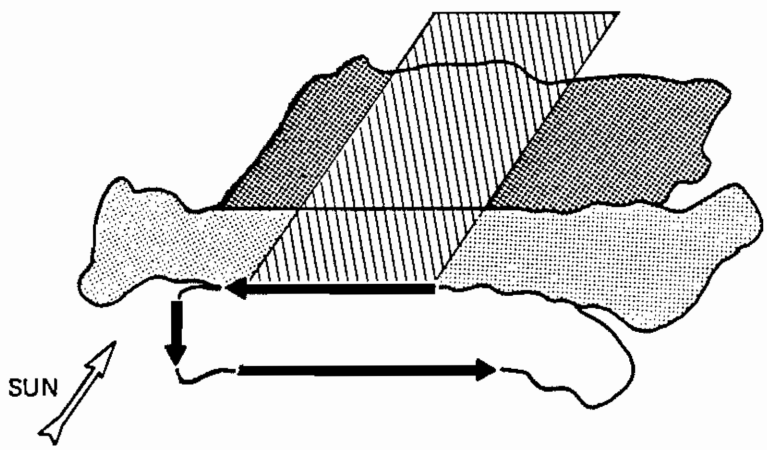


Figure 8: A sunlit roof portion with a **U** structure. The edge elements on the shaded side of the structure are used to look for regions that might be the shaded portion of a peaked roof.

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

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

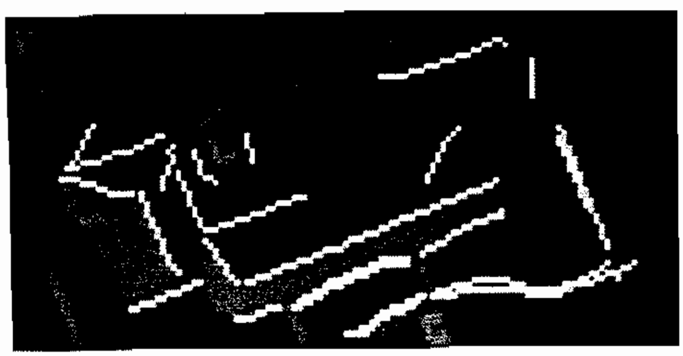
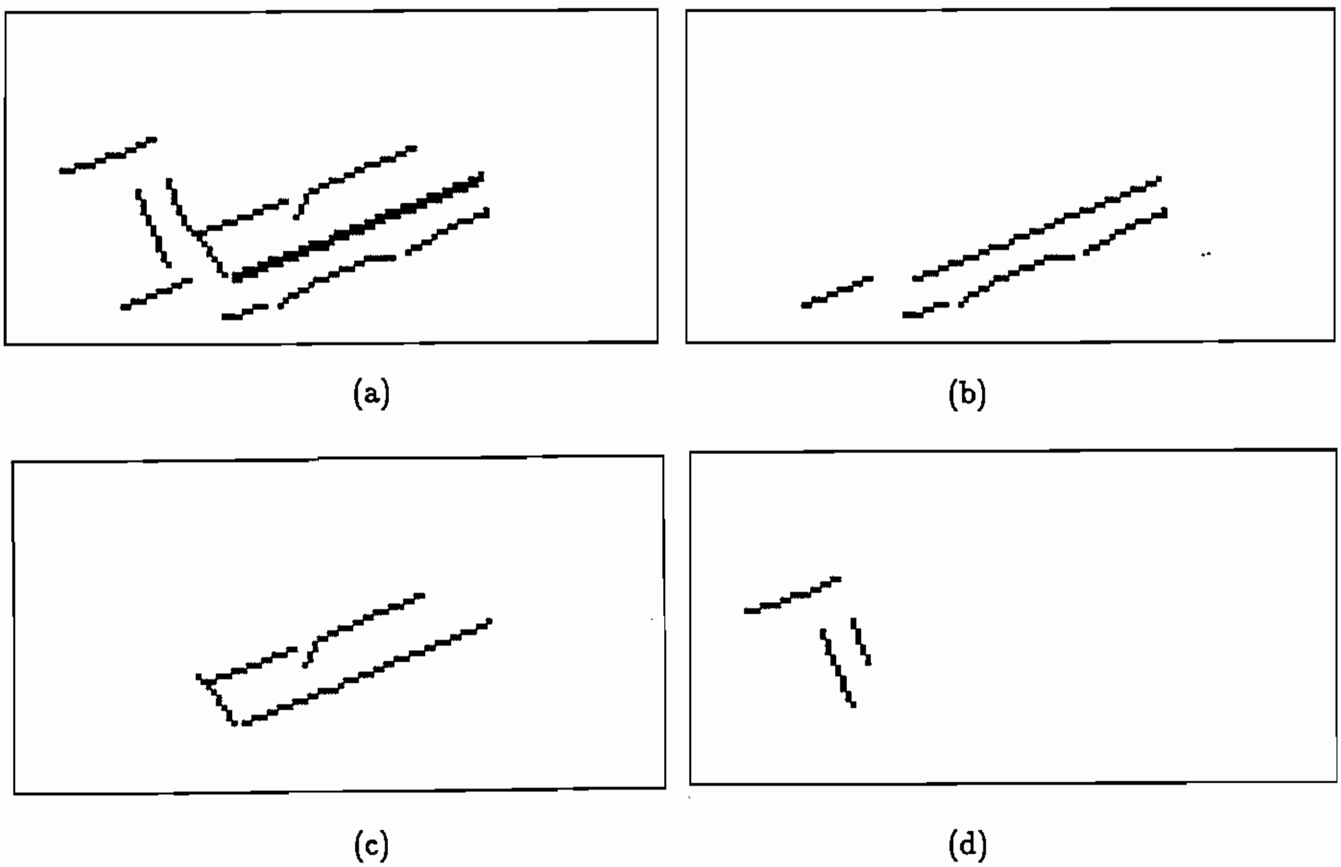


Figure 11: The straight edges used to produce the geometric structures characteristic of the cultural object.



(a) (b) (c) (d)

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.

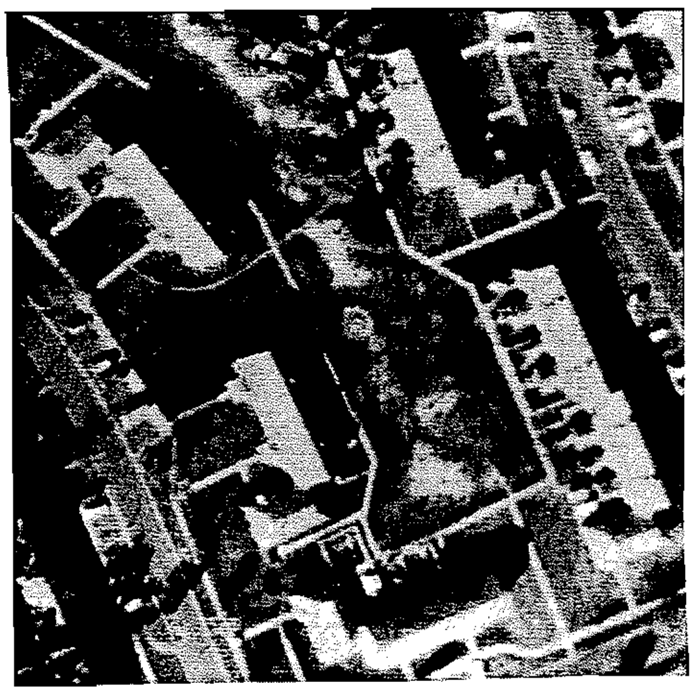


Figure 14: A large image containing the previous example as a subimage.

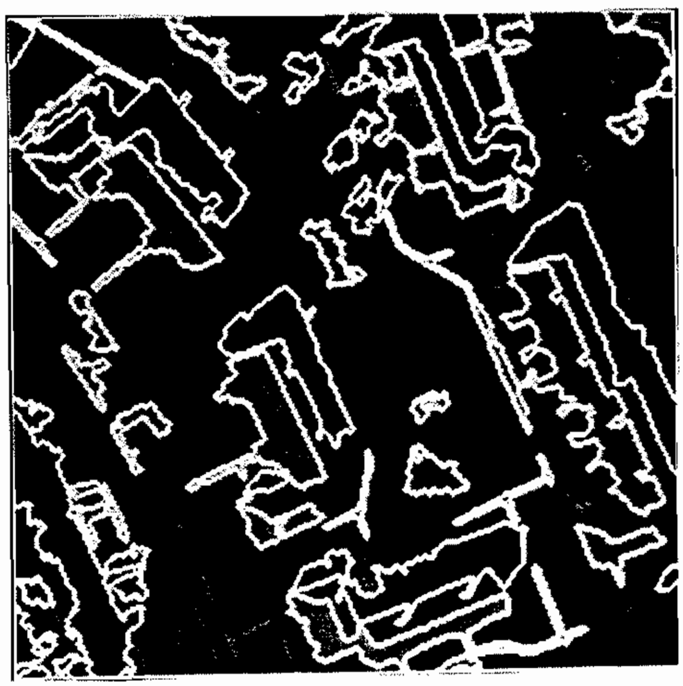


Figure 15: The segmentation boundaries of the large image.

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

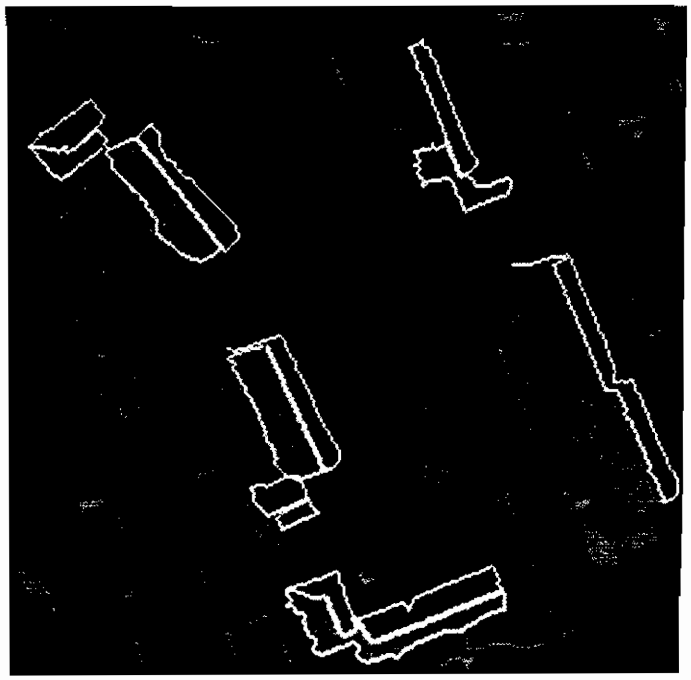


Figure 17: Final results of running the system on the entire image. The initial segmentation produces good candidates for three sunlit roof portions and all shadows. The sunlit roofs, or, conversely, the shadows, then predict the location of the shaded roof portions in the large unsegmentable region.