Pattern Recognition Letters 5 (1987) 243–252
North-Holland

March 1987

Resegmentation using generic shape: Locating general cultural objects

Pascal FUA and Andrew J. HANSON

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

Received 3 November 1986

Abstract: As a step toward automating the ability to locate generic objects in an image, we propose an approach based on model-driven correction of an initial low-level scene partition. To accomplish this, we define generic data structures for geometric shapes, along with robust rules for parsing the image geometry and performing a shape-motivated resegmentation. We successfully apply the system to the task of locating and outlining complex rectilinear cultural objects in aerial imagery.

Key words: Segmentation, generic data structures, aerial imagery.

1. Introduction

People can often perceive and label objects they have never seen before using *generic* functional and structural concepts. One way to emulate this ability is to segment an image and then generate a set of labels for the resulting regions. Syntactic image partitions are indeed a good starting point for generic shape analysis because the resulting regions generally provide relevant, context-free information about unpredictable shapes. Unfortunately, since objects of interest are rarely characterized solely by the statistical signatures upon which standard segmentation methods rely, these methods will always be prone to uncorrectable errors. Thus, although a great deal of attention has been paid to the label-generation step, the weakest link in the process is the segmentation procedure itself.

The goal of this work is therefore to investigate the use of generic shape constraints to refine and correct an initial image segmentation. We will refer to this process as 'resegmentation', since we are effectively revising an initial hypothesis for the scene partition provided by a low-level segmentation algorithm; a shape hypothesis derived from the segmentation is used to predict the corrected outline of a semantic object, and the validity of this prediction is then evaluated.

We use generic, as opposed to specific or template-like, shape models so that we can find instances of unpredictable shapes. This approach provides us with a logically complete semantic extension of the conventional low-level segmentation process that overcomes many of the inherent weaknesses of purely syntactic methods.

Other approaches to object delineation that depend upon low-level partitioning methods will always make substantial errors in their task because they lack knowledge of the objects of interest and of the scene. Most current object-modeling approaches (e.g., GHOUGH (Ballard, 1981) or ACRONYM (Brooks, 1981; Binford, 1982)) use explicit shape 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.

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.

(1984,1985), still impose strong conditions on allowed shapes and context and have insufficient ability to compensate for inaccurate segmentations and incomplete edge maps.

To achieve the goal of reliably outlining objects of interest, we have developed a rule-based re-segmentation system that allows the available semantic knowledge to interact closely with the low-level image-data. The following bodies of knowledge form the core of our approach:

- *Generic geometric models* to describe arbitrarily complex shapes.
- *Models of expected segmentation anomalies* to generate relevant hypotheses for segmentation correction.
- *Noise-tolerant parsing rules* that form the critical interface between real image data and the theoretical models.

We confirm the validity of this method by applying it to the problem of delineating cultural objects in aerial imagery. Our goal, therefore, is to start with challenging images like that in Figure 1 and to locate instances of generic cultural objects.

We begin by defining geometric models to support our task. Next we describe the characteristics of the low-level image partitions we use as input to the system. We then formulate a set of rules used to instantiate these models starting from such image partitions and to obtain the information needed for the resegmentation procedures. Finally, we show some examples of the results obtained when our approach is applied to real images containing complex cultural objects.

2. Shape models

Cultural objects usually have very characteristic rectilinear geometric structures that are easy for people to see but difficult for a machine to extract. To describe such structures of arbitrary complexity, we need an appropriate generic shape model.

First, we replace the standard definition of edge orientation (Nevatia and Babu, 1978, 1980) by a more topologically significant orientation based on image regions: exterior region edges are oriented with the interior of the region on the left of the traversal directions, as are interior boundaries surrounding 'holes' in the region. This orientation may *differ* from the definition of orientation based on the sign of the derivative across the edge when the figure-ground intensity difference changes sign on the object boundary. Region-based orientations enforce topological consistency and support spatial reasoning tasks that are difficult using derivative signs along. (See Fua and Hanson (1985) for more details on this approach.)

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

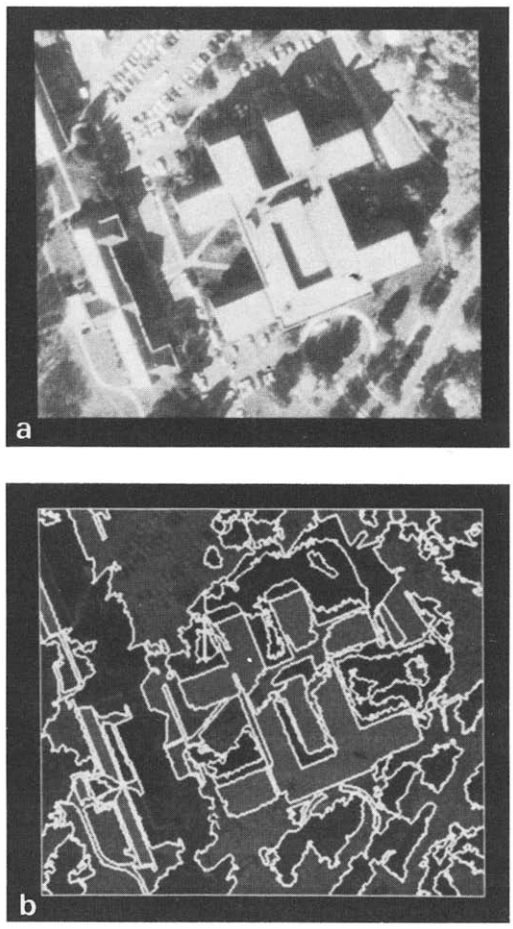


Figure 1. (a) A typical serial image containing a complex building. (b) A syntactic image partition overlaid on the image.

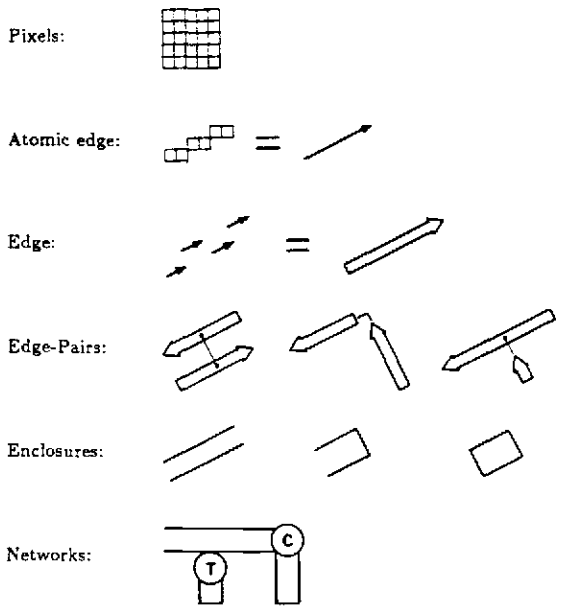


Figure 2. Summary of the definitions and notations used to represent the data structures denoting generic rectilinear objects.

pixels satisfying a straight-edge criterion and having an assigned orientation.

• *Edges* – Sets of collinear atomic edges that appear to represent semantic edges that have been broken by the segmentation process.

• *Related edge pairs* – Pairs of edges associated into rectilinear geometric structures such as parallels, corners, and T's. The parallels play a central role in this system because they are very characteristic of cultural structures and enclose areas that can be used for area based reasoning.

• *Enclosures* – Parallels whose ends may or may not be closed by perpendicular edges. Parallels with end-closures are described as U's and boxes.

• *Networks* – Sets of geometrically related enclosures that will be used to describe rectilinear objects of arbitrary complexity.

In Figure 3, we list the geometric relationships that may be formed among enclosures in order to build the networks. The geometry of these relationships is very similar to the geometry of related edges and can be parsed using analogous rules. All open circles denote a relationship of some kind among enclosures, with the different letters within the circles signifying the type of boundary-closing rules that should be used to complete the particular structure. 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 can be joined, much as a set of collinear atomic edges may be merged as components of a composite edge.

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

• *T relationship.* A parallel that may be linked into a perpendicular composite edge belonging to another parallel without breaking any atomic edge forms a T with the other-parallel.

• *Parallel relationship.* Two sets of parallels that are parallel to another may be independent, or may be evidence for a missing parallel structure between them.

• *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.

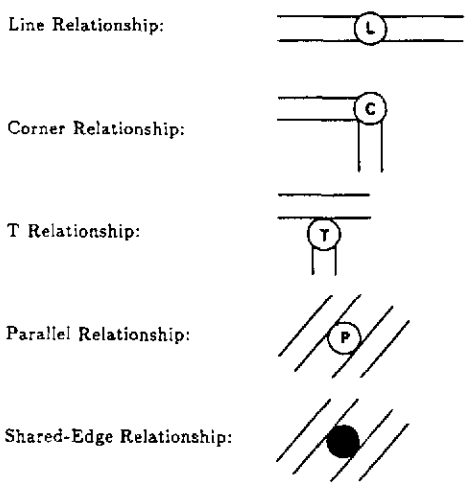


Figure 3. Summary of the relationships among geometric structures that serve as the links in the relationship network characterizing a complex geometric object.

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* con-

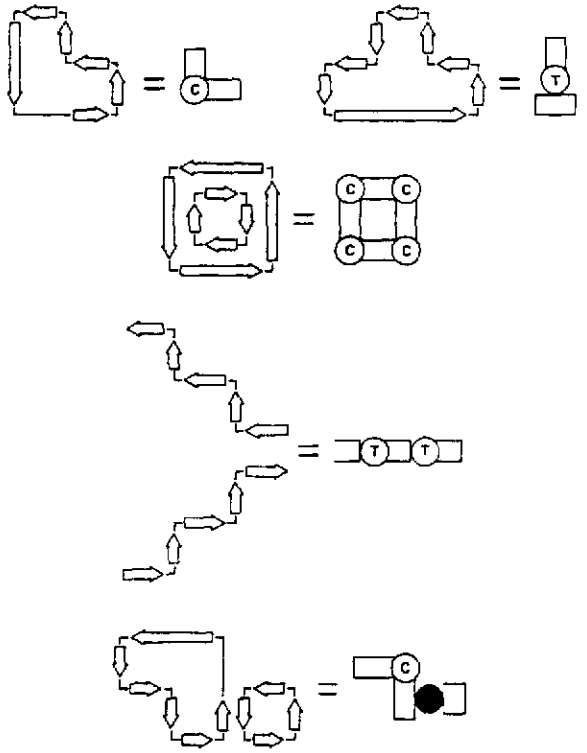


Figure 4. Examples of some typical simple structures that occur in real images and the symbolic notation for their parsed geometry. We show an L, a T, a courtyard, a set of nested T's, and a shared-edge.

cept (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. Note 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.

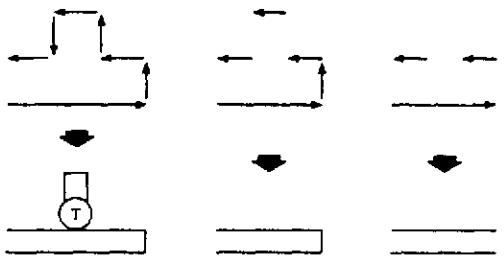


Figure 5. The evolution of interpretations that would be found in a U with rectangular substructure under several changes in image resolution or degradation due to noise, from good in (a) to poor in (c). Atomic edges are shown at the top, with their symbolic interpretation at the bottom.

In Figure 5, we illustrate the behavior of the representation of a parallel structure closed at one end and having a rectangular bump as the data become increasingly noisy or undergo successive coarsening of the image resolution. Longer edges will usually be broken up and shorter ones will be lost. We see that this kind of noise and confusion is 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., those 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. Characteristics of the segmentation

We have chosen to use an Ohlander-style segmentation (Ohlander et al., 1978; Laws, 1984) as initial input to our system. In this work, we assume that the initial segmentation may be inaccurate but still contains significant information regarding the objects of interest.

Because the partition is computed using recursive histogram splitting, it has the following properties:

- The boundaries of regions tend to correspond to high intensity gradients.
- Regions tend to bleed and edges are lost in the middle, resulting in undersegmentation of objects.
- Objects of interest may also be oversegmented, i.e., broken into several pieces.

Resegmentation procedures therefore include:

- Hypothesizing and discovering the missing edges.
- Grouping semantically related areas.

The tendency toward high gradients in the region boundaries permits us to extract reliable straight edges from these boundaries using statistical consistency of the image gradient orientations. Furthermore, these edges can be grouped into geometric structures that typically enclose areas

with uniform statistical properties.

Because significant edges may be missing from region boundaries and merged in the middle of regions because of noise problems, we use the successfully found edges, combined with model-based hypotheses, to look for the missing edges.

Regions corresponding to oversegmented objects are also related using model-based geometric predictions. The presence of these oversegmented objects is in fact one of the major reasons why the region-independent structure-clustering mechanism described below is required.

4. Rules for construction of a geometric object

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 that incorporates knowledge about the segmentation anomalies. 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 neglected.

In Figure 6, we present an abbreviated outline of the layers in the image parsing procedure. The following subsections give more details of the parsing algorithm and some examples of the geometric rules it utilizes.

4.1. Parsing mechanism

The resegmentation process consists of the following steps:

• *Build elementary structures within single segmentation regions.* To provide the powerful context needed for edge parsing, geometric structures are first computed within single segmentation regions. Atomic edges are first extracted from region boundaries, and are then grouped into composite edges that can in turn be grouped, merged, or broken in order to yield geometrically consistent enclosures.

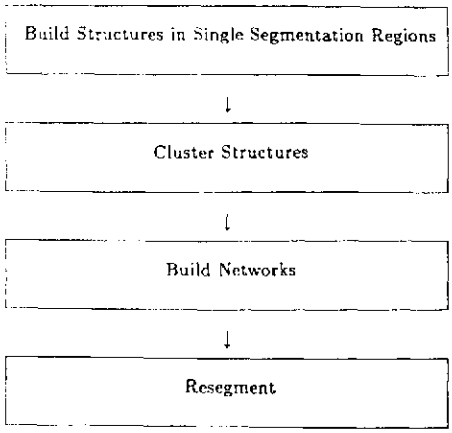


Figure 6. Outline of the rules for the extraction of cultural objects formed from generic rectilinear structures in aerial imagery.

• *Cluster geometric enclosures.* We have used regions as a focus of attention for edge parsing. To accomplish a similar function for structures, we cluster enclosures from arbitrary regions that are potentially semantically related. We associate enclosures that tend to cluster spatially and have either similar statistical properties of image pixels or explainable differences in terms of a semantic model.

• *Build networks.* Enclosures belonging to semantically meaningful objects should be related in a geometrically predictable way. Much in the same way that edges from regions have been grouped into geometric structures, enclosures in the clusters will be chained into consistent networks.

• *Resegment.* These networks can now be used to perform the actual resegmentation: the enclosures are interdependent and geometrically related, and the various types of relationships have special meanings with respect to the kinds of linking operations that may be performed in the final delineation step. By using a path finder such as the one developed by Fischler et al. (1981), the systems can then check the validity of the links, close the open-ended enclosures and outline new regions.

The closed networks delineate generic objects and constitute the final output of the system.

4.2. Examples

In this system, the geometric knowledge is encoded as rules of the form

IF *Pattern Match*

THEN *Operate on Data Structure*.

The rules allow the system to detect inconsistencies and modify geometric relationships, eventually yielding geometrically consistent structures. Although we do not have space here to describe the entire body of rules in detail, we provide the following examples to illustrate their function:

• *Edge-parsing rules.* Consider a set of edges forming a stair-step, as shown in Figure 7. First the horizontal edges are merged into composite edges forming a parallel structure. The process would stop there if the vertical edges were not present, but these vertical edges are used to generate the hypothesis that other vertical edges may have been lost. This hypothesis is then tested by checking whether the parallel structure can be broken in a way that is consistent with the presence of a new edge. If it can be broken, the horizontal edges are broken into sub-edges, yielding the final parse; *two distinct, related enclosures.*

• *Resegmentation rules.* Suppose the system has built the network, formed by an open-ended parallel and a U that form a T, as shown in Figure 8. The T relationship is used to hypothesize the location of the links. The prediction is then checked by the path finder. If the links are found, they will be used to close the whole network, forming a new region; if not, the T relationship will be broken and the system will generate two separate regions, one for each structure.

These two examples illustrate the following features of the system:

• Hypotheses about a geometric structure can be systematically made and refuted.

• The validity of a hypothesis made on the basis of high-level knowledge can be checked by re-examining the low-level data.

Redundancy is an important characteristic of our rule-base design; several different consistent paths of making and breaking associations will lead to the same or equivalent structures. In our edge-parsing example, several reasons might have lead to the merging of two atomic edges into a composite one: the edges being contiguous and having similar directions, their being both parallel to a third edge, or their being part of two L-related

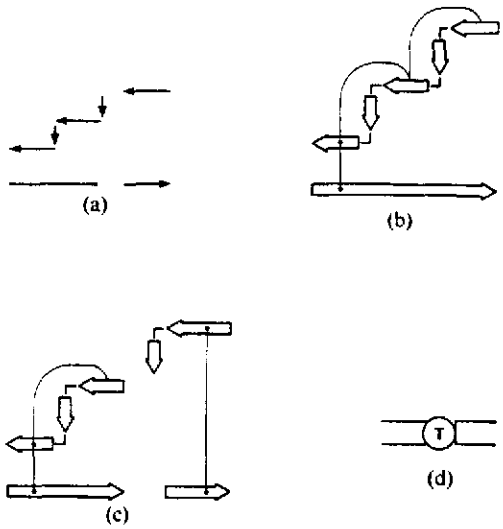


Figure 7. The parsing of a parallel structure with steplike internal structure. (a) The atomic edges. (b) The composite edges merged into a composite parallel. (c) Breaking the composite parallel. A vertical edge is evidence for breaking the original composite edges, thus breaking the parallel structure as well. (d) Final symbolic parse: a parallel on the left forms the stem of a T structure; the T structure is joined to the linking edge of the U on the right.

enclosures. In general, because of noise, not all three cures will be usable, but there is a high probability that at least one will, which helps to make the result relatively stable in the presence of noise and lost edges.

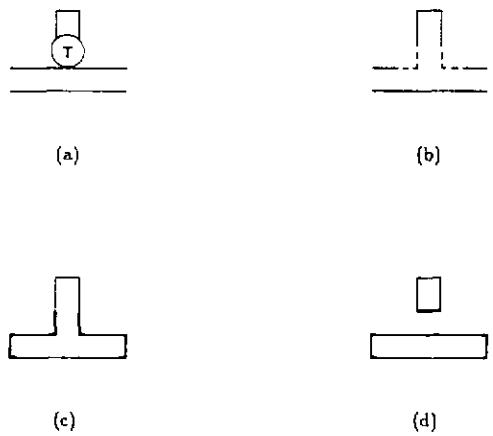


Figure 8. Resegmenting a network. (a) The initial T-network formed by a parallel and a U. (b) The predicted links between the two enclosures. (c) The final resegmented region if the links really exist. (d) The links were not found. The system chooses the alternate hypothesis and closes each structure separately.

5. Results

We now apply the entire procedure to a series of images containing complex cultural objects that would be difficult to extract using more conventional approaches.

For this application, a context model must be provided to supplement the fundamental geometric reasoning capabilities of the system during the clustering phase. The following model has proven sufficient to extract buildings from a wide class of grey-level aerial images:

The different parts of buildings seen in an aerial image at moderate resolution are made of uniform material. Differences in image intensities of these parts are due to differences in the illumination angle or shadowing of slanted surfaces. Three-dimensional structures may, or may not, have detectable shadows.

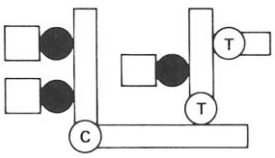


Figure 9. The resultant symbolic representation of the parse network of the entire building structure.

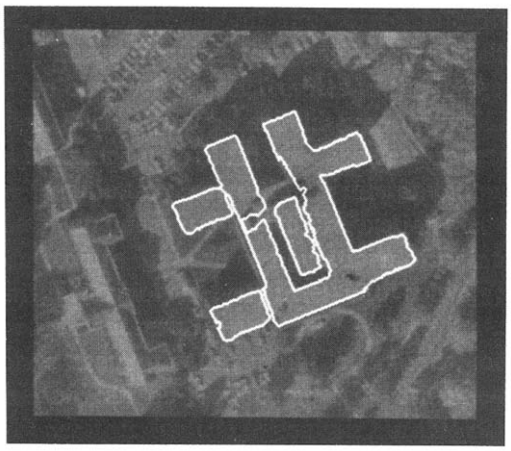


Figure 10. The result of using the parsed geometry to predict region closing paths and joining operations, yielding the final semantically motivated building shape.

In Figure 1a, we show a very complex building scene. Starting from the Ohlander-style segmentation overlaid in Figure 1b, in which the central building is completely broken up, the system extracts geometric structures. Using our model, the system finds the network of associations depicted symbolically in Figure 9 that can be interpreted as a flat roof; recall that this symbolic network stands for a complete representation of the object using the internal data structures of the system. Running the linking procedures, the system delineates the complex building structure, as shown in Figure 10.

Next, in Figure 11, we show a pair of images of the same house digitized from different sources at different resolutions. In one case, the house is segmented into one single region; in the other, into

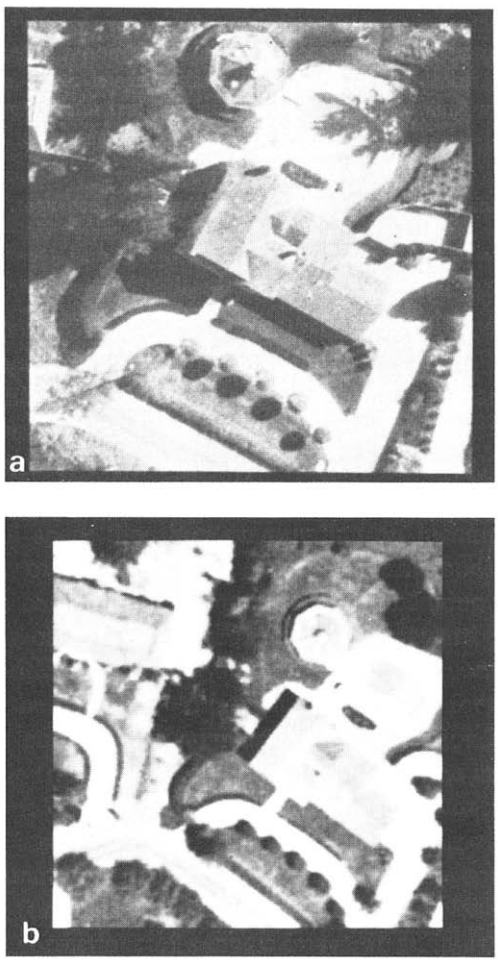


Figure 11. (a) A high-resolution aerial image containing a complex building. (b) Low-resolution image of the same building on a different day.

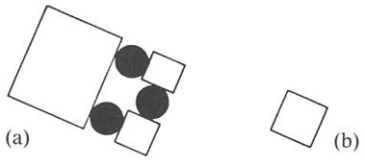
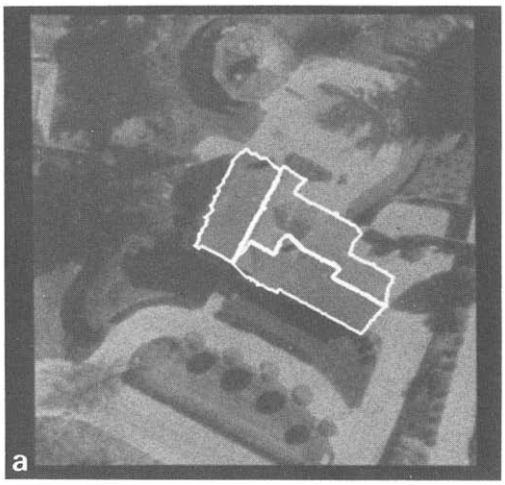


Figure 12. The resultant symbolic representations of the parse network of the building structure at the two resolutions, (a) and (b).

three regions corresponding to each of the various roof parts. Parsing these gives the two symbolic networks of Figure 12, and the resultant structure delineations in Figure 13. The final results are different; however, the three regions in the high-resolution image form a single network and are therefore considered as one cultural object. Based



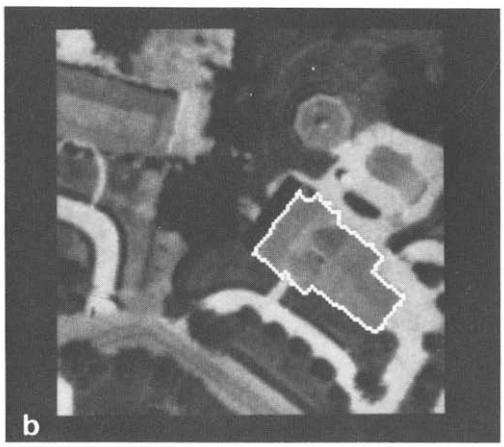


Figure 13. The result of using the parsed geometry to predict region closing paths and joining operations, yielding the final semantically-motivated building shapes at the two resolutions (a) and (b).

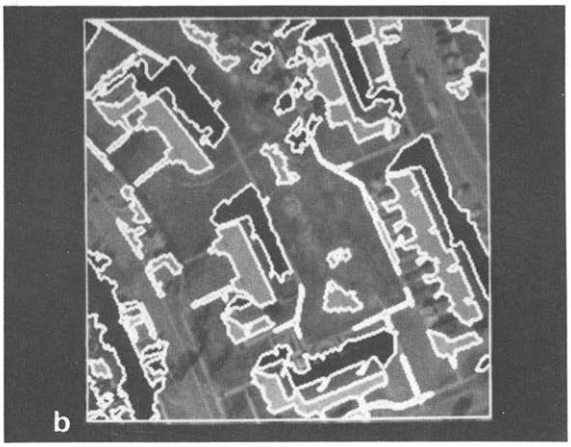


Figure 14. (a) An aerial image containing many shaded buildings. (b) A partition of the image.

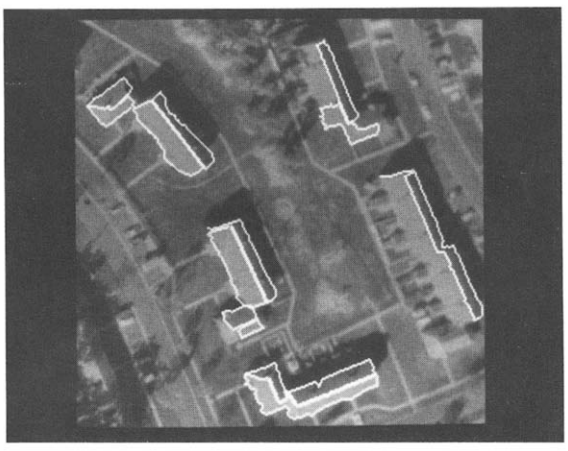


Figure 15. The result of using the resegmentation procedure to outline all the identifiable building areas. At this resolution, the regions corresponding to some substructures of the buildings, had insufficient geometry to be detected.

on semantic grounds, they can therefore be merged. The resulting region is then almost identical to the corresponding region in the low-resolution image. This example illustrates how this system can be used to provide consistent parses of the same objects seen in very different conditions.

Finally, we examine the image in Figure 14a, which contains a large number of buildings with multiple parts and shaded roofs. The image partition in Figure 14b shows substantial problems with the segmentation because the shaded roofs are confused with the background and sidewalks merge with the sunny parts of the roofs. Using the same model as before, the system finds the building portions delineated in Figure 15.

6. Conclusions

In this work we have proposed an approach to the problem of finding instances of generic objects in an image using model-driven resegmentation. Much of the system's effectiveness derives from the close interaction between the low-level and high-level information. The approach was implemented and tested for the case of generic rectilinear cultural objects; we have shown that, in this domain, we can use a single untuned but relevant segmentation to extract instances of complex generic objects. We have thus achieved the following objectives:

- *Object delineation using resegmentation.* Because segmentation regions often do not match objects of interest, we have developed models that allow us to correct the initial segmentation and generate semantically meaningful regions. The resegmentation procedure incorporates significant semantic knowledge about the object domain. Our results correspond very closely to regions containing target objects.
- *Generic shape extraction.* For many important tasks, the exact shapes of objects of interest are not known. We define and use generic models to deal with whole classes of objects.
- *Robust parsing of real image data.* Going from the raw data to its symbolic representation is a difficult task. We have built a rule base that is robust enough to parse incomplete and ambiguous image information.

In future work, we shall relax some of the constraints assumed here. Noncultural and nonrectilinear objects, for example, have boundaries characterizable as smooth or jagged curves. We will add such classes of objects to our analysis by substituting different statistical measures for those used here to extract straight edges. By adding support for the utilization of multiple image partitions, we will enable the system to generate and test a wider variety of shape hypotheses, thereby compensating more effectively for undersegmentation and oversegmentation of target objects.

References

Ballard, D.H. (1981). Generalizing the Hough transform to detect arbitrary shapes. *Pattern Recognition* 13, 111–122.

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

Binford, T.O. (1982). Survey of model-based image analysis systems. *Internat. J. Robotics Research* 1(1), 18–64.

Blum, H. (1973). Biological shape and visual science (Part I). *J. Theoretical Biology* 38, 205–287.

Brooks, R.A. (1981). Symbolic reasoning among 3-D models and 2-D images. *Artificial Intell. J.* 16.

Fischler, M.A. J.M. Tenenbaum and H.C. Wolf (1981). Detection of roads and linear structures in low-resolution aerial imagery using a multisource knowledge integration technique. *Computer Graphics Image Processing* 15, 201–223.

Fua, P. and A.J. Hanson (1985). Locating cultural regions in aerial imagery using geometric cues. *Proceedings of the Image Understanding Workshop*, 271–278.

Laws, K.I. (1984). Goal-directed texture segmentation. Technical note 334, Artificial Intelligence Center, SRI International, Menlo Park, CA.

McKeown, D., W.A. Harvey and J. McDermott (1984). Rule-based interpretation of aerial imagery. *Proc. IEEE Workshop on Principles of Knowledge-Based Systems*, 145–157 (1984); *IEEE Trans. PAMI* 7 (1985) 570–585.

Nevatia, R. and K.R. Babu (1978). Linear feature extraction. *Proc. Image Understanding Workshop*, 73–78.

Nevatia, R. and K.R. Babu (1980). Linear feature extraction and description. *Computer Graphics Image Processing* 13, 257–269.

Nevatia, R. and A. Huertas (1985). Building detection in simple scenes. Intell. Systems Group Technical Report, Univ. of Southern California.

Ohlander, R., K. Price, and D.R. Reddy (1978). Picture segmentation using a recursive region splitting method. *Computer Graphics Image Processing* 8, 313–333.

Rosenfeld, A. (1986). Axial representations of shape. *Computer Vision, Graphics and Image Processing* 33, 156–173.

Shirai, Y. (1978). Recognition of man-made objects using edge cues. In: A.R. Hanson and E.M. Riseman, *Computer Vision Systems*, Academic Press, New York.

Tavakoli, M. (1980). Toward the recognition of cultural features. *Proc. Image Understanding Workshop*, 33–57.